JOURNAL *of* ETHICS
& SOCIAL PHILOSOPHY

VOLUME XX · NUMBER 1
July 2021

ARTICLES

1 Meaning in Life and Becoming More Fulfilled
W. Jared Parmer

30 Friendship as a Non-Relative Virtue
Rachel Z. Friedman

56 Does Convergence Liberalism Risk Anarchy?
Marcus Schultz-Bergin

82 Elusive Reasons and the Motivational Constraint
Benjamin Cohen Rossi

DISCUSSION

111 Why Extending Actions through Time Can Violate a Moral Right to Privacy
Björn Lundgren

JOURNAL *of* ETHICS & SOCIAL PHILOSOPHY
http://www.jesp.org

The *Journal of Ethics and Social Philosophy* (ISSN 1559-3061) is a peer-reviewed online journal in moral, social, political, and legal philosophy. The journal is founded on the principle of publisher-funded open access. There are no publication fees for authors, and public access to articles is free of charge and is available to all readers under the CREATIVE COMMONS ATTRIBUTION-NONCOMMERCIAL-NODERIVATIVES 4.0 license. Funding for the journal has been made possible through the generous commitment of the Gould School of Law and the Dornsife College of Letters, Arts, and Sciences at the University of Southern California.

The *Journal of Ethics and Social Philosophy* aspires to be the leading venue for the best new work in the fields that it covers, and it is governed by a correspondingly high editorial standard. The journal welcomes submissions of articles in any of these and related fields of research. The journal is interested in work in the history of ethics that bears directly on topics of contemporary interest, but does not consider articles of purely historical interest. It is the view of the associate editors that the journal's high standard does not preclude publishing work that is critical in nature, provided that it is constructive, well-argued, current, and of sufficiently general interest.

Journal of Ethics and Social Philosophy
VOL. 20, NO. 1 · JULY 2021
https://doi.org/10.26556/jesp.v20i1.978

MEANING IN LIFE AND BECOMING MORE FULFILLED

W. Jared Parmer

Insofar as meaning as applied to lives is a value, a common question is whether meaning is "objective" or "subjective." When this question has to do with what makes a life meaningful, answering it is a matter of finding out whether only things with objective value can do so.[1] In this context, to say that only things with objective value can make a life meaningful is in part to say that meaning has a necessary objective value condition. A theory that denies this will have to say that things can make a person's life meaningful for her independently of their connection to anything of objective value. Now, presumably, such meaning-makers will do so, at least in part, via their connection to contingent features of a person herself. So any theory that denies that meaning has a necessary objective value condition will be subjectivist where the rubber meets the road.[2] Accordingly, I call a theory *objectivist* just in case it says that meaning has a necessary objective value condition, and I call a theory *subjectivist* just in case it denies this.[3]

1 As opposed to a concern with what the concept or property of meaning consists in. For accounts at those levels of analysis, see Brogaard and Smith, "On Luck, Responsibility, and the Meaning of Life"; Kauppinen, "Meaningfulness and Time"; Martela, "Meaningfulness as Contribution"; and Metz, "The Concept of a Meaningful Life," "The Meaningful and the Worthwhile," "The Meaning of Life," and *Meaning in Life*, ch. 2.

2 The "at least in part" qualification is important: a subjectivist theory can appeal to objective conditions, provided they are not objective *value* conditions. Thanks to Barry Maguire for helping me frame this.

3 Examples of subjectivist views include Calhoun, *Doing Valuable Time*, ch. 2; and Taylor, *Good and Evil*. Darwall (*Impartial Reason*, chs. 11–12) and Wong ("Meaningfulness and Identities") can also be seen as subjectivists, though their theories are distinctly *inter*subjectivist. Non-subjectivist views include Evers and van Smeden, "Meaning in Life"; Kauppinen, "Meaningfulness and Time"; Kekes, "The Meaning of Life"; Levy, "Downshifting and Meaning in Life"; Metz, *Meaning in Life*, ch. 12; Smuts, "The Good Cause Account of the Meaning of Life" and *Welfare, Meaning, and Worth*; Wielenberg, *Value and Virtue in a Godless Universe*; and Wolf, *Meaning in Life and Why It Matters* and *The Variety of Values*. Bramble is commonly called an objectivist because he rejects "The Passion Requirement"

To repeat: this condition has to do with whether *only* things of objective value can make a person's life meaningful for her.[4] So it is logically possible to be a subjectivist about meaning in life by claiming that *some* things without objective value can make a person's life meaningful for her, while allowing that, or being agnostic about whether, objectively valuable things can *also* do so. In other words, one can be a subjectivist just by identifying some contingent features of persons that, independently of those features' connection to anything of objective value, make those persons' lives meaningful for them. In doing so, one identifies a genuinely subjective source of meaning. Because subjectivism about meaning remains rather unpopular among contemporary theorists and viewed by them as straightforwardly refuted, a defensible and compelling subjectivism of even this modest sort should be of interest.

In this paper, I argue that a relatively sophisticated but modest subjectivist theory, the *becoming more fulfilled view*, is both defensible and compelling.[5] The view is that a person's becoming more fulfilled makes her life meaningful for her. Becoming more fulfilled is a process that has being more fulfilled as its hypothetical endpoint. More specifically:

> *The Becoming More Fulfilled View*: A person *S*'s becoming more fulfilled by *x* makes her life meaningful for her when, and only when, *S* aims to do activities $\mathcal{\phi} = \{\phi_1, \phi_2, \ldots\}$ *well*, where $\mathcal{\phi}$ing well

(his language) on meaning in life; on my way of carving up things, however, this is a mistake, since Bramble *also* rejects the necessary objective value condition (what he calls "The Objectivity Requirement"). For him, rather, subjective goodness and objective value figure in *sufficient* conditions for meaning in life ("Consequentialism about Meaning in Life").

4 I will insist on using "for her"–type qualifiers throughout this paper. See section 4 for my reasons. As I use them, these qualifiers do not fix the referent to how meaningful each person *thinks* or *feels* his or her life is. The latter interpretation is not by any means obligatory. Consider the following. In the critical commentary that appears in Wolf's *Meaning in Life and Why It Matters*, Nomy Arpaly says that "being in charge of a beloved goldfish or two can give [a severely mentally disabled] child a measure of fulfillment that would require much bigger projects in a normal adult—but for the same reasons and via the same mechanisms. Thus, in the case of the child it is not strange to say that goldfish keeping gives his life meaning" ("Comment," 89). When considering such a case, it is felicitous to say that a life of goldfish caretaking is *meaningful for* this child. One can do this even while denying that one is making a claim about how meaningful that child thinks or feels her own life is. In fact it is plausible that she lacks the reflective and affective capacities to have such an attitude toward her own life at all. The reader is free, of course, to disagree with the assessment itself. My argument is not an argument for the *truth* of the assessment, but about *what* the content of the assessment *is*.

5 Bramble's view is rather closer to mine in its modest spirit ("Consequentialism about Meaning in Life"). See note 3 above. I thank an anonymous reviewer for encouraging me to take this tack.

a. at least partly constitutes benefiting *x*, and either
b. requires caring more deeply and richly about *x* than *S* has so far, or
c. requires doing more of $\{\phi_1, \phi_2, \ldots\}$ than *S* has so far.

Section 1 lays the groundwork for this view with a discussion of fulfillment and the temporal dynamics of caring, which partly constitute fulfillment. Section 2 motivates the becoming more fulfilled view and spells it out more, though I leave the discussion of how to extend the view to account for *degrees* of meaning for section 4, allowing the discussion of cases there to raise the issue organically.

Section 3 responds to a pair of arguments against subjectivism, the first due to Susan Wolf and the second due to Antti Kauppinen and Aaron Smuts.[6] Discussing a well-known variant of Sisyphus who has every subjective quality that could plausibly matter for meaning, Wolf claims that the fact that his activities are pointless, unproductive, and futile is evidence that his life is meaningless. If she is right, she has identified a counterexample to every plausible version of subjectivism. Her claim is false, however, because Sisyphus's activities are not pointless, unproductive, and futile. Smuts and Kauppinen argue that subjectivism implies, falsely, that no person can be mistaken about how meaningful her own life is. However, subjectivism as such does not imply that. I explain why and then illustrate this with the becoming more fulfilled view.

Taking a step back, the major motivation behind rejecting subjectivism is the thought that it will always produce counterintuitive results: that every variant of subjectivism will count as meaningful a wide range of intuitively meaningless lives. Section 4 addresses this charge head-on and argues that it is not so, at least when it comes to the becoming more fulfilled view. Once we spell out the lives in question in further detail, and we are explicit about exactly in what way we are assessing them, we see that the view produces broadly intuitive results.

1. Being Fulfilled and the Dynamics of Caring

For subjectivism about meaning, a natural place to begin is with the view that a person's being fulfilled makes her life meaningful for her.[7] In this section, I will elaborate this view; in the next, I will motivate going beyond it and spell out my preferred view.

Before I begin, let me head off a misunderstanding: *feeling* fulfilled is not the

6 Wolf, *Meaning in Life and Why It Matters*; Kauppinen, "Meaningfulness and Time"; Smuts, "The Good Cause Account of the Meaning of Life."

7 Though they use the language of valuing rather than caring, Calhoun (*Doing Valuable Time*, ch. 2) and Wong ("Meaningfulness and Identities") can be seen as offering views like this.

same thing as *being* fulfilled. The relationship between these states is analogous to that between, say, feeling afraid and being in danger. Though the relationship between these two states is a matter of philosophical debate, it is a clear mistake to identify them one with another. Feeling fulfilled will henceforth play no role in the plausible subjectivist views I consider, so we can set it aside.

I assume that a person's being fulfilled by some *x* (whether a person, thing, or activity) is a matter of caring about *x* and doing what caring disposes her to do. I take this assumption to be plausible and to generally comport with our intuitions. I also endorse a common view about a person's caring about *x*: she exhibits various familiar emotional, cognitive, motivational, attentional, and physiological dispositions focused on *x*, where these dispositions together constitute *x*'s mattering to her.[8] So, for example, a person's caring about her friend involves dispositions to feel anxiety over his upcoming travails; to believe that his needing her help is a reason for her to do so, and to be motivated accordingly; to notice when he is uncomfortable; to be excited to see him after a long separation; and many more besides. These dispositions constitute this man's mattering to her. Because caring is a complex *dispositional* state, however, being fulfilled by *x* is more *active* than merely caring about *x*: being fulfilled by *x* involves actually doing what caring about *x* disposes one to do. To emphasize this, I will sometimes speak of fulfillment as *caring engagement*.

Being fulfilled by some *x* comes in degrees, which is a function of its components that themselves come in degrees. It should already be clear what it is for a person to do more of what her caring disposes her to do. However, the degree to which a person cares about some *x* requires a little explication. I take this to be a matter of the *depth* and *richness* of her caring about *x*. To put it briefly, depth is a matter of the *intensity* of the responses a person is disposed to manifest in caring about *x*—the intensity of emotions, strength of motivational pull, weight of perceived normative reasons, and variety and extremity of multimodal (visual, aural, etc.) focus involved. By contrast, richness is a matter of the *variety* of the disposed responses. Broadly speaking, deeper and richer caring typically happens as the person's conception of *x* is developed via her continued engagement with it. So, for example, as a person comes to see that her acquaintance gets very anxious in formal social settings, she can become disposed to attend even more intently to his body language than she was before (thus deeper caring), or she

8 See, for example, Helm, *Love, Friendship, and the Self*; Jaworska, "Caring and Internality"; Kolodny, "Aims as Reasons," sec. 8; Maguire, "Love in the Time of Consequentialism"; and Seidman, "Valuing and Caring." Scheffler ("Valuing") and Kolodny ("Love as Valuing a Relationship") build highly similar accounts of valuing. For an in-depth look at *focus* as involved in caring, see especially Helm, *Love, Friendship, and the Self*.

can become newly disposed to whisk him away to a quiet corner at parties (thus richer caring).

These processes of enrichment and deepening can take place such that a person comes to care about *x* when she antecedently did not. This is because caring requires a certain amount of richness and depth in a person's dispositions focused on *x* for her to count as caring about *x* at all: she has to be disposed vis-à-vis *x* to feel, think, attend, and act in at least somewhat intense and various ways. So it is possible to exhibit some such dispositions focused on *x*, which are themselves part and parcel of caring about *x*, without caring about *x*. This will matter shortly.

Returning to the main thread, the basic idea might be that the degree to which a person is fulfilled by various persons, things, and activities in her life is a function of the extent to which she is caringly engaged with them. The latter is, in turn, a matter of how deeply and richly she cares about them, and to what extent she engages with them in what she does.

I will say that a person can be *more* fulfilled by *x* in situation S_1 than by *y* in S_2 to the extent that, either

a. given some activity ϕ she does in S_1 and some activity ψ she does in S_2, she cares about *x* more deeply or richly in S_1 than she cares about *y* in S_2, where her caring about *x* disposes her to ϕ and her caring about *y* disposes her to ψ; or
b. given that she cares equally richly and deeply about *x* in S_1 as about *y* in S_2, she does more activities $\{\phi_1, \phi_2, \ldots\}$ in S_1 than $\{\psi_1, \psi_2, \ldots\}$ in S_2, where her caring about *x* disposes her to $\{\phi_1, \phi_2, \ldots\}$ and her caring about *y* disposes her to $\{\psi_1, \psi_2, \ldots\}$.

In reality, of course, some mixture of the two is often the case.

Accordingly, the view that a person's being fulfilled makes her life meaningful for her can be extended: a person's life is *more* meaningful for her in S_1 than in S_2 because, and to the extent that, she is more fulfilled by the things, persons, and activities in her life in S_1 than in S_2. I will call this extended view *the being (more) fulfilled view* of meaningfulness.

Now, caring engagement often deepens and enriches our caring. As we act for the sake of what we care about, we learn more about how to care about it. We come to see more clearly what its weal and woe consists in and the sorts of situations that affect either, and accordingly become disposed to respond to such situations in the emotional, motivational, cognitive, and attentional ways that constitute richer and deeper caring about it. For example, at one time I cared about doing philosophy for its own sake, but possessed a rather sophomoric view of what doing philosophy was. And yet by doing it I came to see, among

other things, that doing it *well* involves sustained and careful engagement with one's interlocutors. Prior to this discovery, I was not worried about, say, failing to read an important text on the topic I had chosen to write on; I do worry about that now. So the ways in which I am disposed to think and feel, in caring about doing philosophy, have changed—my caring has grown deeper and richer.

The weal and woe of what we care about, it needs to be emphasized, is not obviously a matter of objective value. A person's weal and woe has to do with contingent features of herself, such as what she likes and what she needs to survive. So to see more clearly what the weal and woe of a person consists in is to see what is good *for her*, which does not obviously concern what is objectively good about her or anything else that is objectively good. And the weal and woe of various practices (as distinguished from their *products*), such as making art or doing philosophy, has to do with what it is to make art or do philosophy *well*, which is determined by what these practices are.[9] So to see more clearly what the weal and woe of such practices consists in is to see what is good as an instance of this practice, which does not obviously concern what is objectively good about it as such or about anything else. So the attunement at hand need not, for all we know, be attunement to objective values; objective value, therefore, need play no explanatory role in these dynamics. Or, at a minimum, the objectivist owes us an argument to that effect.[10]

However, just as finer attunement can deepen and enrich a person's caring about something, such attunement can also lead to detachment when her caring depended on misconceptions from which she gets disabused. Suppose, for example, that a woman named Sophie has taken up philosophy in a serious way after exciting exposure to it as an undergraduate when she read the provocative and insightful work of writers such as Nietzsche and Kierkegaard. She finds reading and writing philosophy engrossing, and capable of producing pleasure and frustrating confusion in equal measure. She gets a thrill from every new project she takes up, believing that, as she perseveres, satisfaction or failure lurk just beyond sight. As she progresses with her plans, writing a senior thesis, attending summer programs, graduate school, and so forth, however, her understanding of philosophy as a practice slowly changes. She comes to see, for example, that engaging with her heroes in a respectable way requires extensive archival research and grappling with the turgid prose of Kant and Hegel; that the fruits of such labor will be arguments that, however compelling she finds them at the time, will not be so compelling that none of her opponents can reasonably resist their con-

9 Cf. Thomson, *Normativity*.

10 Cf. the claims made by Metz, *Meaning in Life*, 175–76. My thanks to Nadeem Hussain for prompting me to address this issue.

clusions; and that she herself will, in time, find many of those arguments of hers deeply flawed. As she understands philosophy better in this way, her attachment to it gradually slips loose until the prospect of spending her time and energy on this leaves her cold. She no longer cares about doing philosophy—she struggles to go to the archives, or to write papers, or even to discuss it with friends over beer; she stops worrying about getting papers published, or hoping she will polish off a new theory of this or that; her attention strays when in seminar or when, on rare occasions, she does sit down with a text; and so on. She is *ipso facto* no longer fulfilled by doing philosophy.

Though unfortunate, such moments should be familiar enough. In acting for the sake of someone we care about, we sometimes come to see more clearly who they are in such a way that new divisions arise between us. Sometimes by doing philosophy, or practicing law, or being famous, or getting married, we come to see that these things prove to be other than we thought—perhaps grinding or tedious, or with benefits that lie in places that fail to draw our appreciative gaze. The point here is not that our positive *feelings* of fulfillment, satisfaction, and so on often, even tend to, return to a baseline that disappoints us—the so-called hedonic treadmill. The point, rather, is that to be fulfilled by something or someone is to hazard a great personal risk, for in doing so a person makes herself vulnerable to loss—not just because what matters to her might be lost, but because its mattering to her might be.

2. Becoming More Fulfilled

The being (more) fulfilled view gets some initial grip. It inherits the virtues of subjectivist theories about meaning in life, while eschewing an implausible overemphasis on occurrent *feelings* of fulfillment. Even so, there is some reason to look for an alternative.

Let us continue with Sophie, and suppose that, though she is no longer fulfilled by doing philosophy, she resolves to persist for a while. This she does largely due to her history with it, including not only her past fulfillment by it, but also how her caring about it was subject to learning more about it. We can imagine, in particular, that she thinks that to abandon philosophy now would be a disservice to herself in light of what had mattered so much to her for so long; and this would be a disservice because, as she has experienced, learning more about philosophy has the potential to alter what matters to her—and so, perhaps, the potential to make it matter to her anew.

By persisting in this way, Sophie does so without being fulfilled by doing philosophy. However, her life is more meaningful for her by persisting than it would

be if she abandoned doing philosophy entirely (all else equal). We have, then, a case in which her life is *more meaningful* for her by persisting in philosophy than it would be by abandoning philosophy, even while she is *not more fulfilled* by persisting in philosophy than by abandoning it.

This greater meaningfulness plausibly has, at least in part, a subjective source (outside of, especially, whatever objective value doing philosophy might have). One key factor, recall, is that doing philosophy had mattered to her so much for so long, and this is the context in which she made the resolution she did. Had she made no such resolution, but had rather carried on in a kind of unreflective drift, hemmed in by habit and her prior plans, the meaningfulness for her of continuing to do philosophy would be attenuated. On the other hand, if she had resolved to do so but not against the backdrop of philosophy mattering to her so much and for so long, her persistence would look rather quixotic or arbitrary, which would also attenuate the meaningfulness of doing so.

Cases like Sophie's give us some reason to look for an alternative to the being (more) fulfilled view while remaining within a subjectivist framework. Still, it is true that, by persisting in philosophy, Sophie endeavors in a way that, if "successful" in some sense, will result in her being more fulfilled by doing philosophy than she is at present. And this suggests the following defense of the being (more) fulfilled view: Sophie's life is not more meaningful for her by persisting in philosophy, not directly anyway. Rather, by persisting, she does something that bears an instrumental relationship to meaningfulness: persisting is a way of bringing about her own fulfillment in the future.

However, we should not accept this as the whole story. For one thing, the intuition at hand, that Sophie's life is more meaningful by persisting in philosophy, is preserved even if her endeavors fail to bring about future fulfillment. We could imagine, for example, that before Sophie comes to care again about doing philosophy, she is tragically struck dead by a truck. This would not render her perseverance meaningless for her—it still made her life more meaningful for her than it would have been had she abandoned philosophy entirely.

For another, if it were right that Sophie's endeavors were only instrumentally valuable as far as meaningfulness goes, then the following prescription would be apt: as far as meaningfulness is concerned, she should do whatever is most likely to bring about the most fulfillment—persist in philosophy, run off to Hollywood to become a star, marry that boy who proposed to her on their third date, etc. But this is a strange prescription. It misses the significance of the fact that Sophie's endeavors here are meaningful at least in part because they enact the resolution she has made against the backdrop of what had mattered to her so much and for so long. Now, it might be said that the significance of this fact is that it

makes the particular path she has chosen more likely to result in fulfillment as compared to, say, running off to Hollywood. But this response helps itself to more than the case allows. What is distinctive of the case at hand is precisely that Sophie's caring about doing philosophy turned out to be predicated on a serious misapprehension of what doing philosophy involves. So the fact that philosophy used to matter to her does not make it more likely that philosophy (compared to, say, running off to Hollywood) will come to matter to her again as she learns more about it.[11]

The being (more) fulfilled view treats fulfillment as a state that makes a person's life meaningful, and assesses the value of endeavors like Sophie's in terms of their instrumental relations to that state. But this explanation seems to falter for the two reasons I have just given. So let this be some motivation to look beyond the being (more) fulfilled view.

2.1. Being versus Becoming More Fulfilled

Let me reiterate that, common across the cases I have been emphasizing, people living meaningful lives can be seen as endeavoring in such a way that, should they succeed, they will end up more fulfilled than they presently are. And this is due in part to the fact that success involves deeper and richer caring about what they are doing. Thus, being more fulfilled can indeed be seen as a *hypothetical endpoint* by which subjectivist theorists about meaning understand the endeavors of people living meaningful lives. But it is not, for all that, what *makes* their lives meaningful for them, where their endeavors are merely instrumental thereto.

This is to grant that a person is *becoming* more fulfilled by some *x* only when her endeavors have as their hypothetical endpoint that she is more fulfilled by *x*: she is undergoing a particular kind of process, one that has greater fulfillment as its endpoint. However, I suggest that it is not the endpoint, but the process itself, that makes her life meaningful for her. This, as I elaborate in what follows, is *the becoming more fulfilled view* of meaning in life.

The language of "becoming" emphasizes the processual nature of this subjective source of meaning, distinct from but intimately related to the state-like nature of being more fulfilled. The relationship between being and becoming more fulfilled is analogous to that between *having* more cocktails ready for a party and *making* more of them: the former is a state, the latter is a process that has the former as its endpoint.[12] And, just as it is possible for a person to be making

11 I thank an anonymous reviewer for pressing me to expand my response here.

12 Of course, in this pair, having more cocktails ready is usually the *non*-hypothetical endpoint in the sense that, while making more cocktails, having them ready is indeed what the person

more cocktails for a party over some finite span of time without having (made) more such cocktails at the end of that span or at any later time (perhaps because she realizes halfway through that she is out of bitters), it is possible for a person to be becoming more fulfilled over some finite span of time without being more fulfilled at the end of that span or at any later time. She surely needs to be taking the steps in a sequence of requisite steps such that, should she complete them, she will end up more fulfilled. But this is compatible with not being more fulfilled because she might never take every requisite step, whether of her own doing or the world's. As with any process, becoming more fulfilled can come to a halt before it is complete.

There is a thin sense in which *any* process that has greater fulfillment as its hypothetical endpoint is a process of becoming more fulfilled. However, not just any such process makes a person's life meaningful for her via a subjective source. Imagine, for example, that a benevolent mad scientist labors over Sophie's brain every night while she sleeps, for a very long time, so she ends up caring a great deal about doing philosophy. And imagine, moreover, that he does this precisely because Sophie persists in philosophy. In some loose sense, her endeavors have greater fulfillment by doing philosophy as their hypothetical endpoint—her doing philosophy is part of the explanation as to why she ends up more fulfilled by doing philosophy. But her being in this process does not seem to make her life more meaningful for her via a subjective source.

So let us look for a particular process of becoming more fulfilled that might meet our needs. From here on out, I will use the labels "the process of becoming more fulfilled" and "becoming more fulfilled" to speak *only* about the particular process that makes a person's life meaningful for her via a subjective source. I grant that there are other processes for which those labels might be apt, but they will not be my focus in what follows.

The mad-scientist example helps us get started. What is missing in that case, I suggest, is that the connection between her actual endeavors and the hypothetical endpoint is too *indirect* for the process she is undergoing to make her life meaningful for her via a subjective source. At the same time, however, this connection should not be too *direct*. Namely, it should not be that her endeavors have this hypothetical endpoint because she has her own greater fulfillment as her *aim*. Sophie can be becoming more fulfilled by doing philosophy, where this is a subjective source of meaning for her, without that aim; in the simplest case,

is intentionally trying to achieve. For further discussion of processes in the context of action theory, see, among others, Paul, "Embarking on a Crime"; Stout, *Process, Action, and Experience*; Thompson, *Life and Action* and "Anscombe's Intention and Practical Knowledge"; and Wolfson, "Agential Knowledge, Action and Process."

she will be doing philosophy for its own sake. Indeed, having her own greater fulfillment as one of her aims might be self-defeating if the "paradox of hedonism" is true for fulfillment, which would entail that a person *cannot* have her own greater fulfillment as one of her aims if her endeavors are to have this hypothetical endpoint.

Let me briefly summarize where we are. I suggested moving beyond the *being* (more) fulfilled view of meaning in life to the view that a person's *becoming* more fulfilled by some *x* makes her life meaningful for her. From an examination of an important case that motivated moving beyond the being (more) fulfilled view, I observed the following necessary condition:

> *A Necessary Condition on Becoming More Fulfilled*: When a person is becoming more fulfilled by some *x*, her endeavors vis-à-vis *x* have as their hypothetical endpoint that she is more fulfilled by *x*.

And by examining the mad-scientist example, and considering the live possibility that the paradox of hedonism is true of fulfillment, I motivated the following two constraints on spelling out more completely the process of becoming more fulfilled as it relates to meaningfulness:

> *The Not-Too-Indirect Constraint*: When becoming more fulfilled by some *x*, it cannot be merely that a person's endeavors vis-à-vis *x* would play *some* explanatory role in her downstream greater fulfillment by *x*, were she to attain it.

> *The Not-Too-Direct Constraint*: When becoming more fulfilled by some *x*, it (likely) cannot be that the person, in endeavoring as she does vis-à-vis *x*, *aims* to end up more fulfilled by *x*.

I now proceed to flesh out the becoming more fulfilled view within these latter two constraints.

2.2. The Becoming More Fulfilled View

The key composite of ideas—of acting in a way that has a particular hypothetical endpoint (in a relatively direct way) that the agent need not be directly aiming at—is an interesting and undertheorized area of philosophy of action.[13] Broad-

13 The closest that we have come to addressing this question, it seems to me, is by examining the relationship between *intending* to ϕ and ϕing *intentionally*. Michael Bratman and Al Mele have both denied that a person ϕs intentionally only if she intends to ϕ, and this can be seen as one way to prize apart a person's *aim* from the *endpoints* of her actions—where her aims are constituted by her intentions, and the endpoints of her actions are constituted by what she might intentionally do in virtue of her intentions. However, the relationship

ly speaking, there are two strategies for spelling this out. The first posits that the connection between the endpoint and the person's endeavors is secured by the *attitudes* she has toward that endpoint—such as attitudes that structure her downstream deliberations in such a way that bringing about that endpoint is likely, though not what she straightforwardly plans to do (perhaps, e.g., via higher-order planning states or values).[14] But this strategy would need to thread a very fine needle since, the more the person's own attitudes guide her actions toward the endpoint in question, the more it seems that that endpoint is something she *aims* to do.

The second strategy posits that the connection between the endpoint and the person's endeavors is secured by the features of her endeavors *de re* rather than by her attitudes surrounding and guiding those endeavors. I take this tack. Given the limitations of space and scope, I can only spell out and motivate this view here. Fully fleshed-out arguments *for* it will have to wait, though I will rebut arguments *against* it in sections 3 and 4. Here, in sum, is my idea.

> *The Becoming More Fulfilled View*: A person *S*'s becoming more fulfilled by *x* makes her life meaningful for her when, and only when, *S* aims to do activities $\Phi = \{\phi_1, \phi_2, \ldots\}$ *well*, where Φing well
>
> a. at least partly constitutes benefiting *x*, and either
> b. requires caring more deeply and richly about *x* than *S* has so far, or
> c. requires doing more of $\{\phi_1, \phi_2, \ldots\}$ than *S* has so far.

Importantly, ending up more fulfilled by *x* does not here need to be something *S* aims at *de dicto*; rather, what she aims at *de re* requires ending up more fulfilled by *x*. This strategy thus appeals, as it were, to the deep features of what the person aims to do, independently of her aiming to do it. Let me now flesh out and motivate this idea along a few dimensions.

First, Φing well benefits *x* in the sense that Φing well *(at least) partly constitutes* benefiting *x*. This is meant to rule out cases in which Φing well simply

between intending to ϕ and ϕing intentionally remains rather underspecified—for Bratman, the intentional action must be within the "motivational potential" of the intended action (*Intention, Plans, and Practical Reason*, ch. 8), while for Mele, the intended action must be "relevant" to the intentional action (*Springs of Action*, ch. 8). Moreover, this move has been met with sustained resistance by, for example, Adams, "Intention and Intentional Action"; McCann, "Rationality and the Range of Intention," "Settled Objectives and Rational Constraints," "Intentional Action and Intending," and "Di Nucci on the Simple View"; and Sverdlik, "Consistency Among Intentions and the 'Simple View.'" In any case, I doubt the most perspicuous way to spell out the matter of interest is to begin with a division between intention and intentional action.

14 Cf. Bratman, *Structures of Agency*.

causes something further that, on its own, benefits *x*.[15] Such a causal link, I take it, comes too cheaply to capture what we are after. At the same time, this formulation allows that ϕing well might *wholly* constitute benefiting *x*, though there might be few real-world cases in which that is so.

Second, it should be antecedently clear that nothing can benefit some *x* when *x* cannot fare better or worse. So this account is restricted to all and only *x*s that can fare better or worse. This rules out things like heaps of sand or mathematical truths, but includes any living thing and many nonliving things, provided that ϕing well, for some set of activities ϕ, can (at least) partly constitute benefiting them. I will not endeavor to give comprehensive analyses of faring better or worse, or, concomitantly, benefiting or harming; so much is clearly beyond the scope of this paper. I take it that our ordinary sense of these terms will do for now.

Let me illustrate these two points with examples. Statues can fall apart or corrode, paintings can fade or tear, people can be lonely or sick, institutions can be sclerotic or impotent, practices can lose structure or purpose, and so forth. In ordinary senses of the terms, these things can fare better or worse; accordingly, it is possible to benefit or harm them. For at least some such things, the activities of persons can, when done well, partly constitute such benefit (or harm). While it is implausible that doing anything well can itself partly constitute benefitting a statue, for example, matters are different for people, practices, and, perhaps, institutions. For example, it is plausible that doing philosophy well at least partly constitutes benefitting philosophy; as a practice, philosophy fares better when people are doing philosophy well, and not solely in virtue of the quality of whatever artifacts they produce along the way or however the practitioners (or consumers) of philosophy themselves benefit as a result. Much the same is true of other practices, such as cricket or contemporary dance. For another sort of example, it is plausible that the activities involved in being a good friend, when done well, partly constitute benefitting the person for whom one does them; and this benefit, too, is not solely in virtue of the causal results of those activities vis-à-vis anybody. The activities I have in mind are familiar ones, such as spending time with them and talking through their troubles—activities that benefit the person with whom one spends time or talks with, and not solely in virtue of whether those activities cause something further (such as good feelings or solutions to their problems).[16]

15 It does *not*, however, rule out cases in which ϕing well partly constitutes benefitting *x while also* causing something further that benefits *x*. There is good reason to allow such cases. See note 16.

16 My use of the locution "not solely in virtue of" here is meant to remain agnostic about the possibility that ϕing well partly constitutes a benefit to *x*, where this partial constitution it-

Third, I say that ϕing well requires that S care more deeply or richly about x than she has so far, or that it requires that S do more of $\{\phi_1, \phi_2, \ldots\}$ than she has so far. Since I take the latter disjunct to be clear enough, I will elaborate only on the former. The requirement is not an unrestricted metaphysical necessity, such as that it is metaphysically necessary to care about x to some relatively high degree when ϕing well. I doubt there is any such metaphysically necessary threshold. Rather, the requirement arises due to the kind of agent S is, including her abilities and limitations: it is necessary *for her* to care about x to some degree when ϕing well.

To see why this difference is important, consider again doing philosophy. It is certainly metaphysically possible for some agent to do philosophy well without caring a whit about it. We could imagine that she possesses immense cognitive capacities and very few alternatives, and can be motivated enough to do it on the slightest stimulation. Such an agent might be able to do philosophy well merely by contemplating a question and proceeding to slice and dice the logical space as long as it takes to come to a plausible answer. For us, however, things are obviously not so simple: a variety of alternatives compete for our attention and energy, many philosophical questions leave us unmotivated, and we cannot effortlessly recognize the wide range of options for answering such questions. Accordingly, for us, exploring and offering plausible answers to philosophical questions requires considerable cognitive, attentive, emotional, and motivational resources. In short, doing this work well has to *matter* to us in some way and to a significant degree. There are many practices like this, including sports, artistic endeavors, and "knowledge work." Of course, many practices require considerably fewer resources, such as the less demanding drinking game flip cup; still, to the extent that playing flip cup well requires developing the relevant skills, doing it well has to matter to us in some way and to some degree.

In interpersonal cases, things are similarly straightforward. When doing things well that partly constitute benefiting another person, we must summon cognitive, attentive, emotional, and motivational resources to attend to their

---

self depends on further features—which is plausibly the case when the various constituents of a benefit must form an "organic unity" to really *be* a benefit. This possibility is particularly salient in the friendship example: it is plausible that spending time with one another *does not* benefit each friend *without* concomitant good feelings, while, nevertheless, the benefit is not solely in virtue of those good feelings. This is plausible because it seems that, if we could through some sci-fi mechanism zap each friend into having the good feelings they would have from spending time together, it would still seem as though each friend was not faring as well as they would had they actually spent time together. I thank an anonymous reviewer for pressing me to elaborate how my theory applies to becoming more fulfilled by x when x is not itself an activity.

needs or desires, recognize what we can do to meet those needs or desires, and be motivated to act accordingly; and all of this we must do against a background in which a variety of other options competes for our attention and energy. Often, the requisite resources are considerable; sometimes the person in question needs quite a lot, or our antecedent motivations are too weak. Of course, sometimes this is not so: sometimes doing something well to benefit another person is rather easy. The point is just that the other person must *matter* to us, commensurate with the work that must be done to benefit them.

But I do not mean to overstate matters. This is the fourth and final elaboration I wish to make. When it comes to becoming more fulfilled by *x* in the sense relevant to meaningfulness, what matters is only that the requisite degree (in terms of depth and richness) of caring be *greater* than the degree to which the person has cared about *x* so far. This can be so whether the requisite degree is considerable or not.[17]

Let me now step back a bit. It should be clear that the becoming more fulfilled view postulates a genuinely subjective source of meaning. After all, the process is completed by the person in question ending up more *fulfilled* by the things in her life. Moreover, since the process is completed by ending up *more* fulfilled, it makes essential reference to what matters to her along the way. Finally, and perhaps most importantly, the endpoint of greater fulfillment is secured by what she actively aims to do along the way.

There are at least two aspects to this view that are not fully spelled out, and it is worth being explicit about this. First, I have offered no account of *how* a person comes to be more fulfilled by *x*. In light of the mad-scientist example above, which gave rise to the Not-Too-Indirect Constraint, it seems right that the person comes to be more fulfilled by *x* when she does, through ϕing, where

17 It should be clear that there will be cases in which person *S* already cares about *x* to the requisite degree, and already does well everything involved that at least partly constitutes benefiting *x*. On my account, she will thereby not be becoming more fulfilled by *x* in the sense that makes her life more meaningful for her. To be sure, such instances are typically preceded by becoming more fulfilled by *x*, and hence such a person's life will already be meaningful for her to some degree; I am here granting only that my account has it that her life is not made *more* meaningful for her, with respect to *x*, going forward. My account thus has a consequence worth being explicit about that concerns people living so-called completed lives. These people are already highly fulfilled and doing everything well that partly constitutes benefiting those things and people they care about, and are thus not becoming *more* fulfilled by *anything* anymore. A consequence of my account is that their lives are not made any *more* meaningful for them in virtue of their present engagements. While I am not denying that completed lives can be rather meaningful, this result still might strike some as counterintuitive. I try to ameliorate this feeling in section 4 with my discussion of the piano master. My thanks to an anonymous reviewer for remarks on this consequence.

ϕing *well* meets the above requirements. Moreover, it seems right that her ϕing should play a nondeviant explanatory role of some kind, rather than being merely causally implicated.[18] Beyond this, I doubt that any perfectly general story can be told across all *x*s by which a person can become more fulfilled; the details will matter, and particular accounts ought to give due attention to the particulars of each kind of case.

Second, I have not yet extended the becoming more fulfilled view to account for the *degree* of meaningfulness that stretches of this process might underwrite; this is important since various things can make a life meaningful for someone to varying degrees. I will develop this dimension in section 4, where the issue arises organically.

3. TWO ARGUMENTS AGAINST SUBJECTIVISM

Susan Wolf considers lives that strike her as meaningless—for example, lives entirely devoted to solving crossword puzzles, smoking pot, making handwritten copies of *War and Peace*, and rolling a rock up a hill—in which it is stipulated that the people involved have every subjective quality that could plausibly matter for whether their lives are meaningful. Because they have every such quality, the meaninglessness of their lives is not plausibly explained in terms of an absence of some such subjective feature. This would be a significant strike against subjectivism if it were true.[19]

Of course, Wolf is sensitive to the fact that others might not share her intuitions about these lives, so she tries to offer evidence that these lives *are* meaningless, evidence that is independent from her initial intuitions. She focuses her argument on one such life, trusting (as I will) that it is in all important respects the same as the others. She considers a variation (due to Richard Taylor) of the mythical Sisyphus who is just like the original Sisyphus save that the gods, in a fit of mercy, "[implant] in him a strange and irrational impulse … to roll stones," thereby "[giving] Sisyphus precisely what he wants—by making him want precisely what they inflict on him."[20] Taylor goes on to observe that "Sisyphus' fate now does not appear to him as a condemnation, but the very reverse. His one

18 *Non-deviant* causal explanations are a general philosophical issue in both the philosophy of action and of dispositions, so this requirement is not a problem for my account in particular (see Setiya, "Intention," sec. 2).

19 Wolf, *Meaning in Life and Why It Matters*, 16–18; cf. Bramble, "Consequentialism about Meaning in Life," 448; Metz, *Meaning in Life*, 175; and Smuts, "The Good Cause Account of the Meaning of Life," 541–44.

20 Taylor, *Good and Evil*, 323.

desire in life is to roll stones, and he is absolutely guaranteed its endless fulfillment."[21] Despite the fact that he now takes great pleasure in his task, he feels fulfilled by it, he views his own life not as one of condemnation but the opposite, etc., his life is not meaningful, Wolf argues, because the activity around which his life is built is futile, unproductive, and pointless:

> The reason Sisyphus has traditionally been taken as a paradigm of a meaningless existence is that he is condemned to the perpetual performance of a task that is boring, difficult, and futile. In Taylor's variation, Sisyphus's task is no longer boring—no longer boring to Sisyphus, that is. But it remains futile. There is no value to his efforts; nothing ever comes of them. Even if due to divine intervention, Sisyphus comes to enjoy and even to feel fulfilled by his activity, the pointlessness of what he is doing doesn't change.[22]

It is worth taking a moment to show why this argument, if it works, strikes against the becoming more fulfilled view. A Sisyphus who cares a great deal about rolling a rock up a hill cares about doing an activity that is plausibly of a goodness-fixing kind—there is something that it is to roll a rock up a hill *well* that is plausibly determined by what it is to roll a rock up a hill—and his doing it can change his cares in a way partly explained by his antecedently caring as he does. For example, as he rolls the rock up the hill, he might find out that taking a certain path is faster and requires less effort from him, and thus that rolling it well involves doing that; and he can thereby come to care that he does so on future laps. Provided all of this is true, the becoming more fulfilled view will say that Sisyphus's life is at least somewhat meaningful for him. So if Wolf has provided us with evidence that his life is *not* at all meaningful, the view has a problem. However, she has not: Sisyphus's activity is not pointless, unproductive, or futile.

It helps to see this by keeping an eye firmly fixed on what Sisyphus's goal actually is: to roll the stone up the hill over and over again. His goal is *not* to roll the stone up the hill and place it at the top; the gods' mercy was precisely to give him a desire to do the very thing they condemned him to do. And, it should be noted, he succeeds in his goal: he rolls the stone up the hill over and over again.[23]

21 Taylor, *Good and Evil*, 323.

22 Wolf, *Meaning in Life and Why It Matters*, 17.

23 Sisyphus's goal here is to do what Setiya has called an *atelic*, as opposed to *telic*, activity ("The Midlife Crisis"). Having a telic activity as one's goal is for one's goal to be extinguished upon successfully doing the activity (which is not to say that one cannot adopt the goal, and so do the activity, *again*); when one's goal is an atelic activity, successfully doing

So when Wolf insists his goal remains futile, we should wonder in what sense this is true; it is evidently *not* true if she means to say that he cannot enjoy any success. Nor is it true that his efforts are unproductive, which is to say that his effort produces nothing further. Indeed, perhaps she is elaborating on this point when she says, "There is no value to his efforts; nothing ever comes of them."[24] His success—his doing what he wants deeply to do—produces pleasure and feelings of fulfillment. An uncareful reading of the case can cause us to miss this point. The mercy of the gods is not that Sisyphus is injected with a kind of Feel Good Drug that gives him indiscriminate, warm feelings of pleasure and fulfillment; no, they implant in him a *desire* to live a certain kind of life. The pleasure and feelings of fulfillment follow in the wake of his acting on this desire with success, not his being in an experience-machine-type situation where his pleasure and feelings of fulfillment bear no connection to his active participation in the world.

Finally, there is the charge of pointlessness. Sisyphus's efforts evidently *do* have a point, albeit one that is internal to the activity itself. Many activities are like this: the point of going for a walk is sometimes just to go for a walk; the point of playing tag is sometimes just to play tag. We often endeavor to do these things for their own sakes, not because we hope to achieve something further, not because our activities have some further point.[25] If Wolf means that there is some further point, distinct from the activity itself, that Sisyphus's activity lacks, she is surely right. But she had better say more about why activities lacking in further point cannot make for a meaningful life. On its face, such a claim is implausible because activities that we do for their own sakes often play a part in making our lives meaningful.

So the principal task of the objectivist, vis-à-vis these sorts of cases, remains: she still needs to provide us with compelling evidence that lives like Sisyphus's are meaningless. Now, some theorists take lives like Sisyphus's to be so obviously meaningless, solely on the basis of their intuitions, that they do not take themselves to need any additional evidence.[26] I will discuss this at the end of section 4.

Shifting gears now, it has also been argued that subjectivism about meaning

---

the activity is compatible with the persistence of the goal and the activity (that is, with one's *continuing* to have the goal and to do the activity).

24 Wolf, *Meaning in Life and Why It Matters*, 17; see also Levy, "Downshifting and Meaning in Life," 178; and Martela, "Meaningfulness as Contribution," 240.

25 Of course, one might say that the point of going for a walk even in these cases is to satisfy a desire or to obtain some pleasure, but neither of those things is the intentional object of the person going for a walk in these cases. And even if it were, this exact explanation could be given of Sisyphus's efforts, too.

26 For example, Bramble, "Consequentialism about Meaning in Life," 454; Metz, *Meaning in Life*, 175; and Smuts, "The Good Cause Account of the Meaning of Life," 543–44.

implies that no one can be mistaken about the degree of meaning their own lives have; that people *can* be so mistaken; and therefore that subjectivism is false. For example, as Aaron Smuts characterizes it, subjectivist theories "hold that fulfillment or some other subjective state is what makes a life meaningful. On such views, one's life is meaningful if one finds it meaningful or, we might say, fulfilling."[27] Smuts later says that "the theory implies that no one can be wrong about how meaningful or meaningless [their life is]. . . . But George Bailey's despair [in *It's a Wonderful Life*] gives us excellent reason to reject such a view. On his dark night of the soul, George mistakenly thought that his life was meaningless."[28] Similarly, Antti Kauppinen rejects subjectivism on the grounds that "just as a food can be unhealthy for a person even if she thinks it is healthy, a life can be meaningless for someone even if she thinks it is meaningful."[29]

The problem with this argument is the first premise. Subjectivism in no way implies that a person cannot be mistaken about whether her life is meaningful for her, or how meaningful it is. Subjectivism is just the view that meaning has no objective value requirement—that it is false that only objectively valuable activities can make a person's life meaningful for her—plus the claim that what makes a person's life meaningful for her is, in part, contingent features of the person herself. This is compatible with that person being mistaken either about those features, even within herself, that make her life meaningful for her, or about the fact that it is those features that make her life meaningful for her. Still, a specific moral to draw from this argument is that a *particular version* of subjectivism—the view that what makes a person's life meaningful for her is her thinking that it is—is false. And a more general moral is that, since people can be mistaken about whether their lives are meaningful for them, the right theory of what makes a life meaningful will be built around facts that people can plausibly be mistaken about.

27 Smuts, "The Good Cause Account of the Meaning of Life," 537.

28 I have substituted the bracketed "their life is" for Smuts's original formulation, which says "the theory implies that no one can be wrong about how meaningful or meaningless *they find their life*" ("The Good Cause Account of the Meaning of Life," 544, emphasis added). I take this to be a charitable clarification. If Smuts's point is that subjectivism implies, falsely, that no one can be mistaken about how meaningful they find their own lives to be, then he would have to draw a case in which someone found their life to have a certain degree of meaning, but then possessed a mistaken, second-order opinion about *that*. This is clearly not what Smuts is trying to do.

29 Kauppinen, "Meaningfulness and Time," 356. It should be noted that Kauppinen is rejecting only one variety of subjectivism, the view he attributes to Taylor (*Good and Evil*). However, this objection is the only one Kauppinen levels against subjectivism of any variety before moving on to non-subjectivist alternatives, and so it can be reasonably read as his grounds for rejecting subjectivism as such.

To see that this constraint can be easily met, let me show how the becoming fulfilled view does so. As discussed in section 2, this view says that what makes a person's life meaningful for her is her becoming more fulfilled by some *x*, which is a process in which she aims to do activities ϕ well, where ϕing well at least partly constitutes benefiting *x*, and requires that she be more fulfilled by *x* than she presently is. A person can be mistaken about whether she is undergoing this process, not least because she can be mistaken about the extent to which her activities (done well) benefit various persons, objects, practices, etc. (George Bailey is just such an example: he is mistaken about the extent to which his actions benefit his community.) Moreover, a person can be mistaken about whether doing some activity well requires that she be more fulfilled than she presently is. For example, it is easy enough to be mistaken about how much, and in what ways, philosophy has to matter to oneself to do it well—as when, for example, a person mistakes philosophy for a glorified parlor game, or an all-consuming life project. And of course a person can be mistaken about the extent to which she is already fulfilled by the things in her life, since subtle shifts in our cares—their refocusing, straying, deepening, and so forth—sometimes happen in ways our higher-order reflection does not recognize. And, finally, even if she were not mistaken about such things, she could still be mistaken about the fact that it is this process that makes her life meaningful for her.

4. ON INTUITIONS AND ASSESSMENTS

Finally, subjectivist theories are frequently charged with producing counterintuitive results. Because subjectivism denies that meaning has a necessary objective value condition, activities wholly lacking in objective value can at least in principle make a person's life meaningful for her. Activities like counting blades of grass or eating excrement can and will do so provided whatever conditions the subjectivist theory in question places on meaning can be met for these activities.[30] The charge of counterintuitiveness comes once a particular subjectivist theory is under consideration, and a case is drawn showing how, even when the conditions that comprise that theory are met, the activity intuitively does not make a person's life meaningful.[31]

30 For these examples, see, respectively, Rawls, *A Theory of Justice*, 432; and Wielenberg, *Value and Virtue in a Godless Universe*, 22.

31 The charge is put in its most general terms by Metz, *Meaning in Life*, 175. In the literature, it is typical to bring this argument to bear against Taylor's (*Good and Evil*) theory in particular; see, for example, Smuts, "The Good Cause Account of the Meaning of Life," 543–44; and Wielenberg, *Value and Virtue in a Godless Universe*, 22.

However, there is good reason to tread lightly with the intuitions we are deploying in such arguments. Consider the following quote from Wielenberg, in which he discusses a concert pianist and the excrement eater, both of whom are imagined to be fulfilled by what they do:

> Both the pianist and the grinning excrement-eater are engaged in activity for which they have a genuine passion; each is doing what he [or she] most wants to do. Imagine these two lives, one filled with the sort of activity in which [the pianist] is engaged . . . , the second filled with the grinning excrement-eater's favorite pastime. . . . If you were offered a choice between these two lives, would you be indifferent? Would the two lives seem equally worthwhile to you? If you are like me, the answer is no.[32]

The fulfilled excrement eater is supposed to show that being fulfilled does not make a person's life meaningful. Let us assume, with Wielenberg, that worth and meaning have a relatively tight connection.[33] And let us assume that we would similarly prefer the concert pianist's life, and that hers seems more worthwhile to us than the excrement eater's. The reason to tread lightly is that it is not obvious which sorts of intuitions are being reported in assessments like these.

To see why this matters, notice that the subjectivist can just say that the intuition reported here is that the concert pianist's life is more meaningful *for us*—that is, as lived by you or me. Furthermore, she can explain naturally why we have such an intuition: you and I actually care about doing things like playing the piano and not eating excrement; so you and I, as we actually are, would be fulfilled by playing the piano but not eating excrement.[34] If that is all that is going on in our assessments of cases like these, then subjectivism comes out unscathed; the charge of counterintuitiveness does not stick.

So the objectivist needs us to have intuitions of a rather different sort. One possibility is the intuition that the pianist's life is more meaningful, *period*—that is, not as lived by any person in particular. The objectivist might insist that we are capable of evaluating the degree of meaning in lives in a way sharply discon-

32 Wielenberg, *Value and Virtue in a Godless Universe*, 22.

33 Against this, see Metz, "The Meaningful and the Worthwhile"; and Martela, "Meaningfulness as Contribution."

34 It does not help that we are being asked to imagine that we care about eating excrement. This just highlights that the situation we are assessing is a little more complicated: eating excrement while caring about doing so. That situation can still be one we are assessing from our *actual* point of view, constituted by, among other things, what we actually care about. I can perfectly well acknowledge that, were I to care about eating excrement, I would be fulfilled by it, even while maintaining that I, as I actually am, would not be fulfilled by eating excrement while caring about doing.

nected from any portfolio of cares—and thus capable of having intuitions about meaningfulness, period—and that we can draw cases about which we have just such intuitions. But it is hard to see how this insistence is not question begging. After all, the whole point of subjectivism is that the meaning in a life is inseparable from the contingent features of the person whose life it is—and, when the particular theory is a fulfillment one, those features will ineluctably involve what the person cares about.

A more promising possibility is that we can have intuitions of the following form: that the concert pianist's life is more meaningful *for her* than the excrement eater's life is *for him*.[35] Clearly, having such an intuition must rely on some background metric whereby the meaning-for-the-pianist can be measured against the meaning-for-the-excrement-eater. And this background metric might be determined at least in part by the amount of objective value of the activities in each life, as the objectivist maintains; or it might be determined without that, as the subjectivist does. However, as long as we can have these intuitions before settling the latter question, as I submit we can, then theorists can count on our being able to have intuitions of this form without begging the question. So henceforth I will put the counterintuitiveness charge(s) against subjectivism in these terms.

Start with the following. A particular subjectivist theory will be counterintuitive when it implies, for example, that the concert pianist's life is not more meaningful for her than the excrement eater's is for him.

The becoming more fulfilled view has plenty to say here. On this view, a person's life is made meaningful for her through aiming to do activities well, where doing so is beneficial and requires ending up more fulfilled. Playing piano well meets these criteria: playing piano well benefits the practice of playing piano (among other things), and, at least for a long time, required that the concert pianist end up more fulfilled by it than she was as an amateur or novice. After all, she had to learn a variety of new ways to play and to come to care about playing piano in deeper and richer ways. Accordingly, the concert pianist was becoming more fulfilled by playing piano for a long time, and doing so made her life meaningful for her. Eating excrement does not meet these criteria because there is no coherent notion of doing this activity *well*, nor anything for which doing it well at least partly constitutes a benefit.

Now, the explanation I just gave depended on the proviso that the concert pianist's activities and cares were shaped over time through her aim of playing piano well. So there is a slightly different charge of counterintuitiveness in the neighborhood: that the becoming more fulfilled view implies, falsely, that the

35 Recall that I am not using the qualifiers "for her" and "for him" to refer to how meaningful these people think or feel their lives are. See note 4 above.

concert pianist's life is more meaningful for her when her activities and cares are changed in this way than it would be if they were not. To make the charge most forceful, we could imagine two concert pianists: a journeyman who is becoming more fulfilled by playing piano in these aforementioned ways, and a master who simply continues to play piano well and to care about doing so, with all the richness, depth, and subtlety we expect of her. It is plausible that the master is no longer becoming more fulfilled by playing piano. And it might strike us as counterintuitive to say that, by continuing to play the piano, only the journeyman is making her life any more meaningful for herself.

Yet this claim does not strike me as counterintuitive at all. We can grant that the master's life is already very meaningful for her, and has been made so by the history she has with playing the piano, the excellence she has cultivated thereby, and the intimacy she has developed with it. The question is whether her continuing to play the piano *now* adds anything to that. But we stipulate that her cares never change again in response to playing, and that she does not play the piano well in new ways.[36] In light of that, her personal relationship with playing piano seems also set in stone. It would be reasonable for her to feel that spending the rest of her life doing that would amount to just more of the same, and to look on the journeyman with a bit of envy, wishing she too had such an open future within their vocation. In short, it would be reasonable for her to have a midlife crisis, one that I see no reason not to call a crisis of meaning.

A slightly different way to press the challenge to my view is by appealing to "born" rather than "learned" masters, people who do what they do extremely well, and care very deeply and richly about it, virtually from day one.[37] It might seem that, regarding such people, my view would have it that their lives are not meaningful for them at all because they never become more fulfilled by what they masterfully do. This challenge, however, presupposes a false view of actual mastery, born or otherwise. Consider, first, that certified prodigies like Mozart or the mathematician March Tian Boedihardjo learn a lot about how to do what they do well and how to care about it in deeper and richer ways (though they do so quickly). Second, and more importantly, once they become masters of the state of the art, they usually push the boundaries of their field in new directions. These are interesting cases, to be sure, but not because they cannot become more fulfilled by music or mathematics; rather, I suspect, because they change *what counts as* making music or doing mathematics well, and open up new and exciting ways to do these things. This very fact allows them to continue to become

36 These stipulations need not apply to actual masters—see the discussion a couple of paragraphs down.

37 My thanks to an anonymous reviewer for pressing this line.

more fulfilled by music or math: their own endeavors create new ways to do (well) what they care about and (typically) new ways to care about it.

These dynamics of mastery not only apply to born masters like Mozart, but to learned masters like the pianist from a couple turns back. When it comes to practices like music, art, or philosophy, what it is to do these things well, and which ways there are to do so, are changed through the practice itself—especially the masters' doing of it. And this open-ended refinement of the practice typically brings open-ended refinement of our cares in its train. Because of this, the becoming more fulfilled view does not have it that masters' lives are not made any more meaningful for them when they have such mastery—far from it. Rather, I think the becoming more fulfilled view upholds these practices as those through which an *ideally* or *maximally* meaningful life can be led, one in which becoming more fulfilled can proceed indefinitely.[38]

Let us return to the main thread and level the counterintuitiveness charge in one last way. I have emphasized that in becoming more fulfilled by some *x*, the activities through which a person does so need not be objectively good. It is therefore possible to draw cases in which one person is becoming more fulfilled through an activity that is plausibly very objectively good, and a second person is becoming more fulfilled through an activity that is not. For example, we might compare the concert pianist's life against one entirely devoted to the sophomoric drinking game flip cup, and imagine that both individuals are becoming more fulfilled through their respective activities in the way my view states: both are aiming to do what they do well, and doing so both benefits the practice in question and requires greater fulfillment on their part. It is intuitive that the concert pianist's life is more meaningful for her than the flip-cup player's life is for him. The charge is then that the becoming more fulfilled view cannot get this result.

At its root, this challenges the view's ability to recover intuitive differences in the *degree* to which various lives are meaningful for those who lead them. This is an important challenge that I can only begin to address in this paper. First, notice that playing piano is a more complex activity than playing flip cup, one that admits of greater degrees of subtlety, variation, and innovation. Indeed, greater degrees of such subtlety, variation, and innovation are plausibly *required* to play piano well than to play flip cup well. For this reason, it also plausibly requires

38 The themes broached here are discussed at some length by Neil Levy. For him, *projects* make a life meaningful, which are activities in which "the goal they pursue is not fixed prior to the activity itself. Instead, the goal is gradually defined and more precisely specified in the course of its pursuit, so that the end of the activity is always itself one of its stakes" ("Downshifting and Meaning in Life," 184–85). Levy is, however, an objectivist: for him, "supreme value" is part of what is at stake in projects ("Downshifting and Meaning in Life," 185).

deeper and richer caring to do well—the skilled piano player must attend to and be moved by a wider and more complicated array of considerations, and she must respond to these considerations intensely and sensitively.

These observations support the idea that a person can become more fulfilled through playing piano *to a greater extent than* she can through playing flip cup, but this comparison requires clarification. To make headway on this, my remarks on mastery can help. I suggested that lives devoted to mastering practices that admit of open-ended refinement can be seen as maximally meaningful lives because they permit the process of becoming more fulfilled to proceed indefinitely. Using such practices as a kind of yardstick suggests that how meaningful a life is for the person who lives it is a matter of *how long* she can become more fulfilled through the activities she does. Practices like philosophy, music, or math, which admit of open-ended refinement, allow a person to become more fulfilled indefinitely; simpler and less mutable practices like flip cup allow this to a much more limited extent. This can go some way to explain why the concert pianist's life is more meaningful for her than the flip-cup player's life is for him. And, finally, it can help explain why people who devote their lives to playing flip cup, or rolling a rock up a hill over and over again, or solving classic sudoku puzzles, are missing out (whether by choice or divine condemnation) on lives more meaningful for them than theirs.

5. Concluding Discussion

The aim of this paper has been to outline and defend a novel but modest subjectivist theory about meaning in life. On this theory, a certain way of becoming more fulfilled makes a person's life meaningful for her via a genuinely subjective source. This theory is modest because it remains agnostic as to whether there are nonsubjective sources of meaning, such as engagement with objectively valuable pursuits.

Becoming more fulfilled is, in general, a process with being more fulfilled as its hypothetical endpoint. Of particular relevance to meaning in life is becoming more fulfilled by some *x* in the following way: aiming to do a set of activities well, which at least partly constitutes benefiting *x* and requires the person to end up more fulfilled by *x* than she presently is. I motivated and spelled out this view in sections 1 and 2, contrasting the view specifically against a more standard fulfillment view that says that what makes a life meaningful for a person is her being fulfilled. The particular advantage of the becoming more fulfilled view stems from its emphasis on a particular process (of becoming more fulfilled) as opposed to a particular state (of being fulfilled); my view thereby allows for cases

of meaningful lives in which the person is not, in fact, fulfilled by that life. This is because, like any process, becoming more fulfilled can halt before it is complete, before fulfillment is achieved. I showed that this is an advantage by discussing the case of a person writing philosophy who, though it has come to leave her cold, persists in the hope of recovering her passion for it; writing philosophy makes her life meaningful for her whether or not she in fact recovers her passion.

The remainder of the paper defended the becoming more fulfilled view against charges leveled against subjectivism. Section 3 responded to two arguments, the first by Susan Wolf and the second by Antti Kauppinen and Aaron Smuts. Wolf considers a varieties of lives, dedicated to activities like rolling a rock or solving crossword puzzles, in which she stipulates that every subjective quality is in place that could plausibly matter for meaning. She then suggests that the fact that these lives are pointless, unproductive, and futile is evidence that they are meaningless nonetheless. This clearly would pose a problem for subjectivism in general, and, as I show, the becoming more fulfilled view in particular. I responded by arguing that these lives are not, *pace* Wolf, pointless, unproductive, or futile, and so she has not provided us with such evidence. Of course, such lives might strike one as so obviously meaningless, on the basis of one's own intuitions, that one feels no need to offer any such evidence; I deferred my response to this until section 4. Kauppinen and Smuts, for their part, argue that subjectivism implies, falsely, that no person can be mistaken about how meaningful their own life is. I responded by pointing out that subjectivism as such does not at all imply this. Subjectivism is the view that not only objectively valuable activities can make a person's life meaningful for her, and that contingent subjective features of her are at least part of what does. A person can very well be mistaken about how meaningful her life is on such a view.

Finally, I showed in section 4 that the becoming more fulfilled view is not counterintuitive, once we are careful about what sorts of intuitions we are expressing in our assessments of lives. As a preliminary matter, I argued that we should take care to weigh subjectivism against intuitions about how meaningful a life is *for the very person whose life it is*, as well as intuitions involving comparisons of the same form between multiple persons vis-à-vis their own lives—this is, in short, so as not to beg the question against subjectivism, or to give it too easy a way out. I then showed that the becoming more fulfilled view produces broadly intuitive results. It can say, for example, that lives devoted to activities like eating excrement, watching paint dry, etc., are entirely meaningless for the people living such lives, because those activities cannot be part of the process of becoming more fulfilled. And, even among lives devoted to activities that can be a part of this process, the view can say that some are more meaningful than oth-

ers for the people involved just when and because the activities involved allow for becoming more fulfilled to a greater extent.

The idea of becoming more fulfilled to a greater extent requires further analysis in future work, but I suggested that how masters (born or learned) engage with their craft can be illuminative. In particular, I suggested that their lives are maximally or ideally meaningful (via a subjective source, at any rate) because the process of becoming more fulfilled can proceed indefinitely. Extending this idea, I suggested that *the extent* to which a person can become more fulfilled through some activity is matter of *how long* she can become more fulfilled through it.[39]

Stanford University
wparmer@stanford.edu

REFERENCES

Adams, Frederick. "Intention and Intentional Action: The Simple View." *Mind and Language* 1, no. 4 (December 1986): 281–301.

Arpaly, Nomy. "Comment." In *Meaning in Life and Why It Matters*, by Susan Wolf, 85–91. Princeton: Princeton University Press, 2010.

Bramble, Ben. "Consequentialism about Meaning in Life." *Utilitas* 27, no. 4 (December 2015): 445–59.

Bratman, Michael E. *Intention, Plans, and Practical Reason*. Cambridge, MA: Harvard University Press, 1987.

———. *Structures of Agency: Essays*. Oxford: Oxford University Press, 2007.

Brogaard, Berit, and Barry Smith. "On Luck, Responsibility, and the Meaning of Life." *Philosophical Papers* 34, no. 3 (2005): 443–58.

Calhoun, Cheshire. *Doing Valuable Time: The Present, the Future, and Meaningful Living*. Oxford: Oxford University Press, 2018.

Darwall, Stephen. *Impartial Reason*. Ithaca, NY: Cornell University Press, 1983.

Evers, Daan, and Gerlinde Emma van Smeden. "Meaning in Life: In Defense of the Hybrid View." *Southern Journal of Philosophy* 54, no. 3 (September 2016): 355–71.

39 Earlier versions of this paper were presented to audiences at Stanford University and Humboldt Universität zu Berlin. Special thanks are due to Michael Bratman, Berit Braun, Daniele Bruno, Francesca Bunkenborg, Phillip Fox, Nadeem Hussain, Barry Maguire, Thomas Schmidt, Razvan Sofroni, David Streit, and anonymous reviewers for their very helpful feedback.

Helm, Bennett W. *Love, Friendship, and the Self: Intimacy, Identification, and the Social Nature of Persons*. Oxford: Oxford University Press, 2010.

Jaworska, Agnieszka. "Caring and Internality." *Philosophy and Phenomenological Research* 74, no. 3 (May 2007): 529–68.

Kauppinen, Antti. "Meaningfulness and Time." *Philosophy and Phenomenological Research* 84, no. 2 (March 2012): 345–77.

Kekes, John. "The Meaning of Life." *Midwest Studies in Philosophy* 24, no. 1 (2000): 17–34.

Kolodny, Niko. "Aims as Reasons." In *Reasons and Recognition: Essays on the Philosophy of T. M. Scanlon*, edited by R. Jay Wallace, Rahul Kumar, and Samuel Freeman, 43–69. Oxford: Oxford University Press, 2011.

———. "Love as Valuing a Relationship." *Philosophical Review* 112, no. 2 (2003): 135–89.

Levy, Neil. "Downshifting and Meaning in Life." *Ratio* 18, no. 2 (June 2005): 176–89.

Maguire, Barry. "Love in the Time of Consequentialism." *Noûs* 51, no. 4 (December 2017): 686–712.

Martela, Frank. "Meaningfulness as Contribution." *Southern Journal of Philosophy* 55, no. 2 (June 2017): 232–56.

McCann, Hugh J. "Di Nucci on the Simple View." *Analysis* 70, no. 1 (January 2010): 53–59.

———. "Intentional Action and Intending: Recent Empirical Studies." *Philosophical Psychology* 18, no. 6 (2005): 737–48.

———. "Rationality and the Range of Intention." *Midwest Studies in Philosophy* 10, no. 1 (September 1986): 191–211.

———. "Settled Objectives and Rational Constraints." *American Philosophical Quarterly* 28, no. 1 (January 1991): 25–36.

Mele, Alfred R. *Springs of Action: Understanding Intentional Behavior*. Oxford: Oxford University Press, 1992.

Metz, Thaddeus. "The Concept of a Meaningful Life." *American Philosophical Quarterly* 38, no. 2 (April 2001): 137–53.

———. "The Meaningful and the Worthwhile: Clarifying the Relationships." *Philosophical Forum* 43, no. 4 (Winter 2012): 435–48.

———. *Meaning in Life: An Analytic Study*. Oxford: Oxford University Press, 2013.

———. "The Meaning of Life." *Stanford Encyclopedia of Philosophy* (Summer 2013). https://plato.stanford.edu/archives/sum2013/entries/life-meaning.

Paul, Sarah. "Embarking on a Crime." In *Law and the Philosophy of Action*, edited by Enrique Villanueva, 101–24. Amsterdam: Rodopi, 2014.

Rawls, John. *A Theory of Justice*. Cambridge, MA: Harvard University Press, 1971.
Scheffler, Samuel. "Valuing." In *Reasons and Recognition: Essays on the Philosophy of T. M. Scanlon*, edited by R. Jay Wallace, Rahul Kumar, and Samuel Freeman, 23–39. Oxford: Oxford University Press, 2011.
Seidman, Jeffrey. "Valuing and Caring." *Theoria* 75, no. 4 (December 2009): 272–303.
Setiya, Kieran. "Intention." In *Stanford Encyclopedia of Philosophy* (Fall 2018). https://plato.stanford.edu/archives/fall2018/entries/intention.
———. "The Midlife Crisis." *Philosophers' Imprint* 14, no. 31 (November 2014): 1–18.
Smuts, Aaron. "The Good Cause Account of the Meaning of Life." *Southern Journal of Philosophy* 51, no. 4 (December 2013): 536–62.
———. *Welfare, Meaning, and Worth*. New York: Routledge, 2017.
Stout, Rowland, ed. *Process, Action, and Experience*. Oxford: Oxford University Press, 2018.
Sverdlik, Steven. "Consistency Among Intentions and the 'Simple View.'" *Canadian Journal of Philosophy* 26, no. 4 (December 1996): 515–22.
Taylor, Richard. *Good and Evil*. Rev. ed. Amherst, NY: Prometheus Books, 2000.
Thompson, Michael. "Anscombe's *Intention* and Practical Knowledge." In *Essays on Anscombe's* Intention, edited by Anton Ford, Jennifer Hornsby, and Frederick Stoutland, 199–210. Cambridge: MA: Harvard University Press, 2011.
———. *Life and Action: Elementary Structures of Practice and Practical Thought*. Cambridge: MA: Harvard University Press, 2008.
Thomson, Judith Jarvis. *Normativity*. Chicago: Open Court, 2008.
Wielenberg, Erik. *Value and Virtue in a Godless Universe*. Cambridge: Cambridge University Press, 2005.
Wolf, Susan. *Meaning in Life and Why It Matters*. Princeton, NJ: Princeton University Press, 2010.
———. *The Variety of Values: Essays on Morality, Meaning, and Love*. Oxford: Oxford University Press, 2015.
Wolfson, Ben. "Agential Knowledge, Action and Process." *Theoria* 78, no. 4 (December 2012): 326–57.
Wong, Wai-hung. "Meaningfulness and Identities." *Ethical Theory and Moral Practice* 11, no. 2 (April 2008): 123–48.

Journal of Ethics and Social Philosophy
VOL. 20, NO. 1 · JULY 2021
https://doi.org/10.26556/jesp.v20i1.1263

FRIENDSHIP AS A NON-RELATIVE VIRTUE

Rachel Z. Friedman

THIS ARTICLE takes its bearings from Martha Nussbaum's "Non-Relative Virtues: An Aristotelian Approach," which argues that Aristotle's ethics are grounded in fundamental and hence universal spheres of human functioning in which the various virtues are meant to represent excellent action.[1] In calling attention to this important underlying feature of Aristotle's account, Nussbaum articulates a promising approach to the virtues that can accommodate cultural variations in their expression. The underlying spheres or "grounding experiences" that she identifies allow for the possibility of comparison among different accounts of what good action entails. Once such a sphere has been identified as the proper locus of the virtue in question, those who disagree about the appropriate ways of acting can be seen, not as talking past one another, but rather as "arguing about the same thing, and advancing competing specifications of the same virtue."[2] A virtue ethics based on Aristotle's structure can thus remain grounded in human experience while also critiquing local customs in the name of "a more inclusive account of the circumstances of human life."[3]

The aim of this article is to explore what such an approach to the virtues might look like in practice. To that end, it takes up the case study of friendship, the pinnacle of Aristotle's account of the ethical life. Friendship is a somewhat problematic virtue: Aristotle says it is a virtue or "involves virtue," and there is a strong case to be made that both claims are true. Nevertheless, it is a rich subject for study because, while a constant in nearly all human life, it is also subject to significant variation across and even within cultures (*NE* 1155a4).[4] There has recently been a surge of scholarly interest in friendship and in Aristotle's account

1 Nussbaum, "Non-Relative Virtues."

2 Nussbaum, "Non-Relative Virtues," 247.

3 Nussbaum, "Non-Relative Virtues," 250.

4 All parenthetical references to Aristotle's *Nicomachean Ethics* (*NE*) are to the translation by Terence Irwin. My understanding of the *NE* has also benefitted from the translation of Robert C. Bartlett and Susan D. Collins.

in particular.[5] These analyses, however, typically do not address the question of whether Aristotle's model might be criticized in light of concerns about its cultural particularities and, if so, whether that model could be modified while remaining faithful to his basic framework. This article therefore aims, first, to show how alternative specifications of the same virtue may be put into fruitful dialogue with one another, contributing to an ongoing conversation about how best to characterize excellent action within a given sphere of human existence. Second, in so doing, it aspires to point the way toward a philosophical account of friendship that could bear fruit in contemporary life.[6]

The argument begins with what Nussbaum refers to as the "thin" definition of friendship. This is a definition that identifies the underlying sphere in which friendship constitutes excellent action. I conclude that the best candidate for this role is the sphere of association and argue that Aristotle in fact regards the best kind of friendship as representing the height of virtue within this realm. This step in the analysis isolates the basic human experience or need to which Aristotelian friendship responds and hence the role it plays in a flourishing life. With this in mind, section 2 takes up Aristotle's "thick" definition of friendship and argues that, in addition to encompassing the other virtues, complete friendship entails its own excellent action of sharing a life with someone equal and similar, and wishing her good as one would wish one's own. I expand on this point in section 3 by considering Aristotle's views on the relationship between friendship and erotic attraction. This issue is an important source of variation among accounts of friendship, in particular friendship of the closest and most exalted kind. Aristotle's views on the matter are typically principled and yet, I argue in section 4, they also reflect the limitations of his cultural surroundings. The analysis therefore concludes by suggesting a modification to Aristotle's view—an alternative "thick" definition—that remains within the original thin specification while incorporating what I believe to be most lacking in his account, namely an erotic love of specific traits one finds in a friend as a form of mature longing for the good.

In combining a careful examination of Aristotle's arguments with an evaluation of their continued relevance, this article joins a chorus of scholars who

5 Recent contributions include Hope, "Friendship, Justice, and Aristotle"; McCoy, "Friendship and Moral Failure in Aristotle's *Ethics*"; Salkever, "Taking Friendship Seriously"; and Pangle, *Aristotle and the Philosophy of Friendship*. See also Stern-Gillet, *Aristotle's Philosophy of Friendship*; O'Connor, "Two Ideals of Friendship"; and Annas, "Plato and Aristotle on Friendship and Altruism."

6 In a similar spirit, Salkever also attempts to "bring Aristotle into conversation with … modern conceptions of friendship," but he does not discuss the source of particularity that I identify here or engage at length with the modern alternative ("Taking Friendship Seriously," 70–75).

have lately emphasized that studying the history of virtue ethics has value for contemporary moral philosophy.[7] By placing Aristotle's views into conversation with our own, I hope to show that we have much to gain from engaging with alternative specifications of the virtues. In particular, this exercise can clarify the assumptions and aspirations that shape some of our own practices, and help us to better understand what is involved in choosing one account over another.

1. Situating Aristotle's Virtue: The "Thin" Definition of Friendship

To approach Aristotelian friendship as a virtue in a way faithful to Nussbaum's underlying-experience approach, we begin by looking for the basic sphere of human life in which friendship constitutes excellent action. The most obvious candidate for that sphere is association or sociability.[8] Aristotle enumerates several social virtues in book IV of the *Nicomachean Ethics*, three of which are explicitly concerned with voluntary social interactions. Friendliness, which is the nameless virtue of accepting or objecting to the right things in the right way, "would seem to be most like friendship" (*NE* 1126b21). Truthfulness, the second social virtue, involves judging how to present one's own qualities. Finally, wit or grace describes the proper bearing in general company, and as such requires discrimination, taste, and the understanding that "different people find different things hateful or pleasant" (*NE* 1128a28). All three associational virtues are directed toward everyday dealings with acquaintances, not necessarily close family and friends. Close relations clearly involve a partiality or affection that is not present in friendliness alone. They also seem to point beyond truthfulness, insofar as they rely less on how individuals present themselves than on their characters as they really are. Finally, friends often develop a unique sensibility or way of spending time that might not be appreciated in general company but that brings them great pleasure and need not be uncivilized.

The *Nicomachean Ethics* famously provides another sustained discussion of

7 See Gardiner, "Virtue Ethics, Here and Now," as well as the other contributions to the same volume.

8 In referring to the sphere of association I follow Nussbaum's enumeration of the spheres and their corresponding virtues in her "Non-Relative Virtues," 246. Although I will not rehearse Nussbaum's responses to those who object to the notion of fundamental human spheres, I agree that association with other humans survives the most serious objections to the Aristotelian approach. It is very hard to deny that all humans have a need for contact with others and even some fellow-feeling toward them (although, as Aristotle recognized, the latter is often limited to their own "kind"). It is therefore safe to assume that some broad category of association is a universal feature of life, without which we would have difficulty conceiving of ourselves as human.

what we might identify as the virtues of social interactions: this is the discussion of *philia* in books VIII and IX. Although typically rendered in English as friendship, *philia* is better translated as love or affection and is therefore broader than friendship as English speakers understand it.[9] Despite the breadth of the subject matter, however, this discussion does not rely on the three aforementioned social virtues at all. This suggests that Aristotle intends to make a distinction between the virtues of casual sociability and the virtues of closer relationships. Indeed, a few differences between the two discussions are immediately apparent. First, while the virtues described earlier are relatively self-contained, Aristotle notes from the outset of his account of *philia* that it encompasses other virtues as well. Specifically, the best kind of *philia* encompasses justice, which Aristotle calls "complete virtue in relation to another" (*NE* 1129b27).[10] Complete friendship exists only to the extent that justice does, and indeed relies on justice or something like it (*NE* 1161a10, 1161a35, 1162a30–35, 1165a28–33). In fact, Aristotle says that friendship is the end of justice and that "if people are friends, they have no need of justice" (*NE* 1155a27).[11] Friendship thus demands social excellence of a far-reaching sort while the earlier social virtues do not.

9 See Nussbaum, *The Fragility of Goodness*, 354, and Irwin's glossary entry on *philia* in his translation of the *NE*, 330. On the same subject, David Konstan argues that Greeks made a distinction between *philia* and *philos*, which he finds refers to friends in the narrower sense, but that they lacked a specific abstract term for "friendship." As a result, while one could say the equivalents of "friends" and "love" or "caring behavior," there was no specific word for the relationship between those whom we call friends (Konstan, "Greek Friendship," 74–78, 92). He provides further evidence for the narrower meaning of *philos* in his *Friendship in the Classical World*, 53.

10 Aristotle's discussion of the first social virtues appears immediately before his account of justice, while his discussion of *philia* comes after it. In calling attention to the order of Aristotle's discussion, I rely in part on Irwin's claim that the order of treatment from the beginning of the *NE* through book IX is likely Aristotle's own. See Irwin, "Introduction," xvi.

11 Aristotle says that friendship renders justice unnecessary or is the highest form of it (*NE* 1155a28–29). These two claims appear to conflict with one another: either friendship renders justice unnecessary or friendship is justice. Delba Winthrop argues that in practice friendship is meant to supplement justice while in theory it is meant to replace it. That is, while the teaching on friendship "grounds in a more satisfactory way the institutions and habits which are necessary to ensure a modicum of justice," Aristotle never denies the necessity of justice because complete friendship is so rare (Winthrop, "Aristotle and Theories of Justice," 1214). This reading sheds light on Aristotle's seeming equivocation in the passage just quoted, since it suggests that, while friendship does render justice unnecessary from a theoretical point of view, in the complicated and problematic reality of human affairs justice must persist and friendship is its highest form, realizing its aims most fully while also pointing beyond it. Danielle Allen, writing primarily about political friendship, also points out that friendship both "achieves what justice does" and "extends beyond justice," in part because only friendship fully succeeds in converting rivalry into equity (Allen, *Talking to Strangers*, 121, 136).

Furthermore, while each of the first three social virtues involves careful attention to the demands and potential reactions of others, the best kind of friendship relies on qualities that can be "found in the decent person's relation to himself" (*NE* 1166a10–11).[12] Aristotle makes this claim immediately after relaying the opinion, which he goes on to endorse, that a friend is someone who "wishes and does goods or apparent goods to his friend for the friend's own sake," just as he wishes good things for himself (*NE* 1166a3). This suggests that, whatever friendship is, it goes deeper than the ability to act well toward others in isolated situations. A good friend knows how to be good to herself, without which she cannot know how to wish that same good for another. What seems at first glance to be a prime candidate for excellence within the sphere of sociability, then, has started to look more like excellence simply. While some forms of friendship are clearly possible without complete virtue, these are defined in reference to the best kind of friendship, in which both partners are excellent all around (*NE* 1166a12). The standard for friendship, then, is not virtue in one sphere of human life but rather depends on and manifests virtue in all the spheres. These points cast doubt on the correctness of identifying friendship as excellent action in the sphere of sociability.

There is also another difficulty involved in situating friendship in this way: namely, as A. W. Price puts it, that Aristotle presents *philia* not as a genus unifying its species but as a "range of relationships held together … by reference to a single type."[13] It is well known that Aristotle enumerates three kinds of friendship in both his *Eudemian* and his *Nicomachean Ethics*: that of utility, that of pleasure, and that of virtue.[14] Yet it is not clear how exactly these three kinds of friendship relate to one another. According to Price's reading, Aristotle argues in the *Eudemian Ethics* for a focal relationship between the lesser kinds of friendship and the best or primary kind, but does not show how the former have enough in common with the latter to actually count as friendships; in the *Nicomachean Ethics*, by contrast, Aristotle offers more detail about the three kinds but is less explicit about the pattern into which they fit.[15] On the positive side, this should enable us to focus on primary or complete friendship without trying to incorporate the lesser varieties into a rigorously systematic account. At the

12 See also *NE* 1157b34–1158a2, 1168b1–10.

13 Price, *Love and Friendship in Plato and Aristotle*, 131.

14 In the *NE*, Aristotle variously refers to the last kind of friendship as "complete friendship" (1156b7), "friendship primarily and fully" (1157a31), friendship "without qualification" (1157b4), "friendship of virtue" (1158b10), and the "best type" of friendship (1158b12). I will use these terms interchangeably and also follow the practice of some commentators of calling the best kind of friendship "character friendship." See, e.g., Sherman, "Aristotle on Friendship and the Shared Life."

15 Price, *Love and Friendship*, 134–38.

same time, however, it calls into question whether we will be justified in regarding complete friendship as *the* virtue of right action in the sphere of sociability. Because it means that the best kind of friendship is related only loosely to the two more common kinds, and even more loosely to *philia* broadly construed, it may be too specific and rare a phenomenon to deserve the single definition of right action in such a vast realm of experience.

These considerations might lead one to conclude that the "grounding experiences" approach is at once too narrow and too broad to apply to Aristotelian friendship. It is too narrow because to fit complete friendship into a single sphere risks denying its need to encompass all of the other virtues as well. At the same time, it is too broad because it expands the significance of such friendship across a wider realm of human experience than it may deserve. To the first objection, we can respond that there is no reason why friendship cannot both be a specifically social virtue *and* encompass the other virtues. As a grounding experience, sociability has the unique feature of permeating all aspects of our lives: without the ability to get along with others, it is simply not possible to live life well. Moreover, without others among whom to exercise one's virtues, in particular those such as generosity and justice, there is little reason to possess them in the first place. The idea that a virtue encompassing other virtues is also a fundamentally social virtue thus reminds us of the close connection between living well and living with others—not merely by tolerating or getting what one needs from them, but by appreciating their essential role in human happiness.

The second objection is somewhat more difficult to answer, because it points out that, in identifying friendship as right action within the sphere of sociability, we are limiting the heights of virtue within a very broad area of human life to the very few. Aristotle suggests that only those who are already virtuous can enjoy complete character friendship. The implication of this claim seems to be that everyone else is excluded from social excellence. Yet it is important to recall that possession of the virtues is not an all-or-nothing affair. Of course, to define excellence in a certain way is almost by definition to restrict its achievement to a relatively small number of people. Very few, if any, will be fully virtuous. Nevertheless, Aristotle tells us that many will be capable of some form of character friendship and therefore some degree of social excellence. At the beginning of his discussion in the *NE*, he says that complete friendship is rare because truly virtuous people are few (1156b25). At the very end of his discussion, however, he notes that "the friendship of decent people is decent, and increases the more often they meet. And they seem to become still better from their activities and their mutual correction" (*NE* 1172a10–13). He thereby admits that even those who are not perfect can benefit from the salutary effects of character friendship,

something that resembles the best kind even if it does not fully attain it.[16] We look to Aristotle's model, then, not in order to portray the world in binary terms of the perfectly virtuous and everyone else, but rather to flesh out an account of the heights of human achievement, to see what it might too hastily assume or leave out, and then to shape our own practices, to the best of our ability, in light of what we have found.

Nussbaum distinguishes between the "thin" or nominal definition of a virtue, which is "whatever being disposed to choose and respond well consists in" within a particular sphere, and the "thick" definition, which is the virtue's actual specification.[17] We have now seen that producing a thin definition of the virtue of association is far from a simple task—in fact, there may not be a single thin definition at all. I therefore propose to regard the sphere of association as containing a number of phenomena of which complete friendship is one. Correspondingly, we will also have to see the virtue or virtues associated with Aristotle's complete friendship not as the only social excellence but as social excellence of a particular type, in fact the most elevated or desirable. The latter view finds support if we think of justice as the paradigmatic or most authoritative social virtue.[18] Just as one might have an intuitive grasp of proper social conduct without being just, so too we can imagine someone who is just yet who does not know how to be a good friend.[19]

Having situated complete friendship in this way, we can begin to see what purpose it serves in Aristotle's account of a happy or flourishing life. Of all of the associations Aristotle discusses, complete friendship is the most choice-worthy: it serves at once the many functions that association is thought to serve—utility, pleasure, and most of all virtue—and it does so in the most lasting way possible. It therefore not only protects friends against misfortune and error, and provides them with a pleasant way to pass the time, but it also enables the continued exercise and development of their best desires and qualities. We will thus be able to place alternative accounts of friendship into a dialogue with Aristotle's if they too concern a partnership that does not merely address vital basic needs but also calls on and manifests the heights of both friends' humanity.

16 Price makes a similar point, arguing against a more pessimistic view advanced by John Cooper. Price notes that, just as human beings exist on a spectrum from very good to very bad, so too examples of character friendship may range from total to incomplete (*Love and Friendship*, 158).

17 Nussbaum, "Non-Relative Virtues," 247.

18 See also Nussbaum, *The Fragility of Goodness*, 351–52.

19 While lacking actual friends could be the result of bad luck, lacking the ability to be a good friend is far more likely a failure of character.

2. Well-Wishing and the Shared Life: The "Thick" Definition

The question we must now address is what particular excellence one needs in order to be capable of complete friendship on Aristotle's terms. This corresponds to Nussbaum's "thick" definition, the rich specification with which any alternative account will have to grapple. Price has ably set out and defended the view that the primary or best form of Aristotelian friendship is characterized by wishing one's friend well for her own sake, which distinguishes character friendship from the lesser forms of *philia*.[20] Aristotle makes clear that one should have the same kind if not degree of concern for one's friend that one has for oneself. In addition, wishing a friend well requires more than merely having the right internal state; character friends must share in one another's activities, in particular the distinctively human activities that are constitutive of a good life.[21] Following Price's analysis in significant respects, I will argue that complete friendship requires shared activity because through it each friend is able to reflect upon herself by observing and conversing with the other.[22] This means that one's friend must partake of virtuous activity of the same sort and in the same way that one would wish for oneself, and that one will in turn desire the continued good of the friend just as one would desire one's own good. It is very important on this model that the friend be understood as an independent being capable of acting well on her own volition; otherwise, her action will reflect neither her own virtue nor, by extension, that of her friend. This vision of well-wishing depends on the excellence and relative self-sufficiency of both friends, features about which I will have more to say in the next section. What I wish to bring out here is that the ability to realize Aristotle's vision also requires the recognition that even the greatest excellence is incomplete without others with whom to share it.

Near the beginning of book VII of the *Eudemian Ethics*, Aristotle offers an initial description of the types of friendship and states several essential require-

20 The importance of friendship is, on Price's reading, largely epistemic: in joint deliberation and activity with one's friend, one comes to discover one's own self through the other. This process explains why even good people need friends, since to act virtuously requires being aware both of how an activity comes about and how it is performed. The friend's acting well in turn becomes a part of one's own happiness because one identifies with her choices. One wishes the friend to continue acting well in order to continue to know oneself and to persist in virtuous activity. For the detailed development of this argument, see Price, *Love and Friendship*, 120–24.

21 This point is also emphasized in Salkever, "Taking Friendship Seriously," esp. 68–69.

22 One difference is that I emphasize the pleasure of self-perception more than its instrumentality for virtue, but I also discuss the instrumental role of friendship, below.

ments for friendship of the best or primary kind.[23] First, such friends must mutually recognize each other's goodness of character, since it is on this recognition that their friendship is built (*EE* 1236b1–6). Second, they must feel mutual affection based on their respective goodness, not on any accidental qualities they possess (*EE* 1237a40–1237b5). Third, they must find shared purpose and agreement, which is possible because both friends aim at a conception of the good and have the constancy of character to pursue that conception over time (*EE* 1237b10–35). Fourth, a friend must not only be good absolutely but also good to another, which means actively sharing in the partnership and therefore precludes having a large number of friends (*EE* 1238a4–11). Finally, character friends are beneficial to another not with respect to any particular purpose, but rather for the unqualified good of each (*EE* 1238a30–1238b15).

These observations lead Aristotle to a discussion of equality in friendship and to the conclusion that friends of the best kind must be equal. In unequal friendships, reciprocity between the parties is lost, since a benefactor either does not return the affection of the benefited or does not return it in the same way. As a result, such friendships are prone to discord (*EE* 1239a5–20). By contrast, the aim of complete friendship is shared choice and virtuous activity, which is also pleasant by virtue of being good. This means that the friends must have the same motivations and purposes, which is not possible if what is good for one is not good for another. Moreover, activity consists more in loving than in being loved, so if friends are unequal in their affection their activity necessarily differs (*EE* 1237a35–40). Inequality therefore leads to disharmony and to a reduction or loss of the shared activity that characterizes the best kind of friendship.

Aristotle has still not explained, however, why friendship must revolve around the kind of shared ends that equality makes possible. After discussing a number of other issues, he at last comes to the question of why, if happiness is self-sufficiency and the good human being is as self-sufficient as possible, such a person will nevertheless need or want friends. What human beings most desire, he says, is perception and knowledge, and in particular they desire perception and knowledge of themselves (*EE* 1244b25–1245a11). One attains this knowledge through a friend, who is one's other self. The friend must therefore reflect one's own characteristics and her ends must be one's own, since only then will one be able to observe oneself in observing the friend. It is pleasant to share even base pleasures with friends, but it is "more so to share in the more divine pleasures;

23 All parenthetical references to Aristotle's *Eudemian Ethics* (*EE*) are to the translation by Rackham, with cross-references to that of Inwood and Woolf.

the reason of which is that it is always more pleasant to behold oneself enjoying the superior good" (*EE* 1245a35–40).[24]

Aristotle does not fully explain here why the goal of human life is self-knowledge or why such knowledge demands another person. Even if we grant that the virtuous person must know her own motivations in order to be fully good, the argument for friendship based on this claim is circular: humans need character friends in order to be virtuous, since without such friends they cannot fully know themselves, but they must be virtuous in order to have such friends. Perhaps this is why the following passages and Aristotle's discussion in the *NE* emphasize the pleasure more than the necessity of self-perception. On this account, happiness is activity, and perception or understanding is the definitive activity of human life (*NE* 1170a16–19). The activity of the good human being is excellent and pleasant in itself, so it is particularly pleasant for the good person to observe her own activity (*NE* 1169b29–33, 1170b1–6). Yet the activity of an excellent friend is just as pleasant, if not more so, since it is thought to be one's own, yet it is easier to observe another than to observe oneself (*NE* 1169b34–1170a2). The truly "blessed person" will therefore need virtuous friends in order to fully experience the pleasure of a virtuous life (*NE* 1170a3–4).[25]

Just before this passage in the *NE*, Aristotle offers another indication of the value of friendship to the already virtuous person. Here, he concludes a discussion of selfishness and self-love with the claim that a good friend must be a self-lover. Such a friend always pursues the noble or fine, including by forfeiting money or even actions for the benefit of friends, "since it may be finer to be responsible for his friend's doing the action than to do it himself" (*NE* 1169a34). In describing the self-lover in this way, Aristotle seems to recall that honor-loving paragon of virtue, the great-souled man.[26] Yet there is reason to question whether the great-

24 Cooper, in reconstructing Aristotle's argument for the importance of friendship for the good life, relies heavily on another passage to very similar effect from the *Magna Moralia* (*MM*). Yet this passage does not go much beyond the *EE* insofar as the latter also emphasizes self-knowledge. Cooper stipulates that the self-knowledge obtained through friendship is knowledge "of what actually motivates one's actions, not just of what intellectualized theory of living one is prepared to defend" ("Aristotle on Friendship," 339n23). This argument does not fully explain why the virtuous person, who should already know what motivates him, needs another person to complete his virtue. See Cooper, "Aristotle on Friendship," 320–22.

25 A concise summary of the *EE* and *NE* arguments combined (along with that of the *MM*) is found in Sherman, "Aristotle on the Shared Life," 106.

26 In book III of the *EE*, Aristotle praises greatness of soul as a virtue that relies on all the other virtues (*EE* 1232a32–38). The one who possesses it has the best possible disposition toward those goods that most human beings desire, since of all of them he cares most for honor (*EE* 1232b10–12, 1233a4–7). Aristotle's account in the *NE* is similar: because "it is only the good person who is honorable," those who "lack virtue but have [the goods of fortune]

souled individual is in fact capable of complete friendship. We cannot help but think of this character when, in book VII of the *EE*, Aristotle notes that those who are ambitious of honor tend most of all to seek friendship on the basis of superiority rather than equality (*EE* 1239a26). Some are by nature ambitious, while others are affectionate and prefer to love than be loved. The great-souled clearly belong in the first category. Of the two, however, Aristotle says that the one who loves is both more aware of the good that he possesses and more self-sufficient. He possesses the pleasure of loving simply by doing it, while "being loved is an accident, as one can be loved without knowing it" (*EE* 1239a33).[27] To the extent that he shuns friendships of equality, then, someone who is primarily concerned with honor denies himself a significant part of the activity of friendship as well as the pleasure that accompanies a shared life.

While friendship is the culmination of the virtues, then, it also depends on the recognition that even the most virtuous life is radically incomplete without it. First, without friendship even the best person cannot fully exercise her virtues, in particular those such as generosity and justice that are oriented toward another. Second, she will lack an accurate view of her own activities through which to attain the pleasurable self-understanding that, according to Aristotle, distinguishes human life from other forms. Finally, she will lack an important source of trustworthy guidance for the ongoing activity of living well.[28] Even the friendship of the best person will therefore be aimed at a good that is higher than what each partner brings individually, a good that neither can attain alone.

3. COMPLETE FRIENDSHIP AND *EROS*

My discussion has led to the conclusion that friendship for Aristotle both involves virtue and is a virtue itself: it is the excellent activity of sharing one's life with another, someone with whom one shares the perception of being worthy because similar and equal in goodness. We have thus fleshed out a good deal of Aristotle's "thick" definition of friendship. Yet it is worth highlighting a few features that will in turn provide focal points for the more critical analysis that fol-

are not justified in thinking themselves worthy of great things, and are not correctly called magnanimous; that is impossible without complete virtue" (*NE* 1124a25–29).

27 See also *NE* 1159a28.

28 Aristotle says in *NE* that we "enlist partners in deliberation on large issues when we distrust our own ability" to get the judgment right (1112b11). This could well be what he has in mind when he refers to a good person being guided by his friend (see 1125a1). This need not mean that person's judgment is otherwise faulty. Rather, as Sherman points out, the very act of choosing a friend exposes one's character and capacity for practical reason ("Aristotle on the Shared Life," 97).

lows. Specifically, I would like to dwell on the role—or lack thereof—of *eros* in Aristotle's complete friendship, since this is among the central issues that divides accounts of friendship across cultures and time periods, and even within societies.[29] Erotic relationships represent some of the most intense forms of voluntary association experienced by human beings, particularly in an era where time to associate meaningfully with those who are not lovers, life partners, or family members is limited. In fact, *philoi* who are bonded by some sort of erotic connection are today most likely to meet Aristotle's key criterion of living together, which ensures that friends will come to know one another's characters and allows them to share many of their activities.[30] If Aristotelian friendship cannot be realized within such relationships, then, we will have come up against a serious challenge to the application of Aristotle's model in the contemporary world.

Price, after surveying Aristotle's scattered remarks on the subject, concludes that within both marriage and pederasty Aristotle "envisages the emergence of that reciprocal concern and respect which constitute the best kind of friendship."[31] Even if Price is correct about this, I hope to show in what follows that to the extent that such a friendship might emerge, it would on Aristotle's account require a fundamental alteration of the nature of the original relationship. The reason for this is that Aristotle is reluctant if not altogether unwilling to regard intense desire and the perceived need to possess what one lacks as positive elements of friendship. His complete friendship is in fact best realized among those who feel no burning need at all—who are, rather, content with their characters and consequently able to accept the separate existence of their friends.[32]

In one of Aristotle's more direct statements about *eros* and friendship, found in the *EE*, he says that while *eros* seems to resemble *philia* because lovers desire to live together, they do so "not in the most proper way but in a sensuous manner" (*EE* 1245a25–26).[33] Aristotle has just reported what he goes on to call the "facts

29 For instance, it is one of the questions distinguishing Aristotle's account from Plato's. See Annas, "Plato and Aristotle on Friendship and Altruism," 536–37. For a critique and modification of Plato's view, see Soble, "The Coherence of Love."

30 For a discussion of his requirement, see Nussbaum, *The Fragility of Goodness*, 358–59. Nussbaum suggests that had Aristotle not believed in the inferiority of women "he would very likely have preferred this sharing [of activity among friends] to extend to the sphere of the household as well [as the *polis*]: thus an even more perfect *philia* would be a good marriage, in which the full range of the aspirations and concerns that make up a human life might be accommodated" (*The Fragility of Goodness*, 358). I will have more to say on this subject below.

31 Price, *Love and Friendship*, 249.

32 See also Nussbaum, *The Fragility of Goodness*, 356.

33 Price notes the lexical and syntactic ambiguity of this statement, which can mean either that the lover, unlike the friend, desires not to benefit the beloved but to enjoy seeing him,

of experience" about sociability, namely, that all people find it pleasant "to share good things with our friends, as far as these fall to each, and the best that each can" (*EE* 1245a21–22). Most likely, then, Aristotle's criticism of the partnership of lovers is that what they desire to share in their lives together is not the highest possible good but rather the sensuous good of physical contact. As a consequence, such friends will not attain what Aristotle immediately goes on to identify as the greatest benefit of friendship. While it may be the case that sensual pleasure is the most some people are capable of sharing, the implication of this discussion is that such people nevertheless do not attain complete friendship, and that those who stop here despite being capable of sharing more are not fully realizing their potential.

It may be the case that for Aristotle two friends who choose to pursue physical pleasure in addition to sharing a life of the finest goods can do so without compromising the nature of their friendship. Yet there is another aspect to Aristotelian love, aside from the appetitive or pleasure-seeking side—namely, erotic emotion that adores its object. We must therefore ask about the case of a friendship in which part of the highest good that both friends share is the erotic adoration of what the other possesses, perceived as something that each individually lacks. It is not clear that Aristotle regards the latter as a legitimate part of complete friendship. In fact, there is reason to think that erotic longing of this sort hinders rather than furthers such friendship. Aristotle notes in the *EE* that opposites desire one another because they desire the mean to which their opposite brings them (*EE* 1239b24–1240a4). In this sense, such love of the opposite is love of the good, since each recognizes what she lacks and strives through the other to obtain it. Yet unlike the virtuous person, who desires what is simply good, those who love their opposite only want what "takes them outside their natural state" (*EE* 1239a39). As if they were too hot or too cold, they look to the other to relieve them from the discomfort of the extreme.[34] This passage captures well the sense of neediness and even pain that lovers feel in the absence of their beloved. Indeed, in the *NE*, Aristotle likens erotic passion to an excess, and in the *Politics* he likens Socrates' proposal that the city be made one to Aristophanes' account in Plato's *Symposium* of lovers who "from an excess of affection" desire so much to

or that the lover's approach is dictated not by choice but by physical desire. Price, *Love and Friendship*, 240–41.

34 Aristotle relates a similar view in the *NE*, particularly in discussing Euripedes' claim that when the earth is dry "it longs passionately for rain," and the heaven in turn when filled with water "longs passionately to fall into the earth" (*NE* 1155b2–5). Aristotle goes on to identify erotic friendships among the young with friendships of pleasure and to note that it is their passion that makes their friendships so fleeting (*NE* 1156b1–4). While Aristotle perhaps intends to refer here to mere sexual appetite, this comment is consistent with other remarks in *NE* about the instability of erotic relationships (see *NE* 1157a8–10).

grow close that they fuse together, causing one or both to disappear (*NE* 1158a13; *Politics* 1262b10–15).[35] This last passage distinguishes between the erotic attachment of the lovers and the fellow feeling that holds a city together. In the case of the lovers, excess desire for the other causes each to forget the good; in the case of the city, the danger is not an excess of desire but the denial of affection altogether. Despite their differences, however, both cases suggest that if humans are to achieve healthy relationships, they will have to respect the fact that those they hold dear have separate existences. Such recognition allows each partner to pursue her own good rather than static unity with the other, and to perceive the friend as precious because subject to various forces beyond her own control.[36]

This concern about the denial of separateness may also be why Aristotle says in the *NE* that an "extreme degree of friendship" resembles friendship to oneself—that is, it takes the other as an extension of one's own will, subject to one's control (*NE* 1166b1–2). The danger of *eros* for Aristotle seems to be that, in its overpowering need, it threatens to spur the lover to seek an unrealistic command over fortune and to substitute concern for the beloved's true good with concern for his own needs or desires.[37] Aristotle does suggest that erotic friendships may become character friendship: in the *NE*, he notes that while pederasty involves contrary pleasures, many such lovers "remain friends if they have similar characters and come to be fond of each other's characters from being accustomed to them." Yet he makes it clear that this shift takes place "when the beloved's bloom is fading," and hence when *eros* gives way to affection rooted in time spent together and the mutual appreciation of character (*NE* 1157a8–12).[38]

If we return now to some of the key features of Aristotle's complete friendship,

35 References to Aristotle's *Politics* are primarily to the translation by Lord, with cross-reference made to that of Jowett.

36 Aristotle concludes this passage of *Politics* with the remark that two things above all make human beings cherish and feel affection: that something is their own and that it is precious. This passage is intended as a criticism of a city that denies individuals their separate kin and possessions. Yet we can also interpret the second requirement as referring to the necessary separateness and hence fragility of the lives of those for whom we feel affection. We are more apt to consider someone or something as precious if we recognize the degree to which forces outside of our control can take it away from us. I am indebted in my thinking on this last point to the discussion of Aristotelian friendship in Nussbaum, *The Fragility of Goodness*, 354–69.

37 Nussbaum notes that the accounts of unreformed love in Plato's *Symposium* reflect a general cultural view of *eros* as tied not with mutuality but rather with a longing for the possession of something "seen as valuable and urgently needed" (*Upheavals of Thought*, 483).

38 Aristotle's view seems to refer to pederasty as it was commonly practiced at his time. On the inequality of the relationship and the different emotions of the two partners, including in the act of intercourse itself, see Dover, *Greek Homosexuality*, 52, 84, and 91.

we can see clearly that the neediness and longing of *eros* do not play noticeable roles. First, complete friends should be equal and similar, drawn together not by desire but by the appreciation of one another's goodness. Second, Aristotle assumes character friends will be more or less completely virtuous. This means that they do not feel a *need* for one other so much as they find that the other makes life more pleasant. It is not difficult to see why character friendship works best among those who are each excellent in this way. Within a friendship of virtue, I will not seek to control the actions of my friend, since in doing so I deprive her of her own virtue and therefore of the reflection she provides of mine. In order to resist this temptation, however, I will have to feel confident that she can and will act in the way that I would wish for her, and that my wishes for her are in fact aligned with her good. In other words, I have to genuinely be able to wish for her to live in a way that is good *for her*, and to realize that my own benefit is a derivative thereof. A friendship that is motivated by the sense of privation will likely never reach this mature state, since the one who desires will continue to seek through his friend the direct fulfillment of his own needs.

Aristotle's assumption of goodness and concomitant self-sufficiency does not, as I have already pointed out, exclude the possibility of character friendship among the decent but less than excellent. Aristotle even suggests at one point that character friends might complement one another, since each may possess different qualities and "it is difficult for all to be realized in the case of one person" (*EE* 1245a30–31).[39] Nancy Sherman takes this statement to support an argument for the importance of complementarity and emulation in Aristotelian friendship. Although neither friend may be perfectly virtuous, each will be inspired to develop his character more fully as he sees admirable qualities in the other.[40]

I have suggested that this view does not completely reflect Aristotle's understanding of character friendship, which denotes the rare excellence of the fully virtuous and not such (more accessible) striving.[41] It is also noteworthy that Aristotle devotes little attention to the particular motivations and challenges of character friendship among those who have not attained the standard or may not be fully capable of doing so. This means that he declines to address a significant part of the experience of friendship for those of us who seek close friends not only so that we may perceive ourselves in someone similar, but also for the challenge and self-correction that can accompany a close encounter with ways of thought and being that are significantly different from our own. In other words, Aristotle does not fully account for the intense experiences of desire, self-criti-

39 Cf. *NE* 1172a10–14.

40 Sherman, "Aristotle on the Shared Life," 105.

41 See *NE* 1156b25.

cism, and striving that can come from the development and long-term continuation of a character friendship among those similar yet also deeply different, who share enough to understand and love one another yet whose distinct virtues and takes on the world exist in a perpetual and—ideally—productive tension. One kind of friendship that is likely to exhibit these features is that between long-term romantic partners or spouses.

4. Modern "Companionate" Marriage: A Modification of Aristotelian Friendship?

We have just seen that there is a tension for Aristotle between *eros* and the best kind of friendship. While erotic love may indeed turn into character friendship, thanks to the familiarity that time spent together brings, by then it will have transformed into a different kind of association, one in which the friends no longer feel they intensely need one another. If this conclusion is correct, it will have implications for Aristotle's positions concerning pederasty on one hand and companionate marriage or long-term erotic partnership on the other.

It is quite possible that Aristotle's apparent rejection of pederasty—which is not the same as a rejection of all homosexual *eros*—stems from his views about the role of desire in friendship rather than from the influence of entrenched sexual prejudice.[42] Yet these same views, as principled as they may be, might have differed had Aristotle lived in a society where women were able to more fully develop their potential. That is, he might have found a more positive role for *eros* in friendship had he encountered women who could be the objects not only of male physical desire but of ethical and intellectual desire as well. In that case he may have had to grapple more deeply with the experiences just described, name-

42 Nussbaum has argued that Aristotle's apparent lack of interest in the eroticism of homosexuality—demonstrated by his decision not to include it in his survey of opinions about *philia*—is an injustice both to his method and to his friend Plato, whose lifestyle he might have seriously considered as a way of pursuing the good. That even such a fair-minded man could commit such an oversight is, Nussbaum submits, reflective of the "tremendous power of sexual convention and sexual prejudice in shaping a view of the world" (*The Fragility of Goodness*, 371). While Nussbaum is clearly right that Aristotle's own method requires the careful consideration of human alternatives, it is not obvious that we must read his reticence on the subject of homosexual *eros* as a prejudiced rejection of Plato or lack of interest in his way of life. Aristotle does not affirmatively exclude sexual relations from the best friendship, and given that such friendship would on his view have to exist between two men, we may surmise that he at least contemplated the possibility that their shared activity would include sexual relations. To the extent that Aristotle does diverge from Plato's views on *eros*, I have tried to suggest that the divergence is principled, though by no means immune to critique.

ly, of appreciating the goodness in another who is equal in moral capacity and yet distinct in potentially profound ways.[43]

Aristotle does allow in the *NE* that the friendship between husband and wife may be a friendship of virtue, if both partners are decent. Children, who constitute a common good between them, may also help to render their partnership lasting (*NE* 1162a25–29). Yet Aristotle also makes clear that husbands should rule their wives in a political or aristocratic fashion, because men are naturally more suited to ruling, and that the nature of virtue differs for men and for women (*Politics* 1259a40–1259b9, 1260a1–10). He also suggests that the common good of raising children is not in fact equally shared, since women tend to feel they have labored more for their children and therefore to have more affection for them than men (*EE* 1241b4–9). A further limitation of marital friendship as compared with complete friendship for Aristotle is that Athenian men participated together in political and philosophical activities completely off-limits to their wives, activities that Aristotle was not alone in considering more reflective of virtue than housekeeping and childrearing. His views about the partnership between husband and wife were necessarily shaped, then, by his way of life and in particular by the externally imposed limitations of the women he saw around him.[44]

My point here is not simply that Aristotle might have included marital friendship as a species of primary friendship had he held different views about women. Rather, I wish to go one step further and suggest that had Aristotle been able to better appreciate women's intellectual and moral capabilities, he might have actually modified his views about primary friendship so as to allow a greater role for the erotic emotion of desiring what one lacks as a way of desiring the good. This need not mean adopting a view of *eros* as an inarticulate desire for fusion of the kind described by Aristophanes. Aristotle's sensitive concern for the independent good of the beloved would rightly rule out such an account. Nor would it mean denying the fundamental importance for complete friendship of a shared conception of the good. It would, however, mandate addressing the key features and dangers of this kind of relationship in a way that Aristotle does not.

43 It is not my intention to exclude from this characterization the friendship of long-term lovers of the same sex. Rather, I have chosen to focus on heterosexual *eros* as a prominent but nonexclusive example of a type of longing for a person who is different in significant respects, and which therefore offers a meaningful challenge to the Aristotelian view.

44 Because Aristotle's philosophical method is built on critical examination of traditional opinions, one should not be too quick to accuse him of unthinkingly accepting Athenian prejudices in this or any other regard. Nevertheless, his failure to seriously engage with the kind of long-term erotic relationship I have described does provide evidence that his understanding of friendship was limited by the possibilities for human life, and in particular women's development, revealed to him by his surroundings.

Before I try to identify and address some of those issues here, it is worth noting that Aristotle's theory is more disposed than many others in the history of Western thought to endorse the type of mature erotic relationship I have described. He already accepts some of the distinctive qualities of erotic love, such as particular concern for one individual over others, the relative eclipse of generalized social concern by exclusive attachments, and the independence and mortality of the friend, whose vulnerability represents a potentially destabilizing factor in one's own life.[45] It is therefore plausible that Aristotle would have been willing to incorporate a mature eroticism into his theory had he been able to imagine a relationship of equality and shared activity between spouses of different genders.

In order to situate such an alternative "thick" account within the same conversation as Aristotle's, however, we must first determine that this vision is as complete and choice-worthy as Aristotle understands character friendship to be. We must therefore address several dangers to which such a mature erotic relationship could, from an Aristotelian perspective, be prone. If the following criticisms are apt, we may not have found in long-term erotic partnerships a "thick" model comparable to Aristotle's character friendship.

The central danger for long-term erotic partnership from an Aristotelian perspective is that it will render the friends self-satisfied, complacent, and insular. First, in partnering with someone who possesses what one lacks, one may no longer feel a desire to cultivate those qualities for oneself. Rather than spurring self-correction and development, then, such a relationship could actually hinder those things, allowing friends to become mutually dependent. On a practical level, one sees this tendency at work in the entrenched division of household labor among some spouses, which not only discourages each from developing abilities possessed by the other, but also represents a significant amount of time that they spend in separate rather than shared pleasurable activity. Such an observation might even lead us to conclude that the spousal relationship is more an answer to our physical needs than a means to cultivate our ethical and intellectual capacities.

This concern could be addressed, however, by focusing on the partners' expectations, including and in particular their preconceived roles. If one partner believes that the most he can accomplish is to serve as a complement to his spouse, he will likely rest content with partial virtue and dependence. Yet if he enters into the friendship with an understanding of both his own strengths *and* of his potential for improvement, he will be more likely to continue pursuing a complete account of virtue, aided by the example and encouragement of his

45 For these criteria, see Nussbaum, *Upheavals of Thought*, 461, 470.

partner. He will also be more likely to seek out the kind of shared activity that only two equals can enjoy.[46]

There is, however, another psychological temptation that may be more difficult to combat. Partners who believe that they have discovered the key to fulfillment in their complementarity may devote themselves excessively to one another and to the small world they have created, to the exclusion of engagement with their wider surroundings.[47] Indeed, they might accept one another so completely that they lose the ability to reflect critically on themselves and—insofar as reflecting on one's own choices is a prerequisite for reflection in general—on the human condition. Stephen Salkever argues that the danger of erotic love on Aristotle's account is precisely that it is "unconditional, a quality that blinds us to the inevitable imperfections of composite beings such as we." Such love overwhelms our capacity for reasoned judgment and makes us forget the inherent limitations of human beings, a crucial spur to the shared activity of reflection.[48]

While I do not believe that this concern vitiates the model I am proposing, it does reveal a genuine tension between the wish for security and the drive for continued ethical and intellectual development. Robert Solomon points out that the embrace of another's particularities promotes "long-term care and passionate benevolence" toward that person, an important feature of mature erotic and spousal friendships.[49] One would not wish to deny the power, and indeed the beauty, of such benevolence. Yet unconditional acceptance is at odds with the longing that, I have suggested, justifies considering this form of friendship as a model in the first place. If what one seeks in such unions is the productive

46 The question of how spouses can be said to share activity when they may have different occupations and life goals is an important one that I do not have space to fully explore. Briefly, I think it is important to distinguish between the highest good of a partnership and incidental goods that the partners may or may not share. As long as we do this, we can allow that partners will have very different occupations, provided they are able to take part together in what they understand as the most important or desirable activity of their lives. For some, this will be raising a child. (On this subject, in fact, Aristotle unwittingly provides an argument for the equal sharing of childrearing activities.) For others, it will be political or charitable engagement, artistic creation, or philosophic conversation. Obviously, this assumes a certain amount of free time to spend together, which may be unavailable or less than optimally available for families in which both spouses work multiple jobs outside the home.

47 Such a partnership would become static and stale, like Aristophanes' lovers welded together by Hephaestus. Nussbaum provides a vivid description of this eventuality: "Wrapped in each other's arms, there they lie, for the rest of their lives and on into death, welded into one, immobile" (*The Fragility of Goodness*, 176).

48 Salkever, "Taking Friendship Seriously," 73.

49 Solomon, "Erotic Love as a Moral Virtue," 99.

tension of critical self-reflection and growth, then one may well worry that unquestioned loyalty will hinder rather than promote those ends.

We may be able to address this concern, in part, by modifying our expectations of unconditionality in long-term erotic unions. The sense of being loved as unique, even irreplaceable individuals seems to be an important aspect of what we seek out in such relationships. As Alan Soble points out, however, it is possible to understand erotic love as directed toward the beloved's good qualities without denying that one loves a particular person. In fact, to the extent that such qualities are constitutive of the beloved's identity, there may be no meaningful difference between loving the qualities and loving her.[50] There is also nothing inherently more stable about love for an individual's quirks or particularities than for her virtues. As Soble puts it, "no matter what love is based on, love can be only as constant as its basis."[51] Unconditionality may therefore not be as necessary to the fulfillment of our hopes as it appears at first blush.

In addition, focusing on the process of reciprocal improvement that, in the best case, characterizes such partnerships should alleviate the worry that someone who is loved primarily for his virtues is fungible with others who possess the same traits. Shared deliberation and mutual correction can generate a special kind of affection, grounded in the feeling that one has played a role in the other's development.[52] Erotic partners may therefore grow more attached to one another through the perception that they are responsible for each other's characters, as well as through appreciation of the other's contributions to their own self-understanding.[53]

Above all, perhaps, we should remain clear-eyed about our motivations for entering into such friendships, and in particular about the possibility of competing desires. We may certainly cultivate affection for the traits that render our partner unique, particularly insofar as this helps to sustain our care and commitment over the long term. Yet we should distinguish, within ourselves, between the dispositions that promote compassion and acceptance and those that reflect our desire for critical engagement and growth. We should also prepare ourselves for the possibility that the two may not always coexist in harmony.

50 Soble, "The Coherence of Love," 299.

51 Soble, "The Coherence of Love," 306.

52 See also Milgram, "Aristotle on Making Other Selves." While I have not presented this argument as central to Aristotle's account of character friendship, I do think that it helps to alleviate the concern that the type of erotic relationship I am describing is less secure than one that prioritizes unconditionality.

53 Simon Keller makes a related point in emphasizing the dynamic character of romantic relationships and the fact that long-term partners take on new values and goals through their experiences together ("How Do I Love Thee?" 170).

Finally, one might worry that introducing committed partnership or marriage into our account of friendship would threaten the Aristotelian vision of voluntary association, locking people into a union that may not serve the good of each individually. Perhaps one was initially mistaken about the virtues one most needed to cultivate, or selected a spouse based on an incorrect idea of the good. Perhaps one's partner has taken a turn for the worse, giving in to extreme tendencies in her personality or manifesting vices that were not evident before the friendship began. Marina Berzins McCoy notes that Aristotle recognizes a duty to assist a friend whose character has become compromised, at least if such vice is not incurable. Yet, as she also acknowledges, Aristotle accepts that "it might be necessary for the sake of one's own flourishing to end a relationship with another who threatens it—even a former friend."[54] The ideal of loyalty that is part of our view of companionate marriage or committed erotic partnership would seem to preclude dissolving the friendship even in the face of such a threat—unless it was truly severe and long lasting, by which time the loyal friend may already have been harmed in some way.[55]

At the same time, committed partnership serves a purpose that many contemporary writers on friendship have found to be important. For example, Amélie Rorty argues that the modern view of life as one fashioned by ourselves alone leads to a strong sense of vulnerability, which we look to the continuous devotion of a friend to mitigate. Committed partnership is more likely than casual friendship or even Aristotelian character friendship to meet this psychological need, particularly given our fear that we may lose our best traits. Such friendship provides the modern individual—vulnerable, poorly understood, and in need of sympathetic acceptance—the lasting companionship of someone who steadfastly supports her.[56] Given the fast pace of contemporary life and the high degree of geographic mobility, marriage or its behavioral equivalent may be our best hope

54 McCoy, "Friendship and Moral Failure," 150. O'Connor also notes that the modern ideal of intimacy could reinforce a friendship aimed at low ends, and in this sense have a negative influence on both friends ("Two Ideals of Friendship," 120).

55 Pangle notes that, on Aristotle's view, it is unlikely that friends who start out as virtuous will severely decline, since "virtue is by nature a stable thing, resting not only on deeply ingrained habits but also on clear and unshakable insights" (*Aristotle and the Philosophy of Friendship*, 138). This prospect appears more likely on the view I have proposed, however, according to which friends seek one another out based on the perception that they are not yet completely virtuous and the perceived need to learn from one other.

56 Rorty, "The Historicity of Psychological Attitudes," 80–83. O'Connor also stresses the "relief of our loneliness and support of our sense of self-worth" provided by modern intimate relationships, in contrast to the Aristotelian model ("Two Ideals of Friendship," 111).

of approximating the sustained voluntary life sharing that Aristotle had in mind, enabling us to strive for the good while providing the security we need.

5. Conclusion: Friendship and Dialogue about the Virtues

This article has set out to illustrate how, using Nussbaum's underlying-experience approach, a cross-cultural dialogue about the virtues might take place. Admittedly, my own critique of Aristotle's account is limited, and it is in some sense carried out from within his cultural and intellectual tradition. Nevertheless, it demonstrates a promising approach that may be applied more broadly in the future. First, by specifying the universal sphere of experience underlying the virtue of friendship, we were able to identify its function within that sphere and its role in a flourishing life more broadly. We were then able to consider Aristotle's full account and to identify one area in which his own cultural circumstances shaped and arguably limited his conception. Although proposing to modify Aristotle's thick definition, I preserved the thin account of friendship as fulfilling not merely basic needs but also a deeper longing to realize the best in ourselves by sharing our life with another.

I have suggested that Aristotle's reluctance to incorporate mature erotic longing into the best kind of friendship is attributable, at least in part, to the cultural particularities of women's role in ancient Athens. In proposing an alternative model based on a different understanding of women's abilities and social roles, however, I have opened myself up to the same charge of cultural particularism. It would appear, in fact, that the friendship I am proposing is based on a snapshot of contemporary liberal democratic life, or at least a certain sliver of it. In particular, it assumes equal opportunities for men and women in education, moral development, and the pursuit of fulfilling activities. It also assumes that individuals will have the resources and wherewithal to enter into their partnership voluntarily and out of love for the other's traits, not under pressure from external forces such as familial expectations, financial hardships, or other forms of dependency. As a result of these features, among others, the ideal of complete friendship among spouses is not one that is widely shared across cultures.[57]

57 In his sweeping study of English marriage from 1300 to 1840, Alan Macfarlane relies on anthropological research to argue that the Western ideal of "companionate marriage" is unusual. Elsewhere, he claims, the worlds of men and women are more or less separate, with women sharing their time and interests with female kin and neighbors and men doing the same with male associates. The Malthusian marriage system in England that Macfarlane takes as his subject was based on the ideal of a deep bond between husband and wife, yet as late as the eighteenth century it was considered unusual even in nearby France. See Macfarlane, *Marriage and Love in England*, 154–56.

To some extent, the particularity of my own "thick" account is unavoidable. Yet what I am suggesting does go against the grain of some contemporary thinking on friendship in that it prioritizes not self-disclosure or emotional support, as important as these are, but rather shared activity for the exercise and further development of the virtues.[58] Deeper than that, my account also assumes that the traits to which we are primarily attracted in a partner—those that justify the partnership, if not holding it together through time—are identifiable and repeatable rather than wholly unique to that individual. In this regard, what I have proposed remains at least in part Aristotelian despite its otherwise modern assumptions.[59] Finally, my account is broadly Aristotelian in that it has tried to remain true to his primary specification of friendship as excellent activity within a partnership that most fully expresses each friend's character and potential—in other words, to the idea of friendship as a shared pursuit of the good.

Although my account and Aristotle's are close cousins, they nevertheless challenge one another in valuable ways. For example, it is possible that our hopes for long-term erotic partnerships are in tension, insofar as we choose our partners for their repeatable virtues but seek to be loved for our own particularities. Confronting Aristotle's alternative can open our eyes to such difficulties. It also throws into relief some of our understandable and legitimate reasons for seeking out this combination of features in our closest friendships.

Of course, the alternative specification of friendship that I have offered here can and should be challenged by others wishing to articulate different ways of approaching the same underlying experience. Perhaps some will feel that my account is still too Aristotelian, others that it is too exclusive or individualistic. Regardless, I hope to have shown that the analytic framework at work here has the potential to generate important and productive debates, and that the thoughts in this article will ultimately contribute thereto.[60]

Tel Aviv University
rachelf3@tauex.tau.ac.il

58 The Aristotelian "mirror" view and the modern "secrets" view are succinctly contrasted in Pahl, *On Friendship*, 80–87; and critiqued in Cocking and Kennett, "Friendship and the Self." See also Konstan, *Friendship in the Classical World*, 14–18, which argues that, while frankness and candor were valued in antiquity, in particular because they distinguished true friends from mere flatterers, the ancients did not value the revelation of intimate details about one's life as part of friendship.

59 On this point, compare Langer, *Perfect Friendship*, 114, with Nussbaum, *The Fragility of Goodness*, 306.

60 The author wishes to thank Martha Nussbaum, Anna Schmidt, and two anonymous peer

REFERENCES

Allen, Danielle S. *Talking to Strangers: Anxieties of Citizenship since Brown v. Board of Education.* Chicago: University of Chicago Press, 2004.

Annas, Julia. "Plato and Aristotle on Friendship and Altruism." *Mind* 86, no. 344 (October 1977): 532–54.

Aristotle. *Eudemian Ethics*. Translated by Brad Inwood and Raphael Woolf. Cambridge: Cambridge University Press, 2013.

———. *Eudemian Ethics*. Translated by H. Rackham. Cambridge, MA: Loeb Classical Library of Harvard University Press, 1935.

———. *Nicomachean Ethics*. Translated by Robert C. Bartlett and Susan D. Collins. Chicago: University of Chicago Press, 2011.

———. *Nicomachean Ethics*. Translated by Terence Irwin. Indianapolis: Hackett, 1999.

———. *Nicomachean Ethics* Translated by H. Rackham. Cambridge, MA: Loeb Classical Library of Harvard University Press, 1926.

———. *Politics*. Translated by Benjamin Jowett. Princeton: Princeton University Press, 1984.

———. *Politics*. Translated by Carnes Lord. Chicago: University of Chicago Press, 1984.

———. *Politics*. Translated by H. Rackham. Cambridge, MA: Loeb Classical Library of Harvard University Press, 1932.

Cocking, Dean, and Jeanette Kennett. "Friendship and the Self." *Ethics* 108, no. 3 (April 1998): 502–27.

Cooper, John M. "Aristotle on Friendship." In *Essays on Aristotle's Ethics*, edited by Amélie Oksenberg Rorty, 301–40. Berkeley, CA: University of California Press, 1980.

Dover, K.J. *Greek Homosexuality*. Cambridge, MA: Harvard University Press, 1978.

Gardiner, Stephen M. "Virtue Ethics, Here and Now." In *Virtue Ethics, Old and New*, edited by Stephen M. Gardiner, 1–7. Ithaca, NY: Cornell University Press, 2018.

Hope, Simon. "Friendship, Justice, and Aristotle: Some Reasons to Be Skeptical." *Res Publica* 19, no. 1 (2013): 37–51.

Irwin, Terence, trans. "Introduction." *Nicomachean Ethics*, xiii–xxviii. Indianapolis: Hackett, 1999.

reviewers for their valuable input on earlier drafts of this article, as well as the Edmond J. Safra Center for Ethics at Tel Aviv University for its generous financial support.

Keller, Simon. "How Do I Love Thee? Let Me Count the Properties." *American Philosophical Quarterly* 37, no. 2 (April 2000): 163–73.

Konstan, David. "Before Jealousy." In *Envy, Spite, and Jealousy: The Rivalrous Emotions in Ancient Greece*, edited by David Konstan and N. Keith Rutter. Edinburgh: Edinburgh University Press, 2003.

———. "Greek Friendship." *American Journal of Philology* 117, no. 1 (1996): 71–94.

Langer, Ullrich. *Perfect Friendship: Studies in Literature and Moral Philosophy from Boccaccio to Corneille*. Geneva: Librairie Droz, 1994.

Macfarlane, Alan. *Marriage and Love in England: Modes of Reproduction 1300–1840*. Oxford: Basil Blackwell, 1986.

McCoy, Marina Berzins. *Wounded Heroes: Vulnerability as a Virtue in Ancient Greek Literature and Philosophy*. Oxford: Oxford University Press, 2013.

Milgram, Elijah. "Aristotle on Making Other Selves." *Canadian Journal of Philosophy* 17, no. 2 (June 1987): 361–76.

Nussbaum, Martha C. *The Fragility of Goodness*. Cambridge: Cambridge University Press, 1986.

———. "Non-Relative Virtues: An Aristotelian Approach." In *The Quality of Life, Vol. II: Traditions, Relativism, and Objectivity*, edited by Martha Nussbaum and Amartya Sen, 242–69. Oxford: Oxford University Press, 1993.

———. *Upheavals of Thought: The Intelligence of Emotions*. Cambridge: Cambridge University Press, 2001.

O'Connor, David K. "Two Ideals of Friendship." *History of Philosophy Quarterly* 7, no. 2 (April 1990): 109–22.

Pahl, Ray. *On Friendship.* Cambridge: Polity Press, 2000.

Pangle, Lorraine Smith. *Aristotle and the Philosophy of Friendship.* Cambridge: Cambridge University Press, 2002.

Price, A. W. *Love and Friendship in Plato and Aristotle*. Oxford: Clarendon Press, 1989.

Rorty, Amélie O. "The Historicity of Psychological Attitudes: Love Is Not Love Which Alters Not When It Alteration Finds." In *Friendship: A Philosophical Reader*, edited by Neera Kapur Badhwar, 73–88. Ithaca, NY: Cornell University Press, 1993.

Salkever, Stephen. "Taking Friendship Seriously: Aristotle on the Place(s) of *Philia* in Human Life." In *Friendship and Politics: Essays in Political Thought*, edited by John Von Heyking and Richard Avramenko, 53–83. Notre Dame, IN: University of Notre Dame Press, 2008.

Sherman, Nancy. "Aristotle on Friendship and the Shared Life." *Philosophy and Phenomenological Research* 47, no. 4 (June 1987): 589–613.

———. "Aristotle on the Shared Life." In *Friendship: A Philosophical Reader*, ed-

ited by Neera Kapur Badhwar, 91–107. Ithaca, NY: Cornell University Press, 1993.

Soble, Alan. "The Coherence of Love." *Philosophy and Theology* 12, no. 2 (2000): 293–315.

Solomon, Robert C. "Erotic Love as a Moral Virtue." In *Virtue Ethics, Old and New*, edited by Stephen M. Gardiner, 81–100. Ithaca, NY: Cornell University Press, 2018.

Stern-Gillet, Suzanne. *Aristotle's Philosophy of Friendship.* Albany, NY: State University of New York Press, 1995.

Winthrop, Delba. "Aristotle and Theories of Justice." *American Political Science Review* 72, no. 4 (December 1978): 1201–16.

Journal of Ethics and Social Philosophy
VOL. 20 NO. 1 · JULY 2021
https://doi.org/10.26556/jesp.v20i1.1072

DOES CONVERGENCE LIBERALISM RISK ANARCHY?

Marcus Schultz-Bergin

PUBLIC REASON, or political, liberals argue that coercive social arrangements must be publicly justified in order to be legitimate.[1] On the traditional "consensus" model of public reason most commonly associated with John Rawls, this means that such arrangements must be justified by appeal to a set of "public reasons" that are shared by, or accessible to, all reasonable citizens. On the more recent convergence model, however, an arrangement may be publicly justified so long as every citizen has sufficient reason of her own to accept the arrangement, even if those reasons are not shared by other members of the public.[2] The corollary of this change in justification is that, on the convergence model, non-shared reasons may also function as *defeaters*: in contrast to the consensus model where (for instance) religious reasons play no role in justifying or defeating laws, on the convergence model they may form a patchwork of justification or function all on their own as reasons to reject a law.[3]

This ability to leverage non-shared reasons to defeat laws has led critics, most notably David Enoch, to argue that convergence models are incapable of vindicating liberalism. They argue that someone will have a defeater for every possible law, and so the convergence model leads to anarchy—the view that no law is legitimate.[4] Relatedly, some critics suggest that convergence models can, at

1 There is disagreement among public reason liberals as to what, precisely, stands in need of justification: the basic structure of society, constitutional essentials, laws, moral rules, etc. I use the phrase "coercive social arrangements" as a catchall, but will occasionally reference more precise instances (such as laws) for convenience's sake. Nothing I say, except in section 6, turns on the precise sort of arrangement we are considering.

2 It is worth flagging here that I have only said convergence views permit "non-shared" reasons. I have not claimed convergence views permit "non-accessible" reasons since it is an open debate among convergence liberals whether that is true and exactly what that means. But for the purposes of this paper it does not really matter what position we take on that debate.

3 Gaus and Vallier, "The Roles of Religious Conviction in a Publicly Justified Polity."

4 Enoch, "Against Public Reason." See also Wall, "Public Reason and Moral Authoritarianism."

best, only vindicate the minimal state and, therefore, make libertarians "dictators" since they will defeat any law beyond their minimal preferences.[5]

My aim in this paper is to defend convergence liberalism against this *anarchy objection*. Using Enoch's presentation of the objection as the exemplar, I argue that we should reject both premises of the objection. By and large, I suggest, the anarchy objection rests on misunderstandings or misrepresentations of the convergence model. In the process of responding to the anarchy objection, then, I will also be clarifying key elements of the convergence model. My response is largely a general response on behalf of all convergence models of public reason.

However, in section 6 I will argue that a full response to the anarchy objection requires adopting the "wide scope" position in the *scope debate* among political liberals. Currently, there is debate as to what sorts of coercive social arrangements—constitutions, laws, social-moral rules, etc.—are subject to public justification. Although there are a variety of considerations to this debate, my argument will provide a distinct reason in favor of the view that all coercive social arrangements should be subject to public justification. To that end, consideration of the anarchy objection can help us make progress in the development of convergence liberalism (and political liberalism more generally). I further argue that, in overcoming the anarchy objection, the resulting view of convergence liberalism is much more open to views and policies commonly associated with "social justice liberalism" than is traditionally thought.

1. The Anarchy Objection to Convergence Liberalism

For public reason liberals, the exercise of legitimate authority is constrained by a public justification principle (PJP). The exact content of that principle varies among different theorists, but a general statement of the PJP holds that:

> A coercive social arrangement S is justified in a public P only if each member i of P has sufficient reason(s) R_i to endorse S.[6]

Convergence liberals and consensus liberals differ on how they interpret some of the variables in the PJP. Two of these differences, the interpretations of R and P, are what lead to the anarchy objection. First, convergence liberalism allows non-shared reasons to function within public justification. Second, convergence liberals endorse a *moderate* conception of idealization, rather than the more substantial or radical forms of idealization favored by consensus theorists. These two features together mean that the set of reasons that can function to both jus-

5 Lister, "Public Justification and the Limits of State Action."

6 Vallier and D'Agostino, "Public Justification."

tify and defeat laws is significantly larger than under the consensus conception. Idealized members of the public are understood to hold onto many more personal beliefs and those personal beliefs are relevant to public justification.

Convergence liberals argue that their approach better fits with the underlying motivations for public reason liberalism.[7] For Rawls, the public reason project arose out of two observations: the fact of reasonable pluralism and the freedom and equality of all persons.[8] The freedom and equality of all persons drives the requirement to justify any coercive authority exercised over them while the fact of reasonable pluralism shows the difficulty in establishing that justification. Convergence liberals note that the consensus liberal solution effectively jettisons the fact of reasonable pluralism by substantially idealizing members of the public to the point where they are effectively the same person.[9] Moreover, consensus approaches fail to fully respect persons as free and equal since they require citizens to ignore those beliefs and values that make them distinct individuals.[10] Convergence liberalism, on the other hand, aims to genuinely respect the fact of reasonable pluralism and the freedom and equality of all persons by maintaining a significant amount of evaluative diversity even after idealization.

While this move to moderate idealization and convergence justification may be well motivated, even convergence liberals recognize it has a "troubling dark side."[11] As Christopher Eberle observes, convergence models include "an extremely demanding understanding of what makes for justified coercion—if there is only one coerced citizen who has conclusive reason to reject [a law], then [the law] is morally wrong, even if [the law] would be justified absent that

7 Vallier, "Against Public Reason Liberalism's Accessibility Requirement," "Convergence and Consensus in Public Reason," and "In Defence of Intelligible Reasons in Public Justification." Importantly, I am merely relaying some of the arguments made by convergence liberals against consensus views, not endorsing them. Nothing in my argument will turn on the success of these criticisms.

8 Rawls, *Political Liberalism*, xviii.

9 Rawls effectively says as much about the parties to the original position; *A Theory of Justice*, 118–29. Quong endorses something similar in *Liberalism without Perfection* and "Liberalism without Perfection." Gaus makes something like the "same person" charge in *The Tyranny of the Ideal*, ch. 2. However, it should be noted that Rawls's larger view in *Political Liberalism* adds steps beyond the Original Position that involve less radical idealization. Quong's development of Rawlsian public reason liberalism, though, puts more emphasis on Original Position reasoning and thus gets closer to maintaining the "same person" concern that arose in *A Theory of Justice*.

10 Both Gaus and Vallier make this "alienation" charge: Gaus, *The Order of Public Reason*, 238; and Vallier, *Liberal Politics and Public Faith*, ch. 2.

11 Vallier, "Convergence and Consensus in Public Reason," 268.

lone dissenter."[12] When you combine sufficient evaluative diversity with an extremely demanding requirement of public justification, it seems quite obvious that many, and perhaps all, laws will fail to pass the test. And it is that seemingly obvious fact that motivates the anarchy objection.

David Enoch's statement of the anarchy objection begins with an observation of actual citizens and political debate in modern diverse societies. He suggests that such citizens are a "very varied bunch" with such diverse commitments that "if the justifications offered to them are to engage them as they actually are … then it's hard to believe that there is *anything at all* that can be justified to all."[13] Thus, he argues, without some form of idealization, anarchism results. Perhaps he is right, and in any event, we can accept this conclusion as far as it goes.

But Enoch also believes that convergence liberalism, with its emphasis on moderate idealization, will also result in anarchy. In particular, he argues that if we stick with a moderate theory of idealization, "the price would be anarchism again—for among those who are reasonable in just some very thin sense of this kind, *everything* is controversial … [and] *nothing* is justifiable to all the reasonable in *this* sense."[14] Although he does not offer a substantial argument for this claim—resting it largely on the reader's sympathy to the idea—he is not alone in drawing this conclusion and it appears to have significant intuitive appeal.[15] As such, it is vital for convergence theorists to have a clear response to this anarchy objection.

As I see it, Enoch's argument takes the following form:

1. Among moderately idealized members of the public, everything is controversial.
2. If everything is controversial, then nothing is justifiable.
3. Therefore, among moderately idealized members of the public, nothing is justifiable.[16]

We have good reason to doubt the truth of both premises, and in the course of my response I will argue that both are false. To begin, however, I will accept the truth of premise 1 and focus on premise 2, which I think is the more immediately intuitive premise since it takes much of its support from observation of actual

12 Eberle, "Consensus, Convergence, and Religiously Justified Coercion," 289.

13 Enoch, "Against Public Reason," 117–18.

14 Enoch, "Against Public Reason," 122.

15 E.g., Raz, "Facing Diversity," 46; Lister, "Public Justification and the Limits of State Action."

16 Steven Wall's version of the anarchy objection, leveled specifically against Gaus's convergence theory, has effectively the same structure. See Wall, "Public Reason and Moral Authoritarianism," 168.

political debate and actual citizens. I will turn to evaluating premise 1 in sections 4–7. My argument against premise 2 rests, broadly, on the claim that it depends on two misunderstandings of convergence liberalism. Once these two misunderstandings are clarified, we will see that convergence liberalism is well placed to deal with a society in which everything is controversial.

I will consider each misunderstanding of convergence liberalism separately, first focusing on how the premise misunderstands the notion of *sufficient reason* and, second, how it ignores convergence liberalism's insight that bare social coordination has moral and prudential value for all. With these two elements combined, we will be able to say that all "members of the public"—the name given to the moderately idealized public—despite total disagreement, will have sufficient reason to endorse key institutions of classical liberalism such as a system of private property.[17]

2. DISAGREEMENT AND SUFFICIENT REASON

Aly, Barry, and Chris are discussing dinner options. Aly, a committed vegan, strongly prefers a plant-centered restaurant while Barry is a meat-and-potatoes guy who strongly prefers a meat-centered restaurant. Chris is quite open to various options, caring much more about the social aspect of dinner than the food itself. Aly, Barry, and Chris are, in Enoch's words, a "very varied bunch," and so, on his view, even if they were moderately idealized, there would be no dinner option justifiable for them all.

We can motivate Enoch's position by considering how the dinner debate may go. Aly and Barry, with their strong and conflicting preferences, will stump for their preferred options. Barry may, in this debate, insist that he will only attend a dinner at a steakhouse while Aly will insist that she will only attend a dinner at a new vegan restaurant. In public reason terms, we may be tempted to say, as Enoch does, that only the new vegan restaurant is justifiable for Aly while only the steakhouse (or steakhouses) is justifiable for Barry. If this is indeed the case, then it does appear that nothing is justifiable for them all.

But this analysis of the situation is a mistake. While it is true that Barry and Aly disagree about which option is best—and we will assume such disagreement will survive moderate idealization—it is incorrect to assume that *only* their most preferred restaurant is justifiable for them all. A key element of the convergence model is that members of the public may have sufficient reason to endorse a pro-

17 Gaus introduces the phrase "members of the public" to describe the idealized public but it has been adopted by other convergence liberals as well. See, e.g., Gaus, *The Order of Public Reason*, 267; Vallier, *Liberal Politics and Public Faith* and *Must Politics Be War?*

posal—Aly and Barry may have sufficient reason to endorse a restaurant—that they nevertheless find suboptimal. And while in actual political debate we may, for strategic reasons, suggest that we will only accept our most preferred option, it is incorrect to read that sort of saber rattling into the public justification principle and its requirement that all moderately idealized members of the public must have sufficient reason to endorse a coercive social arrangement.

The aim of public justification is to reconcile the status of all people as free and equal with the authority we, as a society, exercise over each other. But since a society of free and equal people will necessarily be characterized by reasonable disagreement, we cannot expect everyone to agree on what is best. The whole point of a theory of public justification, then, is to provide "an account of how people who disagree on the best may still come to endorse a common rule."[18] Thus, any plausible theory of public justification must, by necessity, allow for the public justification of rules that some (or all) consider suboptimal. This also fits with our general understanding of what it means to live in a society. We commonly think that living with others requires "compromise"—it means not always getting what you believe is ideal.[19] But, of course, we also do not think that living in a society should require you to "compromise on your principles," and so it does not require living in accord with norms or rules that you find wholly unacceptable.

The above reflection on the point of, and underlying motivations for, public justification has direct implications for the construction of the deliberative model we use to investigate what sorts of policies pass the test of the PJP. Although convergence liberals construct the deliberative model somewhat differently, I will mostly draw on Gerald Gaus's construction in *The Order of Public Reason*, as it provides the most developed view and has been more or less adopted by most convergence liberals.[20] In our dinner example, we begin by constructing a ranking of options for each individual: clearly Aly ranks the new vegan restaurant as best while Barry ranks the steakhouse as best, but all three friends have other options on their rankings. To construct these rankings, we do not simply consider "Where do you want to go for dinner?" but instead construct a ranking based on pair-wise comparisons.[21] Like an eye exam where one cannot judge whether a lens correction is better without having another option to compare, the relevant question is "Would you prefer to go to dinner at restaurant A or restaurant B?" and so on for the various options. When the choice situation is modeled in this

18 Gaus, "On Two Critics of Justificatory Liberalism," 204.

19 See, e.g., Wendt, *Compromise, Peace and Public Justification*.

20 Gaus, *The Order of Public Reason*, ch. 5.

21 Gaus, *The Order of Public Reason*, 304.

way, we can start to see how additional options get added to each individual's rankings: Would Aly prefer to go to the steakhouse Barry recommends or would she prefer an exotic-meats restaurant? Would Barry prefer the vegan restaurant Aly recommends or would he prefer Chinese takeout?

But of course some options may be totally unacceptable. Aly, for instance, may be unable to eat a meal at the exotic-meats restaurant, and so we need some way of modeling the idea of an option being totally unacceptable. The commitment to the freedom and equality of all people establishes this baseline. Since no person is naturally under the authority of any other, for any given issue the natural baseline is one where each acts in accord with her or his own judgment and there is no common policy on the given matter.[22] It is always deviations from this baseline of natural freedom that stand in need of public justification. Each exercising her natural freedom need not be justified. And so, in the choice situation Aly, Barry, and Chris face, the natural baseline is that they do not go to dinner together at all. Naturally, if they cannot agree to a common place, they will go their separate ways. And for some of the possible options, such as the exotic-meats restaurant, some of the individuals would prefer to forgo the social dinner. For Aly, the exotic-meats restaurant is worse than no social dinner at all.

Notice the cost Aly pays for ranking the exotic-meats restaurant below the baseline. She, Barry, and Chris had first agreed to go to dinner together. They did this for a variety of reasons, not just because they were hungry. Although Chris is quite laid back about the food choices and cares a great deal about the social element, both Aly and Barry are also motivated by the social element of dinner. To rank the exotic-meats restaurant as worse than no dinner at all is not simply to refuse to eat exotic meats, it is to refuse all the other benefits that come along with a social dinner. Obviously Aly may have to make such a refusal in this case. But the important lesson to draw out is that ranking a proposal as worse than no common rule at all is no simple matter. It is not akin to the sort of "my way or the highway" saber rattling that is characteristic of actual political debate, where we may be blustering for strategic reasons, fully knowing that at the end of the day we will end up with some common rule. To suggest that an option is totally unacceptable is to forgo all the benefits that you attach to having a common rule at all.

Aly, Barry, and Chris all have sufficient reason to endorse any of the restaurant options that they prefer to not going to dinner at all. Given that each of them has some preference for going to dinner with each other—analogous to members of the public having some preference for living in society with one another—they all are likely to have a variety of options on their list. From this,

22 Gaus, *The Order of Public Reason*, 311–13.

we could construct a social ranking, identifying the options that are on all three lists. This would not determine which restaurant they should go to, but it would identify the set of eligible options. At that point, which one they select is not terribly important, for any of the restaurants would be justified for them all.[23]

The preceding discussion illustrates why Enoch's second premise is false. For Aly, Barry, and Chris, it was genuinely controversial which restaurant to patronize. But it did not follow that no option was justifiable for them all. Although Barry may claim, in discussion, that he refuses to eat at a vegan restaurant, to actually refuse such an option is to also refuse the company of his friends, and when the choice is put that way we can see why the vegan restaurant may very well be on his list of acceptable proposals.[24] Now, importantly, this discussion of sufficient reason is not meant to show that there is some justifiable arrangement with regard to *all* controversial matters. In certain situations, it may actually be that there is not a single proposal shared among all members of the public. If Aly and Barry do not really care to eat together, or care much more about getting their way than eating together, then maybe they will not be able to coordinate on any option. But merely disagreeing about what is best does not mean that nothing is justifiable.

There is, however, a further point to make about core liberal institutions. So far, Aly, Barry, and Chris have been considering restaurants that are heavily vegan focused or heavily meat focused. But now Chris suggests an Indian restaurant, which has a variety of meat and vegetable curries. In fact, the menu is split nearly in half between vegan options and meat options. Aly has a wide variety of vegan options to choose from while Barry has a variety of meat options to choose from. What was initially a collective decision over whether to eat vegan or not has now become an individual one. While there must still be a collective decision over the specific restaurant, at least one of the major issues that the friends were get-

23 Convergence models are a bit more complex than this. In particular, most endorse a Pareto rule that holds that if an arrangement is dominated by another arrangement in every member of the public's eligible set, then it is not part of the socially eligible set. Additionally, *how* a member of the socially eligible set becomes the norm can matter. If it is made the norm by a non-publicly justified decision rule, then it is still not publicly justified. But if it is "selected" by way of a publicly justified decision rule or as a result of social evolution, then it is publicly justified.

24 Those familiar with game theory will recognize that this is an example of a (slightly modified) Battle of the Sexes game where players disagree about what they should do but have an interest in doing something together. However, whereas in the standard Battle of the Sexes game players receive zero payout for not coordinating, I leave open that there may be some minimal payout for "eating alone." But the central point of the game, that an interest in doing something together can mean a player receives a positive payout even when they are not doing what they really wanted to do, still holds.

ting hung up on has been sidestepped. Or, more precisely, the decision has been "devolved" from the collective to the individual. This sort of devolving of authority is precisely what makes liberal institutions, such as private property regimes, so well placed to deal with evaluative diversity. As Gaus explains:

> A deeply pluralistic social order can effectively cope with many of its disagreements about what evaluative standards to adopt by establishing a system of private property.... Given the problem of evaluative pluralism ... each Member of the Public has a fundamental interest in instituting a system whereby the natural and social world is divided into different jurisdictions in which the evaluative standards of the ... rightholder ... will be determinative.[25]

Rather than Aly being beholden to Barry's dietary preferences, or Barry being beholden to Aly's, Chris has suggested an option that allows each of them to remain their own dietary master while still gaining the benefits of a friendly social dinner.

In the political realm, the schemes of rights and liberties that characterize a liberal order accomplish the same sort of goal as the Indian restaurant. There can be social coordination around a scheme of rights and liberties, thus establishing a common authority, but that scheme dictates various zones where the evaluative standards of the individual are supreme. In this way, liberal institutions neatly deal with the fact that "everything is controversial." And so, thoroughgoing disagreement need not impugn the public justification of core liberal institutions because such institutions do not involve restricting people's abilities to live their own lives their own way. Rather, they facilitate such an ability by establishing socially recognized jurisdictions of individual authority. Thus, every member of the public will have sufficient reason to endorse some liberal scheme of basic rights and liberties. This is certainly true so long as they have any interest in cooperating or establishing a stable set of social norms. But it remains true (albeit with less force) so long as they have an interest in pursuing their own individual projects and can be understood to see the fact that certain sociopolitical institutions make that possible, likelier, or easier. There will likely still be disagreement over which scheme is best, but so long as there is a non-empty set of socially eligible schemes, then public justification can vindicate the core of liberalism and thereby avoid outright anarchy.

25 Gaus, *The Order of Public Reason*, 375–77.

3. THE VALUE OF SOCIAL COORDINATION

Chris never expressed strong dietary preferences. Rather, her main concern was that she, Aly, and Barry have dinner *together*. She heavily valued social coordination, and so her ranking of various restaurants would likely be influenced more by what she thought both Aly and Barry would agree to than by what sounded tasty to her. Hence her suggestion of the Indian restaurant. There are people like Chris among the general public, who are heavily motivated by a desire for a good social and political community and not so much by a desire to have things "their way." These sorts of people help illustrate the value of social coordination for everyone, and this is important especially for those people who seem to be much more concerned about getting things their way rather than simply coordinating with others. Social coordination, specifically in the form of a liberal scheme of rights and liberties, is good for everyone, and this provides important moral reason for everyone to endorse a common norm over no norm at all, even if they do not believe it is the optimal norm. This shows another way in which we can establish public justification in the face of evaluative diversity.

Let us imagine that Barry has no food at home, and so his only option for eating dinner is to eat out. But Barry also does not like to eat alone. Thus, for him, successfully coordinating with Aly and Chris helps him pursue his individual plan of actually eating. Much of social coordination is like this, especially when the coordination comes in the form of a liberal scheme of basic rights and liberties. Such a scheme establishes a framework under which people may better pursue their own plans and projects. For Barry, agreeing to go to the Indian restaurant—establishing the liberal scheme—allows him to achieve his goal of eating a meat-based dinner with others. He was able to achieve this goal without forcing anyone else to share his goal or otherwise have to forgo their goals. Aly still had her vegan meal and Chris still had a social dinner. Each of these friends had individual aims, and it was the social coordination that made the achievement of these aims possible. This is the moral and prudential value of social coordination, and it is a major reason why diverse individuals have sufficient reason to endorse at least some common authority, and most notably the sort of devolved common authority that is characteristic of liberalism.

The fact of reasonable pluralism is not merely the fact that people disagree. What is more important, socially speaking, is that individuals act based on their beliefs and values, and pursue plans and projects that fall in line with those beliefs and values. If people disagree about fundamental beliefs and values, then it is likely to mean that their ability to pursue their plans and projects will conflict with the ability of others to do the same. If the disagreement remained wholly

"in the head," as it were, then reasonable disagreement would probably not be much of a concern at all. But it is precisely the fact that people act on their beliefs and values, and that doing so may lead to conflict when people disagree, that makes a liberal scheme of rights and liberties so attractive and essential.

A liberal scheme of basic rights and liberties establishes a framework under which people may better pursue their own plans and projects. By establishing a common understanding of the contours of individual jurisdictions, such a scheme drastically improves everyone's ability to form and carry out their plans. This is because social coordination, particularly around a scheme of basic rights and liberties, drastically reduces the uncertainty that usually accompanies evaluative diversity. Take the example of private property. Without any common understanding of a right to private property, people could still collect and store items. However, there is a much greater risk in doing so since one cannot predict—on the basis of a common understanding of the right of private property—whether others will attempt to take what one has collected. This problem is further magnified when thinking about the pursuit of long-term plans and projects, which often include the collection of various items along the way. The successful completion of a long-term project may depend on many smaller steps of collection and maintenance of items, but if one is quite uncertain whether the items she collected last year will still be in her possession next year, it may simply be imprudent to pursue the long-term project at all.

What the above discussion indicates is that social coordination has important prudential value. It makes sense, from a purely self-interested perspective, to grant to others certain basic rights in order to secure those basic rights for oneself. In so doing, it becomes much easier to pursue one's own life plans. Social coordination, particularly in the form of a scheme of basic rights and liberties, creates a background of stability that enhances everyone's ability to pursue their own life their own way. Insofar as everyone is interested in pursuing their own life their own way, everyone has at least some reasonably strong interest in establishing a scheme of basic rights and liberties.

The value of social coordination is not merely prudential. First, most individuals believe their plans and projects are morally valuable. We do what we do because we think it is worth pursuing. And so, a framework that helps us pursue those projects also has moral value, at least from the individual's perspective. Furthermore, a common social norm makes moral relations among free and equal people possible.[26] And here is where the "moderate idealization" that convergence theorists impose on Members of the Public is important, for such idealized persons are understood to view their fellows as free and equal people

26 Gaus, *The Order of Public Reason*, 8–20.

and desire to properly respect that status by establishing and complying with publicly justified social norms. Absent a publicly justified social norm not only is conflict more likely but there are no morally appropriate responses to that conflict—no responses that properly respect others as free and equal; if there are no norms governing property, then when someone steals your property your resentment is inappropriate—it fails to regard the other as a free and equal person.[27]

That social coordination is of both prudential and moral value helps us understand further why the presence of widespread disagreement does not imply anarchy. People are not just interested in society being organized in accord with their own worldview; they are also interested in living in a society with others.[28] This interest follows from both the moral and prudential value of social coordination. But, more generally, as social creatures, human beings are driven to live in a society and so are motivated, to various degrees, to reconcile their differences to maintain a stable society.

Individuals differ on the strength of their commitment to cooperation and of course on how they rank various options under consideration. Some individuals will stick to their preferred action or rule, even as a significant portion of society coalesces around an alternative; others, like Chris, will often join with the crowd, being strongly motivated to establish a common social arrangement or not being strongly committed to the norm she identifies as best. But as work in evolutionary game theory has shown, as more people coalesce around a specific norm or rule, the dissenters have stronger reason to also join in.[29] For instance, insofar as Barry does not want to eat alone, then if Aly hops on board with Chris's recommendation of Indian food, he has even greater reason to accept that option than he did prior to Aly's assent. This implies that even those who are strongly committed to an alternative norm or rule, or are not terribly motivated to reconcile, can be brought on board as more members of society begin to coalesce.[30] Thus,

27 Building in the requirement that one views one's fellow citizens as "free and equal" is common across (nearly) all political liberal theorists. Different theorists justify that idealization in different ways and identify different implications that follow from it. What I have suggested here largely follows Gaus's and Vallier's interpretations, which suggest that the idea of everyone as "free and equal" is baked into our practices of social morality. Rawls, in contrast, draws the idea from the "shared political culture" of a democratic society. See Gaus, *The Order of Public Reason*, ch. 4; Vallier, *Must Politics Be War?* ch. 3; Rawls, *Political Liberalism*, 29–34.

28 Gaus, *The Order of Public Reason*, 398–99.

29 Skyrms, *Evolution of the Social Contract* and *The Stag Hunt and the Evolution of Social Structure*.

30 Gaus, *The Order of Public Reason*, 409–23.

the presence of wide-ranging disagreement about what is best, which could perhaps set people at odds, is offset by their motivation to live together in a society.

What about the person whose projects include directing or constraining the behavior of others?[31] This could be a person whose projects directly include wielding power over others or it could include a person who simply subscribes to certain views about how society should be structured—for instance someone who subscribes to traditional gender roles that fly in the face of sexual equality. To illustrate, we could add two new people to our restaurant example: Dave and Emily. Dave is of the view that unrelated men and women should not mingle in public and so while he, too, wants to eat dinner, he believes it is only acceptable for he and Barry to eat together. The women may also eat together but must do so separately. Emily, on the other hand, sees it as her role to promote the virtue of others by, in this case, ensuring that the group only eats vegan food. Like Aly, she herself wants to eat vegan food but, unlike Aly, she wants everyone else to also eat vegan food. Once we add Dave and Emily into the mix, it starts to look much less likely that the group can find common ground. People like Dave and Emily seem to pose a particular threat to the development of a liberal order.

There are a few things to say about this concern. First, the argument thus far was aimed at showing the possibility of agreement on core liberal institutions, not the possibility of agreement "all the way down." For someone like Dave, strong protection of the freedom of association allows him to form a voluntary community of people who prefer to live in accord with a traditional view about gender norms. This may not be all he wants, but if the alternative is no protection for his ability to form such communities—since plenty of people will have defeaters for making the entirety of society adhere to such norms—then it is likely to be justified for him. But if we were thinking more about a "downstream" policy—say paid parental leave—that can only be justified by the claim that it promotes sexual equality, then perhaps it cannot be publicly justified so long as we have people like Dave in the society. While I think it is unlikely that such a policy can only be justified in that way, to the extent that is true, then it does show that in a sufficiently diverse society some policies that are near and dear to certain progressives may be unjustified.

Someone like Emily, or Dave if he insists that all of society must adhere to his view of gender norms, may appear to pose a more significant problem. But here it is important to remember that it is coercion that stands in need of public justification. Indeed, the paid parental leave policy considered above may simply not be a policy in need of public justification just because it need not be coercive (depending on the details). On the other hand, Emily is advocating coercion—

31 I thank an anonymous reviewer for encouraging me to consider this issue.

she is suggesting that others should be coerced into eating a vegan diet. Similarly, a Dave who insists that unrelated men and women cannot mingle in public is aiming to coerce others. In these situations, it is *their* proposal that must be publicly justified and they quite clearly cannot be in a sufficiently diverse society. Thus, such people pose little threat to others in terms of generating coercion.

Finally, consideration of people like Dave and Emily helps us extract some of the other key elements of convergence liberal theory. For one thing, the theory of moderate idealization that convergence liberals employ can help to eliminate some of the worst sorts of proposals from consideration. Proposals that depend on the rejection of the freedom and equality of all people are simply idealized away.[32] This not only helps us eliminate some coercive proposals but also helps us potentially eliminate certain defeater reasons people may otherwise seem to have. As discussed earlier, convergence liberals do note an important asymmetry in their position—a lone dissenter may be sufficient to defeat a proposal but is never sufficient to adopt one—and they tend to want to minimize restrictions on defeaters. Nonetheless, that does not mean anything goes when it comes to defeaters. Moderate idealization still applies, and so some reasons that actual people will actually profess may disappear under idealization.

So, convergence liberalism, at least at the level of core liberal institutions, can handle concerns related to those who aim to dominate others or who hold views that imply restrictions on the behaviors of others. This does not eliminate the possibility that such concerns may function as *defeaters* for downstream policies and, indeed, beginning in the next section we will focus on moving more toward downstream policies to see what sorts of resources convergence liberals have to handle those sorts of issues. But the existence of such people does not impugn the ability to at least get the basic liberal project off the ground.

4. Classical Liberalism and the Min-archy Objection

Those who level the anarchy objection against convergence liberalism suggest that because moderately idealized members of the public are still quite diverse, no law or social arrangement will pass the test set down by the PJP. The preceding discussion established that this need not be the case. Even in a sufficiently diverse society, everyone has sufficient reason to endorse at least the core institutions of liberalism. This is especially so since core liberal institutions, such as schemes of private property and rights to privacy and freedom of expression "economize on collective justification."[33] Rather than seeking a substantive com-

32 Gaus, *The Order of Public Reason*, 279.

33 Gaus, *The Order of Public Reason*, 374.

mon viewpoint in the face of increasing diversity, liberal institutions only require a minimal common agreement—that within certain realms the evaluative standards of the individual are authoritative. Thus, rather than having to forgo acting in accordance with one's own conception of the good, in order to live in society, liberal institutions secure significant protections for those pursuits.

Thus far I have argued that we should reject Enoch's second premise—it is simply not true that just because everything is controversial nothing will be justifiable to all. But someone leveling the anarchy objection may reasonably retort, "Sure, you have shown that a minimal state can be vindicated against anarchy, but my real concern with convergence liberalism is that it cannot vindicate any more than that minimal state." By setting the bar on justification for coercion so high, those members of the public who strongly oppose state coercion—classical liberals or libertarians—will always block the justification of laws or social institutions that have tended to define the modern liberal state: arrangements that establish the illegality of discrimination and the provision of social welfare, for instance.[34] Indeed, most (perhaps all) of those who level the anarchy objection tend to be in favor of various redistributive or welfare schemes and so, for them, a theory of political legitimacy that rules out their preferred schemes is a theory that should be rejected.

On this version of the anarchy objection, which we may call the "min-archy" objection, convergence liberalism makes libertarians effective "dictators" and, for that reason, should be rejected. In order to respond to this version of the objection, I will now argue that Enoch's first premise—that among moderately idealized persons everything is controversial—should be rejected. In particular, what Gerald Gaus has called the "perspective of agency" is shared among all members and can provide the basis for the possible legitimacy of redistribution and welfare rights or protections. I will further suggest that the min-archy objection should force convergence liberals to refine their model. In particular, there is disagreement among public reason liberals in general, and convergence liberals in particular, over the "scope" of public justification—that is, what stands in need of justification. For some, it is specifically laws that stand in need of justification, while others take a wider perspective and claim that all social coercion stands in need of justification. As I will suggest, it is only this wide-scope model of convergence liberalism that can wholly overcome the min-archy objection and so convergence liberals should extend the PJP to non-state coercion as well.[35]

34 Raz, "Facing Diversity," 47; Lister, "Public Justification and the Limits of State Action," 153–56.

35 Interestingly, I also think that the "narrowest" scope view—that only "constitutional essentials and matters of basic justice" stand in need of public justification—may also overcome

5. Moderate Idealization and Shared Reasons

Each member of the public has her own plans and projects, which are partly determined by her conception of what makes for a good human life (which includes a conception of morality more generally). Because in a pluralistic society we regularly run up against different ways of living, and our own way of living is challenged and open to revision, we all cannot help but understand ourselves as deliberative agents.[36] Thus, as Gaus suggests, despite our significant disagreements, we all share in the perspective of agency. By this he does not mean we view ourselves as "autonomous" agents in any thick sense, but simply that we all see our "actions as following from [our] own deliberations" even if those deliberations are "unreflective, traditional, or superstitious."[37] Insofar as we view ourselves as naturally free, we accept that we are agents. Whether we embrace the sort of individuality Mill advocated for or defer to authority, we are deliberating (at least in a thin sense) about how our lives should go and making choices based on those deliberations.

All people, at the level of moderate idealization, can be understood to share this perspective of agency—they all view themselves as agents, even if they disagree about everything else. This follows quite naturally from the starting point of public reason—that all persons are free and equal—and from recognizing that the whole enterprise of requiring justification assumes that the individual is an agent. If an individual is not an agent, if her actions do not follow from her own belief-value set, her own deliberations, her own choices, then there is simply no reason to worry about justifying coercion to her. It is because coercion, when unjustified, aims to usurp the authority of the individual agent that it must be justified. In this way, it is simply not possible to escape the perspective of agency.

This shared perspective of agency forms the basis, in the convergence tradition, of the justification of the core liberal institutions. Indeed, the perspective of agency was simply assumed in my earlier defense of liberal institutions when I emphasized the value that all people place on pursuing their own projects (i.e., exercising their agency). But the appeal to the perspective of agency does more than this. It is from this noncontroversial shared perspective that we can begin to generate other core elements of the modern liberal state. Now, importantly, that all members of the public can be understood to share this perspective, and there-

---

the min-archy objection. However, I am not familiar with any convergence liberal who adopts that position in the debate. Nearly everyone seems to think Rawls was wrong to restrict the scope of public justification in this way.

36 Gaus, *The Order of Public Reason*, 337.

37 Gaus, *The Order of Public Reason*, 338–39.

fore to share the set of reasons that that perspective generates, does not mean that they will always come to the same conclusions about matters where those reasons are relevant. This shared perspective sits aside various unshared perspectives, and although the shared perspective of agency generates very strong reasons—given its centrality to all other elements of an individual's belief-value set—those reasons still must compete with the reasons generated by various other commitments each individual has.[38] Nonetheless, identifying the sorts of reasons the perspective of agency generates and the sorts of coercive social arrangements those reasons support helps us understand how certain features of the modern liberal state can be legitimate.

Consider, most centrally, claims about welfare. A distinguishing feature of modern liberal societies is a commitment to the welfare of citizens. From the shared perspective of agency, we can begin to see how the eligibility of welfare protections becomes possible. An individual's ability to realize her agency can be impugned in all sorts of ways, including through starvation and ill health. As such, a commitment to one's own agency will lead one to endorse some basic welfare rights. As Gaus claims, "Members of the Public, aware only that they are agents, and of their reasons to maintain their agency and to be successful as agents, would insist not only on the freedom to pursue their agency but also the necessary means."[39] Effective agency is the ability to potentially successfully achieve one's ends and requires more than mere noninterference. This is a position some libertarian thinkers explicitly endorse, and a position that all can be understood to hold under moderate idealization.[40] That is because if one views oneself as an agent, and recognizes that one's agency is central to everything else one values—for one cannot successfully pursue any plans or projects if one's agency is eliminated—then one always has strong reason to preserve and even enhance her agency. In this way, everyone will be committed to basic welfare provisions.

The exact nature of those welfare provisions, however, is indeterminate. The shared perspective of agency provides but one set of reasons in favor of welfare schemes. Other, diverse considerations, such as desert, may count against

38 It should be added that, for the convergence theorist, not only must the shared reasons compete much more fully with unshared reasons, but the moderate idealization means there is a lot less that is "shared" than for the consensus theorist.

39 Gaus, *The Order of Public Reason*, 359.

40 Lomasky, *Persons, Rights, and the Moral Community*, 126. This also seems to be what Nozick has in mind in his discussion of permitting "boundary crossings" and compensation, although he does not clearly endorse a specific proposal. See Nozick, *Anarchy, State, and Utopia*, 78–87.

certain sorts of welfare schemes. And, of course, the strength of one's concern with state coercion may alter one's rankings of various welfare proposals, and render some illegitimate. Nonetheless, at the level of moderate idealization all members of the public have some reason to support welfare protections, despite what an examination of actual political discourse may suggest. We may doubt this when we consider the libertarian who insists on very strong rights to private property and suggests that any sort of taxation or redistribution amounts to theft. However, as I have suggested above, the idealized libertarian would likely still support some basic welfare protections insofar as such protections preserve effective agency.

Moreover, it is worth remembering that even the libertarian is very concerned with the existence of a (minimal) state that protects his property rights. And so here we can leverage the great importance libertarians place on the protection of property rights to understand how a society, composed of libertarians and non-libertarians alike, would come to coordinate around some basic welfare protections, possibly including some redistributive schemes. This is because the public justification of a scheme of private property is inextricably linked to a discussion of welfare rights and the distribution effects of that scheme of private property.[41] This makes sense given that it is the shared perspective of agency that justifies both abstract private property rights and abstract welfare rights. But this implies that, as members of the public rank the various options related to private property rights, they are also considering potential redistribution or welfare schemes. For some members of the public, an extensive scheme of private property rights with only minimal provision for the poor will be understood as exceedingly coercive, so much so that the costs outweigh the benefits and thus such a scheme is outside that member of the public's ranking of proposals.[42]

The forgoing argument draws on two important and related issues in political liberalism: issue individuation and justificatory dependence. In brief, it is an open question whether we should (for instance) consider the public justifiability of laws one by one or whether, at least some of the time, we must consider multiple laws as a single issue. There may be good reason to consider multiple laws as a single issue since there can be various interactive effects between the laws, and so whether a member of the public has sufficient reason to endorse one law will depend on the status of the other. And, of course, we can ask similar questions about a single piece of legislation: Should different parts of that

41 Mill, *The Collected Works of John Stuart Mill*, vol. 2, bk. 2; Brettschneider, "Public Justification and the Right to Private Property"; Gaus, *The Order of Public Reason*, 522.

42 Gaus, *The Order of Public Reason*, 507, 526; Brettschneider, "Public Justification and the Right to Private Property," 132–35.

piece of legislation be considered independently? Or should the legislation be considered as a whole? Here I adopt what appears to be the common view in the literature: it depends. According to Gaus, we ought to consider multiple issues simultaneously—that is, the issues display justificatory dependence—if any member of the public's ranking of proposals would change upon considering the issues together *and* that would also change the social ranking of proposals.[43]

The justification of private property rights and the justification of abstract welfare rights are interdependent. Given the previous argument of the inextricable link between private property rights and abstract welfare rights, it should be clear that there will be at least some members of the public whose rankings of private property schemes will change depending on the justificatory status of abstract welfare rights (and vice versa). Since some people may reasonably hold that extremely strong protections on private property can only be justified if there are sufficient welfare guarantees, the social ranking of proposals will also vary depending on whether we consider the two issues separately or together. Thus, we cannot consider the justification of private property rights in isolation from consideration of welfare rights. Therefore, in constructing our social ranking, proposals with no welfare protections will be eliminated, as presumably will proposals with no protections for private property. We will then be left with a set of proposals that mixes, in various ways, protections for private property and welfare guarantees.

It is crucial here to remember how the eligible sets of proposals are constructed: the question is, would this proposal on the matter be better than no coordination at all? And so, given the libertarian's deep concern with the protection of property rights and the justificatory dependence of private property protections and welfare rights, we can see that when the alternative is no protection for private property, everyone is likely to have sufficient reason to endorse some welfare protections. Importantly, however, on this view it is not obviously the case that the state must be the guarantor of those welfare protections.[44] It is an open question whether there may be non-state-based means to instantiate the abstract welfare rights. Nonetheless, given that instantiating any abstract private property protections will almost certainly require the state, and indeed if we imagine the question to be "What scheme of private property rights should the state enforce?" then it is very likely that some of the proposals in the social ranking will involve the state in both protecting private property rights and providing for basic welfare.[45]

43 Gaus, *The Order of Public Reason*, 495.

44 I thank an anonymous referee for encouraging me to clarify this point.

45 Additionally, if one adopts a view where *only* state-based coercion stands in need of public

And so, even if Enoch were correct that nothing can be justified to all if everything is controversial, he is wrong to assert that among moderately idealized members of the public everything is controversial. The perspective of agency is not, in itself, controversial. And although precisely what it will be taken to support will vary among individuals, it can nonetheless provide the basis of establishing some of the core features of modern liberal welfare states and, therefore, provide the basis for escaping the minimal state.

6. Social Coercion and the Wide Scope of Public Justification

The preceding discussion of welfare rights can be understood as part of a broader concern some may have with convergence models: that they cannot justify the sorts of social justice policies that many contemporary progressives advocate. Policies relating to the welfare of the poor are but one subset of concerns we may have in society. We may also be concerned with the protections of various vulnerable populations and so advocate for anti-discrimination policies. And yet some in our society have suggested that anti-discrimination and related civil rights policies, including the landmark Civil Rights Act, are instances of unjustified coercion.[46] The claim here is that such protections involve coercing private citizens, or privately held businesses, to change their practices in ways incompatible with their personal commitments. Requiring businesses to serve people regardless of race, for instance, coerces business owners because it tells them they must act in a certain way or risk punishment. More generally, the worry may be that anti-discrimination policies are unlikely to be publicly justified and so even if convergence models do not risk anarchy, they risk what many would identify as a significantly unjust society.

This sort of concern is well founded if applied to certain versions of the convergence view. If only laws, or state-based coercion, stand in need of public justification, then it does make it difficult to justify anti-discrimination policies.[47] For instance, policies that make it illegal to refuse service to people based on race, or sexual orientation, are straightforwardly coercive.

---

justification, and that includes state-based protections for private property, then it is obvious that the state will be involved in the related welfare protections as well. But, as I argue in the next section, we should not hold that only state-based coercion stands in need of public justification.

46 Both Rand Paul and Ron Paul have said things to this effect. See, e.g., Rand Paul's 2010 interview with the *Courier-Journal* at https://www.courier-journal.com/videos/opinion/2014/07/09/12376813.

47 Vallier's position in *Liberal Politics and Public Faith* is an example of a narrow convergence view. He expands the view in his later work.

But here the remedy for the problem is to recognize that it is not just state-based coercion that stands in need of justification. Neither Rawls nor Gaus restricts the principle of public justification to state-based coercion.[48] For as Mill argued, the forum of public opinion, which may be exercised by the majority, can represent a "social tyranny more formidable than many kinds of political oppression, since, though not usually upheld by such extreme penalties, it leaves fewer means of escape, *penetrating more deeply into the details of life*, and enslaving the soul itself."[49] Following this insight, a "wide" convergence model expands the scope of public justification to social coercion as well. And in so doing, we can take account of many of the social justice concerns that may otherwise worry us.

Socially enforced discrimination is coercive.[50] It can significantly restrict the life prospects of various citizens and seems quite obviously to fail to regard many citizens as free and equal. If we follow Mill (and, as I have suggested, Rawls and Gaus) and place such social discrimination within the purview of public justification, then we get a better view of the relationship between discriminatory social arrangements and anti-discrimination laws. In particular, we can evaluate the coercive social arrangements as themselves unjustified—for surely those being discriminated against do not have sufficient reason to endorse them. But, further, we can evaluate the anti-discrimination laws, depending on how they are crafted, as either coercion neutral or net coercion reducing. This is because the laws themselves, while having coercive components, would presumably correct for the unjustified coercion already present in the society. Of course, the details will matter here and not just any anti-discrimination legislation would be acceptable. The central point is that this sort of explanation of the role and effect of anti-discrimination laws fits better with our intuitive understanding of the laws.

Anti-discrimination legislation is one way state coercion can be a remedy for unjustified social coercion, but other "social justice" causes may also be plausibly understood as calls for the state to take coercive action to reduce or eliminate unjustified social coercion.[51] Here, the right way to understand the situation is that (at least) those who are being (for instance) discriminated against by way of various social norms do not have sufficient reason to endorse that social coer-

48 Rawls, *Political Liberalism*, 269; Gaus, *The Order of Public Reason*.

49 Mill, "On Liberty," 220.

50 While I think this claim is largely noncontroversial, Colin Bird defends it in detail ("Coercion and Public Justification").

51 For instance, I argue elsewhere that this route can make sense of calls for stricter animal protections by indicating that various property norms surrounding animals are actually unjustified.

cion. The social coercion of discrimination can come in various forms, but often involves norms that limit opportunities and access for certain groups of people and is backed by the threat of social sanction or violence. For instance, informal segregation involves social norms that restrict where certain people may go. That there already exists social coercion for a given issue alters the coercion costs of state-based intervention. The mistake some self-proclaimed, liberty-loving politicians make, in rejecting anti-discrimination legislation, is to view the situation without the legislation as noncoercive (regarding the relevant issue). It is not, in fact, noncoercive; it is just that they are not the ones who are being coerced.[52]

Importantly, the argument I have just given in favor of (some) social justice policies directly counters the conclusion most (perhaps all) convergence liberals draw about their own view. They often claim that their theory shows a "classical liberal tilt" and Gaus goes so far as to claim that "principles of social justice [are] not stable under full justification."[53] Now, I am not impugning all of the specific arguments he makes there regarding specific abstract social justice principles; he may be right as far as those go. But the argument I have given in this section gives us a different way to think about social justice issues than he discusses. He focuses on whether basic principles of social justice may be justified early in the "order of justification," where other basic rights are justified. The argument I made here does not depend on social justice principles being justified in this way, but instead suggests that we can often think about calls for "social justice" as being calls for a *net reduction* in coercion by emphasizing that *existing* (usually social) coercion is illegitimate. In this way, I am employing what Gaus calls the "testing conception" of public justification that focuses on evaluating whether *existing* social norms are morally acceptable rather than making positive arguments for the moral acceptability of new forms of coercion.[54] In this way, I am bringing more in line the strong anti-coercion view commonly associated with classical liberalism and the strong pro-equality view commonly associated with social justice liberalism. In effect, I agree that an emphasis on public justification will have a "classical liberal tilt," but reinterpreting what that means by emphasizing the social coercion as well as the state-based coercion. In short, whereas many

52 Importantly, this claim is not universal. Many instances of social discrimination are certainly coercive, as when individuals are excluded from certain employment or educational opportunities due to race or gender, but merely limiting freedom—as when private organizations restrict membership—need not be construed as coercive. I thank an anonymous referee for encouraging me to clarify this claim.

53 Gaus, *The Order of Public Reason*, 521. Although this quotation can be found later in the book, he refers to the arguments he makes earlier, 359–68.

54 Gaus, *The Order of Public Reason*, ch. 21.

have taken convergence liberalism to be quite hostile to policies and views commonly associated with social justice liberalism, the approach I have taken shows that there is plenty of room in convergence liberalism for social justice liberalism, suitably interpreted.

Before moving on, it is worth further highlighting the implications of this argument. Both Mill's liberalism and contemporary "social justice" liberalism are often identified, by political liberals, as "comprehensive liberalisms" that depend, for instance, on controversial views about human nature or motivation or thick conceptions of political concepts like equality. This is problematic, we are told, because not all members of a diverse public will hold those controversial views and thus the liberalism that results will not be publicly justified and hence, in Rawls's terms, will not be stable for the right reasons. Political liberalism has often been seen as a competing approach to these sorts of comprehensive liberalisms, working to eliminate any controversial metaphysical or ethical commitments in favor of purely "political" conceptions of the person (or citizen) and justice, among other concepts. What I have suggested, however, is that the views are much closer than many political liberals have suggested. Not only must political liberals integrate some of the deeper insights about the nature of society and social norms that comprehensive liberals like Mill identified, but convergence liberals in particular must reconsider the relationship between their view and various calls for social justice in modern society. While convergence liberalism may still have a "classical tilt," the tilt is not as great as Gaus and other convergence liberals have suggested.

7. CONVERGENCE LIBERALISM AND THE MODERN WELFARE STATE

The modern liberal state does much more than merely protect private property rights. While we should not think that all that the modern liberal state does is justified, we may also reasonably worry that a view of political legitimacy that vindicates only the minimal state has got something wrong.[55] The various welfare institutions of the modern liberal state do not seem obviously illegitimate, and so it should at least be an open question whether, in any given society, they—or other versions of them—can be justified.

I have suggested that a convergence liberalism with a wide scope of justification does, indeed, leave it an open question as to whether welfare institutions and other social justice policies are justified. Critics are correct to oppose convergence models that restrict the scope of public justification to state action,

55 Rawls, *Political Liberalism*, 262–65; Lister, "Public Justification and the Limits of State Action."

as these inappropriately ignore Mill's insight regarding social tyranny. Thus, in moving forward with the public reason project, our focus should be on models with a wide scope.

A vindication of core liberal institutions over anarchy and of something approximating the modern liberal welfare state over the minimal state should render the min-archy objection toothless. In the vast majority of cases where we may want to forgo public justification in the name of justice—we do not think we need to justify anti-racism policies to the racist—a model of convergence public reason can make sense of how we can have both public justification and justice.

Enoch is a staunch critic of the entire public reason project. He has posed various objections to the consensus model, objections that convergence theorists are largely happy to accept. But he is wrong to conclude that the public reason project is hopeless and should be abandoned. Although the consensus model may fail, the convergence model remains an attractive alternative. Not only can it vindicate classical liberalism, but it can also vindicate the modern liberal welfare state. In a contemporary society, characterized by significant, but reasonable, disagreement, the convergence model of public reason can provide the framework for distinguishing legitimate from illegitimate coercive social arrangements.[56]

Cleveland State University
m.schultzbergin@csuohio.edu

REFERENCES

Bird, Colin. "Coercion and Public Justification." *Politics, Philosophy and Economics* 13, no. 3 (August 2014): 189–214.

Brettschneider, Corey. "Public Justification and the Right to Private Property: Welfare Rights as Compensation for Exclusion." *Law and Ethics of Human Rights* 6, no. 1 (December 2012): 119–46.

Eberle, Christopher J. "Consensus, Convergence, and Religiously Justified Coercion." *Public Affairs Quarterly* 25, no. 4 (October 2011): 281–303.

Enoch, David. "Against Public Reason." In *Oxford Studies in Political Philosophy*,

56 Earlier versions of this paper, or ideas contained within, were presented at the Rocky Mountain Ethics Congress and Adelphi University. I am grateful to the participants at each for their comments. I would also like to thank Michael Weber, Kevin Vallier, and John Basl for feedback on earlier drafts, as well as the referees and editors of this journal.

vol. 1, edited by David Sobel, Peter Vallentyne, and Steven Wall, 112–42. Oxford: Oxford University Press, 2015.

Gaus, Gerald. "On Two Critics of Justificatory Liberalism: A Response to Wall and Lister." *Politics, Philosophy and Economics* 9, no. 2 (May 2010): 177–212.

———. *The Order of Public Reason: A Theory of Freedom and Morality in a Diverse and Bounded World*. New York: Cambridge University Press, 2012.

———. *The Tyranny of the Ideal: Justice in a Diverse Society*. Princeton: Princeton University Press, 2019.

Gaus, Gerald, and Kevin Vallier. "The Roles of Religious Conviction in a Publicly Justified Polity: The Implications of Convergence, Asymmetry and Political Institutions." *Philosophy and Social Criticism* 35, nos. 1–2 (January/February): 51–76.

Lister, Andrew. "Public Justification and the Limits of State Action." *Politics, Philosophy and Economics* 9, no. 2 (May 2010): 151–75.

Lomasky, Loren E. *Persons, Rights, and the Moral Community*. New York: Oxford University Press, 1990.

Mill, John Stuart. *The Collected Works of John Stuart Mill*, vol. 2, *The Principles of Political Economy with Some of Their Applications to Social Philosophy*, edited by J. M. Robson. Toronto: University of Toronto Press, 1965.

———. *On Liberty*. In *Collected Works of John Stuart Mill*, vol. 18, *Essays on Politics and Society*, edited by J. M. Robson, 213–310. Abingdon, UK: Routledge, 2014.

Nozick, Robert. *Anarchy, State, and Utopia*. New York: Basic Books, 2013.

Quong, Jonathan. *Liberalism without Perfection*. Oxford: Oxford University Press, 2011.

———. "Liberalism without Perfection: Replies to Gaus, Colburn, Chan, and Bocchiola." *Philosophy and Public Issues* 2, no. 2 (2012): 51–79

Rawls, John. *Political Liberalism*, 2nd ed. New York: Columbia University Press, 2005.

———. *A Theory of Justice*, rev. ed. Cambridge, MA: Belknap Press, 1999.

Raz, Joseph. "Facing Diversity: The Case of Epistemic Abstinence." *Philosophy and Public Affairs* 19, no. 1 (Winter 1990): 3–46.

Skyrms, Brian. *Evolution of the Social Contract*. Cambridge: Cambridge University Press, 1996.

———. *The Stag Hunt and the Evolution of Social Structure*. Cambridge: Cambridge University Press, 2003.

Vallier, Kevin. "Against Public Reason Liberalism's Accessibility Requirement." *Journal of Moral Philosophy* 8, no. 3 (January 2011): 366–89.

———. "Convergence and Consensus in Public Reason." *Public Affairs Quarterly* 25, no. 4 (October 2011): 261–279.

———. "In Defence of Intelligible Reasons in Public Justification." *Philosophical Quarterly* 66, no. 264 (July 2016): 596–616.

———. *Liberal Politics and Public Faith: Beyond Separation*. New York: Routledge, 2014.

———. *Must Politics Be War? Restoring Our Trust in the Open Society*. New York: Oxford University Press, 2019.

Vallier, Kevin, and Fred D'Agostino. "Public Justification." In *Stanford Encyclopedia of Philosophy* (Spring 2014). https://plato.stanford.edu/archives/spr2014/entries/justification-public.

Wall, Steven. "Public Reason and Moral Authoritarianism." *Philosophical Quarterly* 63, no. 250 (January 2013): 160–69.

Wendt, Fabian. *Compromise, Peace and Public Justification: Political Morality Beyond Justice*. London: Palgrave Macmillan, 2016.

Journal of Ethics and Social Philosophy
VOL. 20, NO. 1 · JULY 2021
https://doi.org/10.26556/jesp.v20i1.1210

ELUSIVE REASONS AND THE MOTIVATIONAL CONSTRAINT

Benjamin Cohen Rossi

THE MOTIVATIONAL CONSTRAINT on normative reasons says that a consideration is a normative reason for an agent to act only if it is logically possible for the agent to act for that reason, or at least to be moved so to act.[1] The claim figures Zelig-like in philosophical debates about practical reasons: on hand, occasionally prominent, but never the focus of discussion. However, because it is entailed by a number of prominent views about normative reasons—including various forms of internalism and some views that closely connect reasons to good practical reasoning—its truth or falsehood has important implications.[2]

Mark Schroeder and Julia Markovits have recently criticized the motivational constraint on the grounds of "elusive reasons": reasons for some agent to act that are such that it is logically impossible both that they are normative reasons for that agent and the agent is moved to act for those reasons.[3] The type of elusive reason most discussed in the literature is what I call "blindspot reasons." Blindspot *propositions* are contingently true propositions that some agent cannot—"cannot" denoting logical impossibility—believe truly.[4] Some of these propositions seem to be reasons for action. In Schroeder's example, Nate likes surprise parties, but only if they are a genuine surprise; thus, the fact that there is a surprise party in the living room seems like a reason for Nate to go into the living room. This reason is also a blindspot proposition for Nate. He cannot truly believe that there is a surprise party in the living room since, if he did, it would

1 Most discussions of elusive reasons say that any reason for action requires the logical ability to *act* for that reason, but I will use the weakened version requiring mere *motivation* to act for that reason. The principal reason for this is that, on the weakened formulation, it is clear how the motivational constraint is logically entailed by some common claims about normative reasons.

2 See section 4 for a discussion of these views.

3 Schroeder, *Slaves of the Passions*; and Markovits, *Moral Reason*.

4 Sorenson, *Blindspots*.

not be a surprise. So, that fact is a blindspot *reason* for Nate. In general, a blindspot reason p is a normative reason for A to ϕ that is such that there is no possible world in which (a) A believes that p and (b) p is true. But two plausible assumptions make this kind of reason challenging for the motivational constraint: that normative reasons are facts or true propositions, and that being moved to act for a reason p requires believing that p.[5] Given these assumptions, blindspot reasons fail to satisfy the motivational constraint. If Nate is moved to go into the living room because there is a surprise party in the living room, then there is no longer a reason for him to go into the living room.

In response to this criticism, a number of philosophers have attempted to reconcile blindspot reasons with the motivational constraint. Neil Sinclair claims that given a certain plausible conception of "being moved to act for a reason," blindspot reasons present no difficulty for the motivational constraint.[6] Michael Ridge and Sean McKeever argue that, for any blindspot reason, there is a consideration *identical to it* that can satisfy the motivational constraint.[7] Another defense, parts of which can be found in Ridge and McKeever and Hille Paakkunainen, claims that, for every blindspot reason, there is a consideration that bears a certain relation to that reason such that on a plausible conception of the motivational constraint the blindspot reason satisfies it.[8] My aim in this paper is to show that these conciliatory strategies fail to overcome the challenge posed by elusive reasons. First, I examine each strategy and argue that it is unsuccessful on its own terms. Second, I adduce another type of elusive reason not heretofore discussed in the literature, and argue that these strategies also cannot make *this* kind of reason consistent with the motivational constraint. Finally, I defend the existence of this kind of reason against an important objection.

5 If reasons are worldly facts (see section 2.1), then we need a slightly different account of blindspot reasons. On this account, blindspot reasons are obtaining facts that *correspond* to, or perhaps *make true*, propositions that some agent cannot believe truly. Although I will not argue for the point here, I do not believe the assumption that normative reasons either are or correspond to true propositions is essential to the argument against the motivational constraint from blindspot reasons, although dropping that assumption would require modifying the account of blindspot reasons to some extent.

6 Sinclair, "On the Connection between Normative Reasons and the Possibility of Acting for Those Reasons."

7 McKeever and Ridge, "Elusive Reasons."

8 McKeever and Ridge, "Elusive Reasons"; and Paakkunainen, "Can There Be Government House Reasons for Action?" See section 2.2 for the formulation of Motivational Constraint*.

1. INTERPRETING "*A* IS MOVED TO ACT FOR THE REASON THAT *P*"

1.1. The Motivational Constraint and Motivating Reasons

One straightforward interpretation of the motivational constraint is:

> *Motivational Constraint*: If *p* is a reason for *A* to ϕ, then there is a logically possible world *w* in which *A* is moved to ϕ for the reason that *p*.

Clearly, the key notion in this formulation is "*A* is moved to ϕ for the reason that *p*." Traditionally, the locution has performed double duty, standing for the concept of *being moved for a motivating reason* and *being moved for a motivating reason that is also a normative reason*.[9] More precisely: sometimes the claim that "*A* is moved to ϕ for the reason that *p*" is true just in case

1. *A* believes that *p*,
2. *A* regards *p* as a reason to ϕ, and
3. the attitudes described in 1 and 2 nondeviantly cause *A* to be moved to ϕ.

Notice that, on this interpretation, *p* need not be a normative reason for *A* to ϕ. This interpretation is equivalent to the semi-technical locution, "A motivating reason for *A* to ϕ is *p*." Thus, I will call this the "Motivating Reason" interpretation of "*A* is moved to ϕ for the reason that *p*." Sometimes, however, this locution is true just in case

1. *p* is a normative reason for *A* to ϕ,
2. *A* believes that *p*,
3. *A* regards *p* as a reason to ϕ, and
4. the attitudes described in 2 and 3 nondeviantly cause *A* to be moved to ϕ.

On this interpretation, one of *A*'s *motivating* reasons for ϕ-ing, *p*, must also be a *normative reason* for *A* to ϕ.[10] Call this the *Motivating + Normative* (M + N) interpretation of "*A* is moved to ϕ for the reason that *p*."

Now, on the Motivating Reason interpretation, blindspot reasons are not counterexamples to Motivational Constraint. On this interpretation, Motiva-

9 Cf. Sinclair, "On the Connection between Normative Reasons and the Possibility of Acting for Those Reasons," 1218; McKeever and Ridge, "Elusive Reasons," 130–33.

10 Some formulations of the M + N interpretation include that *p* is true as an additional condition (see, e.g., Sinclair, "On the Connection between Normative Reasons and the Possibility of Acting for Those Reasons," 1219). I am assuming here that if it is true that *p* is a normative reason for *A* to ϕ in *C*, then *p* is true. See note 5 for further discussion.

tional Constraint would simply assert that the fact that *p* is a normative reason for *A* to ϕ entails that there is some possible world in which *A* is moved to ϕ for the reason that *p*, but *p* need not be a normative reason for *A* to ϕ in that world. But the tension between Motivational Constraint and blindspot reasons followed from the assumption that in the possible world in which *A* is moved to ϕ for the reason that *p*, *A* both believes that *p* and *p* is a normative reason to ϕ (hence, *p* is true; see note 10). If the latter assumption is dropped, then there is no difficulty reconciling blindspot reasons and Motivational Constraint. For example, the world in which Nate is motivated to go into the living room by the reason that there is a surprise party in the living room may be a world in which that consideration is not a normative reason for him to act; but this world still satisfies the description of the possible world described in the consequent of Motivational Constraint with the Motivating Reason interpretation of "*A* is moved to ϕ for the reason that *p*" plugged in.

However, it is implausible to interpret Motivational Constraint's *being moved to act for reasons* clause along the lines of the Motivating Reason interpretation. If we did, Motivational Constraint would probably be trivially true: *any* consideration *could*, in *some* set of circumstances, move an agent to act. The *point* of Motivational Constraint is to put a constraint on normative reasons: it is to assert that normative reasons, *as such*, must be capable of motivating. By contrast, Motivational Constraint on the Motivating Reason interpretation merely asserts that propositions that are normative reasons in the actual world can motivate in some possible world. Put another way, Motivational Constraint should be read as a *de dicto* claim: it says that, necessarily, a normative reason *as such* must be able to motivate. Motivational Constraint on the Motivating Reason interpretation is a *de re* claim: it says that if some proposition counts as a normative reason, then, necessarily, *that proposition* must be able to motivate.

Thus, we should opt for the M + N interpretation of "*A is moved to* ϕ for the reason that *p*" in the consequent of Motivational Constraint. Of course, if we opt for this interpretation, then we are immediately confronted with the problem of elusive reasons. On this interpretation, it is a *requirement* of the truth conditions for "*A is moved to* ϕ for the reason that *p*" that *p* is a normative reason for *A* to ϕ and *A* believes that *p*, and it is precisely this combination of conditions that seems to be ruled out in the case of blindspot reasons.[11]

11 One response to the problem of elusive reasons is to deny that they are reasons. Paakkunainen suggests two lines of argument in this vein: first, that apparently elusive reasons are not reasons for action, but reasons of another kind; and second, that our intuitions about cases of elusive reasons track other normative phenomena, such as other reasons that *are* accessible to the agent ("Can There Be Government House Reasons for Action?" 58). The only

1.2. Neil Sinclair's Strategy

Neil Sinclair's contribution to this debate is to suggest that there is another plausible interpretation—indeed, multiple possible interpretations—of "*A* is moved to ϕ for the reason that *p*." On his favored interpretation, "*A* is moved to ϕ for the reason that *p*" means that:

1. *p* is a normative reason for *A* to ϕ,
2. some agent, *X*, believes that *p* and regards *p* as a reason for *A* to ϕ (where it is possible that $X \neq A$), and
3. the attitudes described in 2 nondeviantly cause *A* to be moved to ϕ.[12]

The key point is that, on this view, "*A* is moved to ϕ for the reason that *p*" does *not* entail that *A* believes that *p* or regards *p* as a reason to ϕ; rather, it entails merely that *someone* holds these attitudes. Call this the *Proxy* M + N interpretation. The heart of Sinclair's argument for this interpretation is a modified version of Schroeder's case in which another person, LeTrain, believes that there is a surprise party in the living room and believes that this fact is a reason for Nate to go into the living room, but chooses to tell Nate only that there is a reason for him to go into the living room. As Nate trusts LeTrain, he acquires the belief that there is reason for him to go into the living room on the basis of LeTrain's testimony. This belief then moves him to go into the living room. Sinclair asserts that, in this case, Nate is moved to go into the living room for the reason that *there is a surprise party in the living room.*

How does Sinclair's proposed conception of *being moved to act for a reason* help resolve the problem of elusive reasons? If there is some logically possible world in which a reliable advisor is motivated by Nate's blindspot reason to advise Nate appropriately, and if in this world that advice nondeviantly causes Nate to be moved to go into the living room, then on the Proxy M + N interpretation of "*A* is moved to ϕ for the reason that *p*," in that world Nate is moved to go into the living room *for* his blindspot reason. If it is logically possible for Nate to

response I can give here is to point out that, on any of the most prominent accounts of normative reasons, elusive reasons plausibly count as such. For example, given that he enjoys surprise parties, the fact that there is a surprise party in the living room certainly seems to count in favor of Nate's going into the living room. It is also a fact that helps explain why going into the living room would promote one of his desires, or help realize something that is valuable for Nate, such as non-perverse enjoyment.

12 See Sinclair, "On the Connection between Normative Reasons and the Possibility of Acting for Those Reasons," 1221. Sinclair adds another condition: that the attitudes described in 2 and 3 are appropriately sensitive to the fact involved in 1. Nothing I will say hinges on this additional condition.

be moved to act for his blindspot reason, then Nate's blindspot reason satisfies Motivational Constraint. There seems to be nothing *logically inconsistent* about such an advisor playing this motivational role; so, assuming Proxy M + N, Nate's blindspot reason satisfies Motivational Constraint. If something analogous is true for every blindspot reason, then Sinclair has a perfectly general strategy for reconciling Motivational Constraint with blindspot reasons.

Sinclair's principal argument for the Proxy M + N interpretation is that Nate's act of going into the living room on the basis of LeTrain's testimony is dependent on a sequence of mental processes that can plausibly be considered an instance of non-defective reasoning in which the fact that there is a surprise party in the living room figures as a premise. The nature of the dependence of Nate's act on this reasoning is roughly as follows: LeTrain's belief that there is a surprise party in the living room and his belief that this fact is a reason for Nate are premises in a piece of practical reasoning motivating LeTrain to act; these beliefs are nondeviantly causally connected to Nate's action, and they are sensitive both to the truth and the "reasonhood" of that fact, where "sensitivity" is a matter of the truth of certain counterfactuals. Sinclair suggests that if Nate's action is dependent in this way upon an instance of non-defective reasoning in which his blindspot reason figures as a premise, this is sufficient to establish that Nate is moved to act for that reason.

Arguably, this line of reasoning has some counterintuitive consequences. Suppose that Nate and LeTrain, who are roommates, are out on the town the night before the surprise party, which is to take place the following morning. Sometime past midnight, LeTrain decides that it is time to get home; if they are out too late, Nate will oversleep and miss his party. Nate refuses. LeTrain knows that, in his drunken state, Nate will likely fall asleep as soon as he gets home—more precisely, as soon as he flops onto the sofa—but he needs to induce Nate to agree. So, LeTrain tricks him: he calls an Uber and, as it arrives, tells Nate they are heading to another bar, not home. Nate enthusiastically jumps into the Uber. Of course, the real destination is home. Given that getting a decent night's sleep is necessary for Nate to enjoy his surprise party, the fact that getting into the Uber will take Nate home is reason for him to get into the Uber. Yet Nate's *motivating* reason to get into the Uber is that doing so will take him to another bar. Since the propositions that *the Uber will take Nate home* and *the Uber will take Nate to a bar* are contraries, his motivating reason contradicts his normative reason. Yet Sinclair would have us believe that because LeTrain's reasoning involving the belief that the Uber will take Nate home nondeviantly causes Nate

to get into the Uber, Nate gets into the Uber for the reason that it will take him home. But this is at the very least a very odd way of talking.[13]

Sinclair also argues that if the fact that there is a surprise party in the living room is not the reason for which Nate is moved to go into the living room, then its role in bringing about that outcome must be understood either as a form of causal explanation or as an explanation in terms of motivating reasons. However, both of these options are unsatisfactory. The merely causal explanation cannot account for the role of this consideration in someone's reasoning, and the motivating reason explanation is not available because in order for some fact to be a motivating reason for some agent, that agent must believe that fact. Sinclair argues that this explanatory problem can easily be resolved if we hold that Nate is moved to act for this reason in the Proxy M + N sense, so his being moved to act for a normative reason does not require that he doxastically grasp that reason. However, another way of accounting for the blindspot reason's deliberative and motivational roles is to characterize it as a reason for which LeTrain is moved to act in the M + N sense. On this account, the fact that there is a surprise party in the living room is a reason for which LeTrain is moved to tell Nate that there is reason for Nate to go into the living room. This consideration plays a particular role in LeTrain's reasoning, so it is not a mere cause of his behavior; LeTrain believes this reason, so it can play the role of a motivating reason; and it is a normative reason for him to do something, namely, to tell Nate that there is a reason for him to go into the living room. On the basis of his belief that LeTrain is a reliable source, Nate acquires the belief that there is a reason for him to go into the living room—a belief that refers to his blindspot reason. This belief is his motivating reason to act accordingly. The story I have just told does not require us to claim that Nate was moved to act for his blindspot reason, yet it fully accounts for the role of his blindspot reason in reasoning and motivation.

1.3. The Theoretical Fitness of M + N and Proxy M + N Interpretations

Partly because "*A* is moved to ϕ for the reason that *p*" is a semi-technical locution, I very much doubt that my dispute with Sinclair can be resolved on the basis of intuitions: the choice between the M + N or Proxy M + N interpretations seems to be a matter of conceptual legislation rather than joint carving. The best way to proceed, then, is to consider the *roles* that the idea of being moved to act for a

13 This case is in some respects similar to a case discussed by McKeever and Ridge ("Elusive Reasons") involving a conservative who acts on the basis of a reliable liberal friend's advice to give to a cause that he would balk at if he knew more about it. I believe my case is stronger because it involves not just a normative reason that the agent would reject, but a motivating reason that formally contradicts the agent's normative reason.

reason play in philosophical theorizing, and then to consider which interpretation of that idea is better suited to play them. I will consider two important areas in which the concept of being moved to act for a reason plays an important theoretical role: theories of moral creditworthiness and theories of rational action. I will then argue that the M + N conception is better suited to these theoretical roles than the Proxy M + N interpretation.

One area where the concept of acting for a reason plays an important role is in theorizing about the conditions of the *creditworthiness* of actions, a broad evaluative category of which the *moral worth* of actions is a species. Roughly, to say that a person is creditworthy for some action is to say that the action reflects well on them from some point of view—prudential, moral, etc. In the moral domain in particular, it is a familiar idea that people can do the morally right thing without being morally creditworthy for it, so moral creditworthiness requires something *in addition* to doing the morally right thing. A standard move is to claim that, for moral creditworthiness, one must not only act rightly but must also "act for the morally right reasons"; as Kant put it, one must not merely act *in accordance with* one's moral reasons, but also *from* them.[14] Given that acting for one's moral reasons entails being moved to act for one's moral reasons, we could offer the following account of moral creditworthiness:

> *Moral Creditworthiness*: *A* is morally creditworthy for ϕ-ing iff *A* ϕs as a result of being moved to ϕ for the morally right reason(s).

The question is: What interpretation of "being moved to act for reasons" is the best candidate for cashing out the notion of being moved to act for the *morally right* reasons in Moral Creditworthiness? Suppose we opt for Sinclair's Proxy M + N interpretation. The problems for this suggestion emerge if we consider a modified Nate and LeTrain case. Suppose that there is an evil demon who will seriously injure someone if Nate is not surprised by a surprise party in the next thirty minutes, and LeTrain knows this. Luckily, there is a surprise party in the living room. LeTrain, regarding this fact about the demon's conditional intention as a reason for Nate to go into the living room, tells Nate there is reason for him to go into the living room. However, he does not say that there is a *moral* reason for him to do so. Nate goes into the living room as a result of being moved to do so by LeTrain's testimony. By the lights of the Proxy M + N interpretation, Nate is moved to go into the living room for the reason that the demon conditionally intends to injure someone. Thus, Nate has done the thing he morally ought to have

14 For discussion, see Arpaly, "Moral Worth"; Stroud, "Moral Worth and Rationality as Acting for Good Reasons"; Markovits, "Acting for the Right Reasons"; and Way, "Creditworthiness and Matching Principles."

done, and he did it because he was moved to act for the morally right reason. According to Moral Creditworthiness, Nate is therefore morally creditworthy for going into the living room—despite having no idea that anything morally significant was at stake in doing so! Something has gone awry.

One widely accepted way of cashing out the notion of acting for the morally right reasons is that one's motivating reasons must *match* one's moral reasons. This yields the following account of moral creditworthiness:

> *Matching Reasons*: *A* is morally creditworthy for ϕ-ing iff the motivating reasons for which *A* is moved to ϕ (and subsequently ϕs) are the moral reasons for *A* to ϕ.[15]

Notice that this account of creditworthiness uses the M + N concept of being moved to act for a reason. The account basically claims that one is morally creditworthy for an action only in case one is moved to act for a reason—in the M + N sense—that is a *moral* reason for one so to act.[16] Thus, in many discussions of moral creditworthiness, the operative conception of being moved to act for reasons is the M + N conception, *not* the Proxy M + N conception. Furthermore, if the idea of an agent's motivating reasons *matching* her normative reasons is key to interpreting "being moved to act for the morally right reasons" as a necessary condition of creditworthiness, then Sinclair's conception is arguably not suitable for the following reason. The Proxy M + N interpretation would count someone as being moved to act for a normative reason even when their motivating reasons do not match their normative reasons, and *a fortiori* do not match their *moral* reasons. In the Nate and LeTrain case, Nate's reason for being moved to act is something like that *there is a reason for me to act*, or perhaps that *LeTrain, a reliable and virtuous friend, told me there is a reason for me to act*. Arguably, neither fact is a normative reason for Nate to act. Sinclair argues that counting the former as a normative reason in its own right leads to an infinite regress, and Ridge and McKeever argue that facts about testimony are never themselves normative reasons.[17] Even if they are normative reasons for Nate to act, neither is the same reason as that *the demon conditionally intends to injure someone*. Thus, if Matching Reasons captures something crucial about the idea of being moved to act for the morally right reasons, it is not plausible to cash out the latter in terms of Sinclair's conception of being moved to act for a reason.

15 For discussion, see Markovits, "Acting for the Right Reasons"; Way, "Creditworthiness and Matching Principles."

16 This is a merely necessary condition, since for moral creditworthiness one has to actually *act* in addition to being *moved to act* for the right reason.

17 McKeever and Ridge, "Elusive Reasons," esp. 116–20.

Not everyone accepts that acting for the morally right reasons is sufficient for moral creditworthiness.[18] The details of the arguments against the sufficiency of the consequent of Moral Creditworthiness need not occupy us here. The important thing for our purposes is that if satisfying the consequent of Moral Creditworthiness is not sufficient for moral creditworthiness, this gives Sinclair an opening to claim that the concept of being moved to act for reasons and the concept described on the right side of the biconditional in Matching Reasons are distinct concepts, rather than the latter being a way of cashing out the former. In this way, Sinclair might claim that being moved to act for reasons can be understood along the lines of his conception without yielding counterintuitive results. We can call the resulting account Moral Creditworthiness*:

> *Moral Creditworthiness**: *A* is morally creditworthy for ϕ-ing iff (1) *A* is moved to ϕ (and subsequently ϕs) for the morally right reason(s) and (2) the motivating reasons for which *A* is moved to ϕ are the moral reasons for *A* to ϕ.[19]

This account would not yield that Nate is morally creditworthy if he acts on LeTrain's vague advice. Instead, it would say that Nate acts for the morally right reasons, but his *motivating* reasons do not match those reasons, so Nate is not morally creditworthy. Thus, while an account of moral creditworthiness using Sinclair's conception would not necessarily yield counterintuitive verdicts, amending the account so as to eliminate these verdicts *would* commit him to saying that someone might fulfill the condition of *being moved to act for the morally right reasons* even if her motivating reasons do not match her normative reasons. The matching requirement would be a *further* requirement on creditworthiness, rather than an *interpretation* of the concept of *being moved to act for the right reasons*.

However, this move is foreclosed for Sinclair by the idea that the moral creditworthiness of action must not depend on luck. I will presently argue that, for this reason, Sinclair's conception is unsuited to play the role in theorizing about

18 Cf. Way, "Creditworthiness and Matching Principles." It is unclear to me whether in making his case against the sufficiency of Matching Reasons, Way intends to show that acting for the morally right reasons is not sufficient for creditworthiness, or that his "Matching Principles" is a rival *interpretation* of "Acting for the morally right reasons." Way seems to suggest that he is open to rejecting the equation of moral creditworthiness and acting for the morally right reasons ("Creditworthiness and Matching Principles," 213n9).

19 Given that the case against the sufficiency of Moral Creditworthiness may turn on the claim that Matching Reasons is not sufficient for moral creditworthiness—thus, it assumes that Matching Reasons is an *interpretation* of "acting for the morally right reasons"—this revised account would probably require an additional condition, such as Way's Matching Principles.

moral creditworthiness *even if* it does not entail counterintuitive verdicts about cases.

One plausible desideratum of any account of moral creditworthiness is that one's moral creditworthiness for an action does not depend to a significant degree upon luck. As Jonathan Way puts it, "You are not creditworthy for a response if the way in which that response was motivated could quite easily have led you to do the wrong thing."[20] Given this "no-luck" constraint, and assuming that being moved to act for the morally right reasons is at least a *necessary* condition on moral creditworthiness, a conception of being moved to act for the right reasons must foreclose the possibility of being moved to act for the right reasons *accidentally*. Sinclair's conception does not do this. On his view, a person may be moved to act for the morally right reasons if there happens to be a reliable and virtuous advisor who, grasping those reasons, is able to advise the agent to act without describing the reasons that support acting this way and chooses to do so. Surely, the advisee could quite easily have done the wrong thing: she might not have followed the advice properly, and the advisor might not have been on hand, might not have been able to advise properly, or might not have chosen to give advice.[21] But then this conception of being moved to act for reasons makes moral creditworthiness depend on luck, in violation of the no-luck constraint.

Another area of philosophical inquiry in which the idea of being moved to act for reasons plays an important role is theories of "rational agency," the conditions under which it is rational for agents to act. The relationship between rationality and reasons is a matter of significant debate, but suppose that something like the following is true:

> *Rationality and Reasons* (RR): *A* ϕ-ed rationally only if *A* had reasons that made it reasonable for *A* to ϕ and *A* was moved to ϕ "for" some (sub-)set of those reasons.[22]

The concept of being moved to act for a reason appears in this claim as a necessary condition on the *ex post* rationality of an action.[23] Let us consider how Sin-

20 Way, "Creditworthiness and Matching Principles," 218.

21 As I mentioned in note 12, Sinclair includes in Proxy M + N the condition that X's attitudes are appropriately sensitive to, because counterfactually dependent upon, the fact that *p* is a normative reason for *A* to ϕ. This condition rules out the possibility of the advisor making certain kinds of mistakes, such as being mistaken about whether *p* is a genuine normative reason for *A*. It does not, however, rule out the possibilities described here. Thus, *A*'s being moved to ϕ as a result of the advisor's testimony might still be objectionably lucky.

22 This is Comesaña and McGrath's formulation of the relation between rationality and reasons ("Having False Reasons," 62).

23 For the distinction between *ex post* and *ex ante* judgments of rationality, see Comesaña and

clair's conception of being moved to act for a reason fares in this role. Since RR offers only a necessary condition of rational action, we do not encounter quite the same problems as we did in the case of Moral Creditworthiness.[24] Yet the same basic point applies here: even if satisfying RR does not make an action rational, it is obscure exactly how the fact that *someone else* grasped the normative significance of, and was motivated by, *A*'s reasons is supposed to help satisfy a condition of the rationality of *A*'s action. Consider again the case involving LeTrain and Nate having a night on the town. LeTrain has told Nate that the Uber will take them to the next bar, when in fact, it will take them home. Suppose that the consideration that the Uber will take him to the next bar is not a reason for Nate to get into the Uber because it is false; furthermore, even if it were true, it would not be a reason to get into the Uber because going to the next bar would not fulfill any of Nate's reflective non-perverse preferences and is not morally required. On RR together with the M+N conception of acting for reasons, Nate's getting into the Uber *for* this reason should turn out to be irrational, as it is not based on—in the sense of "based on" meaning "*motivated* by"—any real or apparent normative reason. Rather, it is based on a false proposition that, even if true, would not provide a reason for Nate to act. And this *seems* to be the most intuitive verdict on this case: Why, after all, should we credit Nate for acting *rationally* when, for all he believed, his action was not recommended even by any consideration that *would* be a normative reason to act if it were true? By contrast, on Sinclair's conception, Nate's getting into the Uber would not turn out to be irrational even if none of his beliefs genuinely supported acting in that way, since he would nevertheless be moved to act for the normative reason that the Uber will get him home. Again, this just does not seem right. It is not *Nate* who has the beliefs in the light of which getting into the Uber is rational; it is LeTrain.

The general point is this: whichever way we wish to spell out the rationality of action, to say that an action is "rational" is to say something about the relation between the action and the *agent's view* about the circumstances under which she acted.[25] But using Sinclair's conception as a way of cashing out the concept of being moved to act *in light of* some beliefs would commit us to saying that, in

---

McGrath, "Having False Reasons," 61.

24 For the different view that rational action is action for good reasons or *apparent* reasons, see Parfit, "Rationality and Reasons." For the view that rational action is action *on* good reasons but not *for* them, see Arpaly, *Unprincipled Virtue*.

25 Stroud makes precisely this point to criticize Arpaly's account of rationality in *Unprincipled Virtue*: "I myself am very loath to accept [*contra* Arpaly] … that an agent's beliefs about and evaluation of the reasons in favor of options *A* and *B* are irrelevant to the rationality of her opting for (say) *B*" ("Moral Worth and Rationality as Acting for Good Reasons," 454).

some cases, the beliefs that make the action rational, and in the light of which the agent was moved to act, do not in any way represent the agent's view about the circumstances under which she acted. So, to call an action "rational," on this conception, would not necessarily say anything about the relation between the action and the agent's view about the circumstances under which she was moved to act. At that point, it is tempting to say that we would be no longer talking about *rationality*, or at least not about the concept of *rationality* with which the theory of rational agency is typically concerned.

I have shown that Sinclair's conception of acting for a reason is not best suited to two important theoretical roles for which the concept of being moved to act for a reason is used. Of course, it is open to Sinclair to grant this point and reply that there is no conclusive reason to think that we must use the *same* conception of being moved to act for a reason in every theoretical area. This is true, and I do not intend these considerations of theoretical fit as a knockdown argument. Still, if the M + N conception is operative in all other areas in which the concept *being moved to act for a reason* figures, or *at least* if the Proxy M + N conception is *never* operative in these areas, then the burden of proof is on Sinclair to show that the M + N conception is not operative, or at least that the Proxy M + N conception *is* operative, in Motivational Constraint.

2. A DIFFERENT REASON?

2.1. Identical Reasons

Ridge and McKeever propose a different solution to the problem of elusive reasons. Like Sinclair, they claim that the logical possibility of a reliable advisor playing a certain role in causing Nate to be moved to go into the living room can help blindspot reasons satisfy Motivational Constraint. However, according to them, in the Surprise Party case involving LeTrain there is a normative reason for which Nate is moved to act in the M + N sense; hence, *contra* Sinclair, Nate's motivating reason is also *a* normative reason. Furthermore, they argue that, assuming reasons are a certain kind of coarse-grained fact, this reason—the reason for which Nate is moved to act in the M + N sense—is *identical to* Nate's blindspot reason. Hence, Nate's blindspot reason satisfies Motivational Constraint after all.

Some of what follows hinges on the distinction between *facts* understood as *states of affairs*—complexes of concrete individuals, objects, and properties *that obtain*—and *true propositions* understood as structured complexes of concepts *that are true*. Many share the view that reasons are facts rather than true propositions, although there appears to be nothing approaching a consensus on

the matter.[26] I will not assume that reasons are either facts or true propositions for the purposes of this paper; my only assumption is that the object of belief is a proposition, not a fact. Except where noted, when I use the word "fact" to describe or refer to a reason for action, this can be taken to mean either a true proposition or an obtaining state of affairs. I will use the term "worldly fact" to designate *only* obtaining states of affairs.

According to Ridge and McKeever, in the Nate and LeTrain case one normative reason for Nate to go into the living room is the fact that *his so doing will promote the worthy end(s) that his friend's advice is actually tracking.* I will call this the "R&M reason." I do not deny that this fact is a normative reason for Nate to go into the living room or that it is *not* a blindspot reason for Nate; from this it follows that it could be a reason for which Nate is moved to act in the M + N sense. While thus far we have been referring to the fact that *there is a surprise party in the living room* as Nate's blindspot reason, Ridge and McKeever actually point to *another* blindspot reason in the Surprise Party case: that *Nate will be pleasantly surprised if he goes into the living room*. For them, this reason bears the relation of "Russellian identity" to the R&M reason. As I define this notion,

> *Russellian Identity*: Two reasons *p* and *q* are Russellian identical iff
> 1. *p* and *q* are worldly facts and *p* is identical to *q*, or
> 2. *p* and *q* are propositions and *p* and *q* correspond to or are made true by the same worldly fact.

Understood as worldly facts, *going into the living room will promote the worthy end(s) that Nate's friend's advice is actually tracking* is the same worldly fact as *Nate will be pleasantly surprised if he goes into the living room*; understood as true propositions, they correspond to the same worldly fact. So, Ridge and McKeever argue that the R&M reason is Russellian identical to one of Nate's elusive reasons.[27]

26 Sometimes the distinction is obscured by the equivocal use of the term "fact." For example, T.M. Scanlon says that reasons are facts, but by "facts" he means "reflections of true thoughts." On this understanding of "fact," Scanlon is closer to those who believe that reasons are true propositions. Dancy says that reasons are "facts, or, better, states of affairs," but by "facts" he means what I mean by "worldly facts," with the important qualification that he may think some states of affairs *that do not obtain* are reasons. See Scanlon, *Being Realistic about Reasons*; and Dancy, *Practical Reality*, esp. 116.

27 What about the reason that is the target of Sinclair's discussion, that *there is a surprise party in the living room*? McKeever and Ridge could plausibly identify a reason similar to the R&M reason that is Russellian identical to *that* elusive reason: for example, that *there is something in the living room that explains why Nate's going into the living room will promote the worthy end(s) that his friend's advice is actually tracking.* If what I say below is correct, then McKeever

Although the relation of Russellian identity can hold between worldly facts or propositions, it is actually important to Ridge and McKeever's argument that reasons are worldly facts *rather than* true propositions. On this assumption, any two *Russellian* identical reasons are identical reasons. So, if reasons are worldly facts, then the R&M reason and one of Nate's elusive reasons—that *Nate will be pleasantly surprised if he goes into the living room*—are identical *reasons*. Ridge and McKeever argue that if these reasons are identical then when Nate is moved to act for the reason that *going into the living room will promote the worthy end(s) that Nate's friend's advice is actually tracking*, it follows that he is moved to act for the reason that *he will be pleasantly surprised if he goes into the living room*. If that is true, then the latter satisfies Motivational Constraint: Nate's being moved to act for the R&M reason *just is* his being moved to act for the blindspot reason. So, on the assumption that reasons are worldly facts, Ridge and McKeever appear to have a plausible strategy for reconciling Motivational Constraint with blindspot reasons. If, on the other hand, reasons are true propositions, then Ridge and McKeever's solution fails. If reasons are true propositions, then even if the R&M reason and the elusive reason are Russellian identical, they are not identical *reasons*. Thus, on the view that reasons are true propositions, the fact that one is moved to act for the R&M reason does not entail the fact that one is moved to act for the blindspot reason.

In fact, I believe that, even assuming that reasons are worldly facts and the R&M reason is the same worldly fact as the elusive reason, it does not follow that when Nate is moved to act for the R&M reason, he is moved to act for the elusive reason. Recall that the M + N interpretation requires that the agent possess two propositional attitudes: the belief that <*p*> and the attitude of regarding <*p*> to be a reason (<*p*> is the proposition that *p*). The fact that *being moved to act for the reason that p* ought to be analyzed in terms of these two propositional attitudes plus their causal role in bringing about an action, along with the fact that propositions are generally taken to be the referents of *that*-clauses, suggests that a *proposition* is the referent of the clause "the reason that" in the statement "*A* is moved to ϕ for the reason that *p*." In claiming this, I am not simply insisting, *contra* Ridge and McKeever, that normative reasons are propositions and not worldly facts. From the claim that "*p*" in the statement "*A* is moved to ϕ for the reason that *p*" is the proposition <*p*>, it does not follow that the term "reason" exclusively refers to propositions. Consider the fact that the referent of "the belief that" is a proposition, yet this does not entail that the term "belief" exclusive-

and Ridge's strategy for solving the problem of elusive reasons fails for *any* elusive reason and not just the one that is the target of their discussion.

ly refers to a proposition.[28] Now, although the proposition <he will be pleasantly surprised if he goes into the living room> may be *Russellian* identical to the proposition <going into the living room will promote the worthy end(s) that my friend's advice is actually tracking>, they are not identical propositions: they clearly consist of different concepts. It follows that the fact that one is moved to act for the reason that <he will be pleasantly surprised if he goes into the living room> neither entails, nor is entailed by the fact that one is moved to act for the reason that <going into the living room will promote the worthy end(s) that my friend's advice is actually tracking>. Therefore, even assuming that reasons are worldly facts, establishing that the R&M reason and Nate's blindspot reason are the same worldly facts is not sufficient to establish that being moved to act for the R&M reason entails being moved to act for the blindspot reason. Thus, Ridge and McKeever's strategy does not ensure that Nate's blindspot reason satisfies Motivational Constraint. *A fortiori*, it fails as a general strategy for reconciling blindspot reasons with Motivational Constraint.

It may be objected that my argument against Ridge and McKeever's strategy depends upon the particular wording I have chosen for Motivational Constraint. That wording includes the phrase "*A* is moved to ϕ for the reason that *p*," and in that phrase *p* plausibly refers to a proposition. But suppose we reworded Motivational Constraint as the claim that if *p* is a reason for *A* to ϕ, then there is a logically possible world *w* in which *A* is moved to ϕ *because p.* There, *p* plausibly refers to a worldly fact.[29] My response to this is that "because *p*" is ambiguous as between a purely *causal* reading and a *motivational* reading. On the causal reading, *p* refers to a worldly fact: we might say, for example, that I went to bed early because I took melatonin, which made me feel sleepy. But "because *p*" could also have a motivational reading, according to which *A* is moved to ϕ because *A* believed that *p* and regarded *p* as a reason to ϕ, and these attitudes nondeviantly caused him to be moved to ϕ. On this reading, *p* refers to a proposition. For example, I might say that Lois Lane climbed to the rooftop of the building because Superman was there. Even if that is true, it might yet be false that Lois Lane climbed to the rooftop of the building because Clark Kent was there. This suggests that, in this use of the locution "because *p*," *Superman is there* is the propositional object of some attitudes in the light of which, and because of

28 Nevertheless, it does suggest that, like "belief," which sometimes refers to an attitude and sometimes to the propositional object of that attitude, "reason" must at least *sometimes* refer to propositions, even if normative reasons—reasons that count in favor of acting—are all worldly facts. This yields a disjunctive view of reasons in some ways similar to that proposed by Hornsby, "A Disjunctive Conception of Acting for Reasons."

29 Thanks to Sean McKeever for raising this objection in correspondence.

which, Lois Lane performed the action. And it seems to me that everyone in this debate, including Ridge and McKeever, interpret Motivational Constraint using the *motivational* reading, regardless of whether they use the "for the reason that *p*" or "because *p*" locutions. If that is true, then the suggested rewording does not save Ridge and McKeever's strategy.

2.2. The Derivative Reasons Strategy

Another strategy for reconciling Motivational Constraint with blindspot reasons claims that a *plausible* conception of the Motivational Constraint allows a blindspot reason to satisfy it so long as the reason is *suitably related* to some other reason for which that agent can be moved to act.[30] In other words, this strategy hinges on arguing for a revised version of the Motivational Constraint, which we can call Motivational Constraint*:

> *Motivational Constraint**: If *p* is a reason for *A* to ϕ, then there is a logically possible world *w* in which
>
> 1. *A* is moved to ϕ for the reason that *p*, or
> 2. *A* is moved to ϕ for a reason *q* that bears relation(s) R to reason *p*.

The strategy then argues that *derivativeness* is a relation that satisfies the second disjunct of Motivational Constraint*. To illustrate this relation using Paakkunainen's example: suppose ϕ-ing would destroy the only crop-yielding field in a village, and destroying the field would lead to much suffering.[31] Both facts seem to be decisive reasons against ϕ-ing, but the latter seems more fundamental than the former in that the former is a reason not to ϕ *because* it stands in some important relation to the latter. Here is one possible formulation of this derivativeness relation:

> *Derivative Reason*: A reason *p* is derivative of a reason *q* just in case *p* is a reason in virtue of standing in some important relation to *q*.[32]

30 Aspects of this approach can be found in McKeever and Ridge, "Elusive Reasons"; and Paakkunainen, "Can There Be Government House Reasons for Action?" However, I do not want to attribute the strategy to either for a couple of reasons. McKeever and Ridge do not clearly invoke the relation of derivativeness; they use the word "parasitism," and it is unclear to me whether they mean this word to pick out what I mean by "derivativeness." And Paakkunainen's principal response to *elusive reasons* is to quickly cast doubt on the reality of such reasons. She invokes the idea of derivative reasons to handle another set of putative reasons that make trouble for her Deliberative Constraint. So, neither of them quite makes the argument I am laying out here.

31 Paakkunainen, "Can There Be Government House Reasons for Action?" 87.

32 This formulation can be found in Nair, "How Do Reasons Accrue?" 64–65.

The strategy argues that the second disjunct of Motivational Constraint* is satisfied if either *q* is derivative of *p* or *p* is derivative of *q*. For example, plausibly, the non-blindspot reason that *Nate will be pleased if he goes into the living room* is derivative of the blindspot reason that *there is a surprise party in the living room*, since the former is a reason in virtue of *being explained by* the latter.[33] So, if we accept Motivational Constraint* and that for every blindspot reason there is either (a) a non-blindspot reason for which the agent can be moved to act and from which the blindspot reason derives or (b) a non-blindspot reason for which the agent can be moved to act that is derivative of the blindspot reason, then we seem to have a way of reconciling blindspot reasons with Motivational Constraint*. The argument hinges on the claims that (1) the Motivational Constraint* is a plausible conception of the motivational constraint and (2) derivativeness satisfies the second disjunct of Motivational Constraint*. It is to an evaluation of these two claims that I presently turn.

2.3. The Argument for Motivational Constraint and Derivativeness as a Satisfier*

The derivative reasons strategy raises the fundamental question of whether Motivational Constraint*—and in particular, its second disjunct—is itself a plausible conception of the motivational constraint. On its face, it would seem that taking the latter seriously would rule out the second disjunct of Motivational Constraint*: if the motivational constraint is the claim that the existence of a reason implies the possibility of being moved to act for *that* reason, then the possibility of being moved to act for *some other reason* just does not cut it. However, a proponent of Motivational Constraint* has at least one argument in its favor. The argument is that an agent ought to be able to satisfy the motivational constraint with respect to a blindspot reason *by* being able to be moved to act for some non-blindspot reason so long as the agent's *inability* to act for a blindspot reason does not entail that she is unable to appropriately respond to anything of normative significance in her circumstances. Admittedly, this talk of "normative significance" is a bit wooly, so let me try to make precise—if not less metaphorical—one thing we might mean by it.

It is a familiar idea in ethics that reasons for action not only *count in favor of* acting, but that they do so with a certain *strength* or *weight*. We might represent these weights as natural numbers, so that any reason for action can be represented as a four-place relation between a fact *p*, an agent *A*, an action ϕ, and a natural

33 We might argue, too, that *Nate will be pleasantly surprised if he goes into the living room* is derivative of the R&M reason in virtue of helping to explain it (*Nate will be pleasantly surprised if he goes into the living room* is a reason because it helps to explain the fact that *going into the living room will promote the worthy end that LeTrain's advice is actually tracking.*)

number n: $R(p, A, \phi, n)$. In many cases, where p is a reason to ϕ with strength m, and q is a reason to ϕ with strength n, p and q support ϕ-ing with a strength of $M + N$; the accrual of these reasons is an increasing function of the weights of the individual reasons.[34] For example, that *Mark promised to go to dinner at Rick's with Mel* and *Rick's is Mark's favorite restaurant* seem to be two reasons for Mark to go to dinner at Rick's whose accrual supports that action more strongly than either of the reasons individually. I will call reasons that accrue additively in this way "independent" reasons.[35] However, there are some reasons that do not accrue additively. For example, in the LeTrain case, the accrual of the reasons *going into the living room will promote the worthy end(s) that Nate's friend's advice is actually tracking* and *there is a surprise party in the living room* is not additive: the accrual of both reasons does not count in favor of going into the living room more strongly than either of these reasons individually. Similarly, the reasons that *ϕ-ing would destroy the only crop-yielding field in a village* and *destroying the field would lead to much suffering* seem not to accrue additively. Notice, too, that, while in both examples the reasons do not accrue additively, their accrual also does not have *less* strength than either reason individually.[36] We might therefore roughly characterize such pairs of reasons as lacking in *distinct* weight or strength. I will call such reasons *overlapping* reasons: a reason p to ϕ with strength m and a reason q to ϕ with strength n are *overlapping* just in case $m = n$ and the accrual of p and q support ϕ-ing with strength m.[37] Moreover, there are often *more than two* reasons that overlap one another. For example, that *there is a surprise party in the living room*, that *Nate will be pleasantly surprised if he goes into the living room*, that *Nate will be pleased if he goes into the living room*, and that *going into the living room will promote the worthy end(s) that Nate's advice is actually tracking* all seem to be reasons for Nate to act that do not have distinct strength, and so overlap. I will call any group of overlapping reasons consisting of two or more reasons a "cluster" of overlapping reasons.

Now, one proposal for cashing out the idea of "normative significance" in the argument for Motivational Constraint* is in terms of non-overlapping *clusters* of overlapping reasons.[38] Intuitively, the idea is that all features of an agent's cir-

34 Nair, "How Do Reasons Accrue?" 63.

35 Nair, "How Do Reasons Accrue?" 63.

36 To use Nair's example, that *it's hot outside* and *it's raining outside* may together support not going for a run with less strength than either of these reasons individually ("How Do Reasons Accrue?" 59). Thus, not all nonindependent reasons are overlapping.

37 See Nair, "How Do Reasons Accrue?"; and Lord and Maguire, "An Opinionated Guide to the Weight of Reasons," esp. 12.

38 If S and A are two clusters of overlapping reasons, S and A are *nonoverlapping* just in case

cumstances that are normatively significant for her are captured by the set of *distinctly weighted*, or non-overlapping, reasons for her to act.[39] This idea can be expressed in terms of clusters of overlapping reasons: an agent is able to be moved to act on every *normatively significant* feature of her circumstances if she is able to be moved to act for at least one member of all the non-overlapping clusters of her overlapping reasons.[40] So, the argument for Motivational Constraint* is as follows. So long as some reason *p* is a member of a cluster of overlapping reasons, and an agent is able to be moved to act for one of these members, she is able to be moved to act on that normatively significant feature of her circumstances captured by *p*. And if she can do this, she ought to satisfy the motivational constraint with respect to *p*.

Having established Motivational Constraint* with this argument, the derivative reasons strategy can invoke the notion of overlapping reasons to help clinch the argument that all blindspot reasons satisfy Motivational Constraint*. One premise of that argument that has so far not been supported is the claim that *derivativeness* is a relation that satisfies the second disjunct of Motivational Constraint*. The argument just offered for Motivational Constraint* implies that one relation that does satisfy its second disjunct is the relation of "overlapping-ness." So, the derivative reasons strategy can now claim that whenever two reasons are related by *derivativeness* they are *overlapping*, thus substantiating the claim that the former satisfies the second disjunct of Motivational Constraint*. In general, if one reason derives its normative force from another, then it will have no *independent* normative strength or weight. This seems to be borne out in our analysis of the Surprise Party case. We said that one non-blindspot reason for Nate to go into the living room is that *Nate will be pleased if he goes into the living room*, and this reason is *derivative of* the elusive reason that *there is a surprise party in the living room.* Intuitively, both reasons support going into the living room with the same strength or weight, but their accrual seems to support going into the living room no more and no less than either one individually. So, they are overlapping. Thus, if Motivational Constraint* is correct; and if for every blindspot reason there is either (a) a non-blindspot reason for which the agent can act and from

---

there is no member of *S* that is also a member of *A* and vice versa.

39 For ease of exposition, this statement excludes instrumental reasons, which do overlap but are not reasons to perform *the same act* as the reasons that overlap with them. See Nair, "How Do Reasons Accrue?" 60.

40 Perhaps normatively significant features include facts that are not themselves reasons but affect the strength of reasons, which we call *modifiers*, or their instantiations, which we call *defeaters* and *enablers*. For an overview of these concepts, see Lord and Maguire, "An Opinionated Guide to the Weight of Reasons," 11.

which the blindspot reason derives or (b) a non-blindspot reason for which the agent can act and derivative of the blindspot reason; and if all reasons related by *derivativeness* are overlapping, then all blindspot reasons satisfy the second disjunct of Motivational Constraint*.

*2.4. Evaluating the Argument for Motivational Constraint**

The key idea motivating the argument for Motivational Constraint* is that all the features of an agent's circumstances that are normatively significant for her are captured by the set of *distinctly weighted*, or non-overlapping, reasons for her to act. For example, that *Mark promised to go to dinner at Rick's with Mel* and *Rick's is Mark's favorite restaurant* are non-overlapping reasons for Mark to go to dinner at Rick's, and, intuitively, they are normatively significant for Mark in a distinct way. But this is not the only way to individuate normatively significant features. For many philosophers, reasons are closely connected to explanation: for example, on the two most prominent non-fundamentalist views about reasons, reasons are facts that explain how actions promote desires *or* why actions have some other normative status.[41] Let us assume the latter view for ease of exposition. Someone attracted to this kind of view will, I think, naturally understand the normatively significant features of an agent's circumstances in terms of the facts that explain how actions promote valuable states of affairs, and will claim that only an agent who grasps the *complete* explanations of how actions promote valuable states of affairs will grasp all of the normatively significant features of her circumstances. I will call this the "explanatory" notion of normative significance, as opposed to the "non-overlapping reasons" notion of normative significance. But adopting such a conception of normatively significant features will require accepting that an agent does not always grasp all of the normatively significant features of her circumstances by grasping at least one member of all the non-overlapping clusters of her overlapping reasons. In many cases, individual reasons will only be partial explanations: for example, that *Nate likes surprise parties* only partially explains why Nate will be pleased if he goes into the living room. That *Nate likes surprise parties* and that *there is a surprise party in the living room* are overlapping reasons, but they make distinct explanatory contributions to the fact that *Nate will be pleased if he goes into the living room*. As this example suggests, even overlapping reasons can constitute distinct parts of a complete explanation of the target fact. Similarly, the fact that *by ϕ-ing one will destroy the only crop-yielding field in the village* overlaps with the fact that *destroying the only*

41 For examples, see Markovits, *Moral Reason*; Broome, *Rationality through Reasoning*; Schroeder, *Slaves of the Passions*; McKeever and Ridge, *Principled Ethics*; and Finlay, "The Reasons that Matter."

crop-yielding field in the village would lead to much suffering, which itself overlaps with the fact that *ϕ-ing will lead to much suffering*. But on theories of reasons that point up their explanatory role, the first two facts are reasons not to *ϕ* in virtue of being distinct parts of an explanation of the third fact. And since they are distinct parts of that explanation, on this view they have distinct normative significance. Thus, if we adopt the explanatory conception of normative significance, then Motivational Constraint* is not supported by the argument from normative significance.

If the argument from normative significance does not support Motivational Constraint*, then in the absence of another compelling argument for Motivational Constraint*, we ought to revert back to the conception of the motivational constraint spelled out in Motivational Constraint. So, whether or not the argument from normative significance supports Motivational Constraint* depends upon our choice of conceptions of normative significance. I have not *defended* the explanatory conception over the non-overlapping reasons conception; instead, I have simply offered it as a plausible alternative, and I suspect that it is not the only one. Given this, it seems that the available arguments are insufficient to support the move from Motivational Constraint to Motivational Constraint*. I have already argued in this paper that blindspot reasons do not satisfy Motivational Constraint. Thus, the available arguments are insufficient to vindicate this strategy for reconciling blindspot reasons with the Motivational Constraint.

3. Another Elusive Reason

So far, philosophers have only discussed *blindspot* reasons in print. However, there is clearly logical space available for other kinds of elusive reason. Recall Motivational Constraint:

> *Motivational Constraint*: If *p* is a reason for *A* to *ϕ* in circumstances *C*, then there is a logically possible world *w* in which (a) circumstances *C* obtain and (b) *A* is moved to *ϕ* for the reason that *p*.

Now recall the Motivating + Normative (M + N) interpretation of "*A* is moved to *ϕ* for the reason that *p*":

1. *p* is a normative reason for *A* to *ϕ* in *C*,
2. *A* believes that *p*,
3. *A* regards *p* as a reason to *ϕ*, and
4. the attitudes described in 2 and 3 nondeviantly cause *A* to be moved to *ϕ*.

Given Motivational Constraint and the M+N interpretation, a reason could count as *elusive* so long as it is logically impossible for at least two of the conditions of the Motivational Constraint or the M+N conception to jointly obtain. Clearly, that leaves sizeable logical space for elusive reasons. In this section, I will explore only a small corner of that space. We have seen that a *blindspot* reason *p* is such that, given the factivity assumption, it is logically impossible that it is a normative reason for *A* to ϕ and *A* believes that *p*. Another kind of elusive reason is a reason *p* such that it is impossible that *p* is a normative reason for *A* to ϕ and *A* regards *p* as a reason to ϕ. Here is an example of such a reason. Suppose I like to kick cans down the road, but only if I do not take myself to have any reason for doing so. Plausibly, that I like to kick cans down the road is a reason for me to kick *this* can. However, because as soon as I regard a consideration as a reason for me to kick *this* can down the road, that consideration is no longer a reason for me to do so, the fact that I like to kick cans down the road is a reason for me to act that is such that it cannot be a reason for me to act *and* I regard it as a reason for me to act. Therefore, it is an elusive reason.

One might object that if I only like to kick cans down the road for no reason, it is impossible for me to do what I like to do *intentionally*, and this might suggest that the fact that I like to kick cans down the road is not a reason for me to do so. Let us assume *arguendo* one of the key premises of this objection, that *p* is a reason for *A* to ϕ only if *A* is able to ϕ intentionally. The other key premise is that *A* is able to intentionally ϕ only if *A* regards some consideration as a reason for ϕ-ing. This seems like a plausible principle. But if I am assembling some IKEA furniture and I select, from among a small heap of identical screws, one particular screw, I may not pick up *that* screw for any reason *because* I may not regard any consideration as a reason for picking up *that* screw.[42] Nevertheless, picking up *that* screw is surely an action I perform intentionally. Therefore, the principle that intentional action requires acting for reasons is false, at least on one interpretation of that principle.

In fact, I believe there is a plausible interpretation of this principle, but it does not support the objection. To see this, we must distinguish between action-universals (or *act-types*) and action-particulars.[43] An action-particular is a

42 One objection to this claim is that an agent can regard the same consideration as a reason to pick up any particular screw: namely, that my furniture requires a screw. I concede that this is a possibility. My claim is only that, in order to be properly motivated to pick up *that* screw, an agent *need not* regard any consideration as a reason to pick up *that* screw: she need only regard some consideration as a reason to pick up *some* screw. Thanks to an anonymous reviewer for the *Journal of Ethics and Social Philosophy* for this objection.

43 In some respects, in this discussion I follow Peter van Inwagen's discussion of the distinc-

concrete event that satisfies whatever conditions are necessary and sufficient for intentional action. Such actions are typically denoted by sentences like "Jones shot Smith," "Anna turned on the lights," and "Sully landed the plane." On the other hand, an action-universal is a *type* of intentional action; we denote such actions using phrases like, "mowing the grass," "killing a president," and "driving a car." Just as the proposition *the man mowed the grass* can be made true by any number of arrangements of concrete particulars, action-universals like *mowing the grass* can be instantiated by many different action-particulars. For example, the action-universal *mowing the grass* can be instantiated by many different particular instances of grass mowing. Now, a person may regard some consideration as a reason for an action-universal, but not some action-particular that is one of its (i.e., the action-universal's) instantiations. I may have a reason to mow the grass (because I want a tidy lawn), but not to mow the grass *right now* (because it is raining, and my John Deere would turn the lawn into a quagmire). This distinction yields two theses:

1. *A* is able to intentionally instantiate an action-universal only if *A* regards some consideration as a reason for its instantiation (i.e., the action-particular).
2. *A* is able to intentionally instantiate an action-universal only if *A* regards some consideration as a reason for the action-universal.

The lesson of the IKEA furniture story is that 2 may be true and 1 is false. I surely regard myself as having a reason to pick up a screw, an action-universal (for example, the instructions call for a screw at this point). It *might* be true that in order to intentionally pick up *this* screw I must regard some consideration as a reason for the corresponding action-universal. So, my intentionally picking up *this* screw may satisfy the consequent of 2, since I do regard some consideration as a reason for the action-universal. Still, *contra* 1, I need not regard some consideration as a reason for picking up *this* screw in order to do so intentionally. If we apply these points to the kicking-the-can case, we can say that I like to *kick cans down the road* (an action-universal). Furthermore, that I regard this or some other fact as a reason for me to kick cans down the road may underwrite the possibility of intentionally instantiating this universal by kicking *this* can down the road. But, in this case, I can instantiate the action-universal of kicking a can down the road only if I do not regard any consideration as a reason for an action-particular that instantiates this action-universal. Thus, if I intentionally kick

---

tion between responsibility for event-particulars and event-universals ("Ability and Responsibility").

this can down the road, I may satisfy the consequent of 2, but I cannot satisfy the consequent of 1. Since 1 is false, this is no problem for the case.

We can now consider whether any of the strategies discussed in this paper could reconcile Motivational Constraint with elusive reasons like my reason to kick this can down the road. Sinclair's and Ridge and McKeever's solutions both rely in different ways on the possibility of a reliable advisor who *gives* the agent at least a *motivating* reason to act. On Sinclair's view, the agent can be moved to act *for* their elusive reason so long as the *advisor* has certain attitudes toward that reason. On Ridge and McKeever's view, the agent can be moved to act *for* their elusive reason so long as the motivating reason the advisor gives them is *Russellian-identical* to their elusive reason. Since the fact that I like to kick cans down the road is a reason for me to kick *this* can only if I do not regard this *or any* fact as a reason for me to kick *this* can down the road, it is not clear how the possibility of a reliable advisor, whose advice consists of considerations that *I* can come to regard as a reason to kick this can down the road, could help reconcile my elusive reasons with Motivational Constraint. LeTrain and I are walking down the road, and LeTrain sees a can in the gutter. He knows that I like to kick cans down the road, albeit only for no reason, and concludes that the fact that I like to kick cans down the road is a reason for me to kick *this* can down the road. It seems, however, that LeTrain cannot give me a *reason* to kick this can down the road, since if I come to regard this reason as a reason for me to kick this can down the road, it will cease to be a reason. The same considerations show that the derivative-reason strategy does not help my elusive reason satisfy the Motivational Constraint, even on a revised Motivational Constraint* conception of the latter. That *I would be pleased if I kicked this can down the road* is a reason in virtue of being *explained by* the fact that *I like to kick cans down the road*. But, just like the latter, the former consideration is a reason for me to kick *this* can down the road only if I do not regard it as such, given that I *only* like to kick cans down the road if I do it for no reason. But that means that any reason derivative of the fact that *I like to kick cans down the road* is also elusive.

4. CONCLUSION

I have argued that elusive reasons present a serious challenge to the motivational constraint. With respect to one kind of elusive reason, *blindspot* reasons, I have argued that three strategies for reconciling such reasons with the motivational constraint fail. I then introduced a second kind of elusive reason that, I argued, also cannot be made consistent with the motivational constraint using any of the three strategies previously examined. That the motivational constraint appears

to be endangered by elusive reasons is a significant result, as it appears to endanger a number of popular views about reasons for action.

Consider what Kieran Setiya calls *Internalism about Reasons*: the fact that *p* is a reason for *A* to ϕ only if *A* is capable of being moved to ϕ by the belief that *p*.[44] Setiya argues that a wide range of views about normative reasons are best interpreted as arguing for or assuming Internalism about Reasons.[45] Another version of practical reasons internalism says that *p* is a reason for *A* to ϕ only if, if *A* were in circumstances *C*, *A* would be moved to ϕ by the belief that *p*.[46] As noted in section 2, being moved to act for the reason that *p* entails having the belief that *p* and its causing movement to ϕ. If, in order for *p* to be a reason for one to act, one must be *capable* of being moved to ϕ by the belief that *p*, *or* it must be the case that if one were in circumstances *C*, one *would* be moved to ϕ by the belief that *p*, then it must be logically possible for one to be moved to act for the reason that *p*. Thus, both of these views entail Motivational Constraint.

Motivational Constraint is not only relevant to internalist theories of reasons. It is also entailed by views that connect reasons to good practical reasoning. Setiya claims that the fact that *p* is a reason for *A* to ϕ just in case *A* has a collection of psychological states, *C*, such that the disposition to be moved to ϕ by *C* and the belief that *p* is a good disposition of practical thought, and *C* contains no false beliefs.[47] Jonathan Way argues that for the fact that *p* to be a reason for *S* to ϕ is for there to be a good pattern of reasoning from the belief that *p*, perhaps together with other correct attitudes that *S* has to ϕ-ing.[48] Both of these views clearly entail Motivational Constraint. For Way's view to hold, it must be logically possible that one ϕs for the reason that *p*; and this entails that it is logically possible that one is *moved to* ϕ for the reason that *p*. On Setiya's view, *p* is a reason to ϕ only if on the basis of *C* and *p* one is able to be moved to ϕ. This clearly entails that *p* is a reason to ϕ only if it is logically possible that one is moved to ϕ for the reason that *p*.[49]

44 Setiya, "Introduction," 4–5.

45 Including Nagel, *The Possibility of Altruism*; Williams, "Internal and External Reasons"; Korsgaard, "Skepticism about Practical Reason"; Velleman, "The Possibility of Practical Reason"; Setiya, "Introduction," 2–18.

46 For examples of views that satisfy this schema, see Manne, "Internalism about Reasons"; Joyce, *The Myth of Morality*; McDowell, "Might There Be External Reasons?"; Korsgaard, "Skepticism about Practical Reason"; Darwall, *Impartial Reason*.

47 Setiya, *Reasons without Rationalism*.

48 Way, "Reasons as Premises of Good Reasoning."

49 Way and Whiting have examined the problem of blindspot reasons and offered a response that relies on particular interpretation of the motivational constraint ("Reasons and Guidance"). I hope to respond to that argument in later work. This paper relies on a reading of

The strategies I have discussed for reconciling elusive reasons with the motivational constraint are not the only ones available, but they are some of the best attempts committed to print thus far. If what I have said is right, then vindicating the motivational constraint—and with it, the theories of practical reasons discussed in this section—must rest on another strategy.[50]

Duke University School of Law
brossichi@gmail.com

REFERENCES

Arpaly, Nomy. "Moral Worth." *Journal of Philosophy* 99, no. 5 (May 2002): 223–45.

———. *Unprincipled Virtue.* Oxford: Oxford University Press, 2003.

Broome, John. *Rationality through Reasoning.* Oxford: Oxford University Press, 2013.

Comesaña, Juan, and Matthew McGrath. "Having False Reasons." In *Epistemic Norms: New Essays on Action, Belief and Assertion*, edited by Clayton Littlejohn and John Turri, 59–78. Oxford: Oxford University Press, 2014.

Dancy, Jonathan. *Practical Reality.* Oxford: Oxford University Press, 2000.

Darwall, Steven. *Impartial Reason.* Ithaca: Cornell University Press, 1983.

Finlay, Steven. "The Reasons that Matter." *Australasian Journal of Philosophy* 84, no. 1 (2006): 1–20.

Hornsby, Jennifer. "A Disjunctive Conception of Acting for Reasons." In *Disjunctivism: Perception, Action, Knowledge*, edited by Adrian Haddock and Fiona MacPherson, 244–61. Oxford: Oxford University Press, 2007.

Joyce, Richard. *The Myth of Morality.* Cambridge: Cambridge University Press, 2001.

Korsgaard, Christine. "Skepticism about Practical Reason." *Journal of Philosophy* 83, no. 1 (January 1986): 5–25.

Lord, Errol, and Barry Maguire. "An Opinionated Guide to the Weight of Rea-

---

the motivational constraint that is important *because* it is entailed by so many common claims about reasons.

50 I would like to thank Robert Audi, Ted Warfield, Jamie Dreier, Sara Bernstein, Meghan Sullivan, Sean McKeever, and two anonymous reviewers for the *Journal of Ethics and Social Philosophy* for helpful comments and criticisms. Thanks also to participants at the 2018 Kentucky Philosophical Association Annual Meeting, the 2018 Great Lakes Philosophy Conference, and the 2018 Wisconsin Philosophical Association Annual Meeting.

sons." In *Weighing Reasons*, edited by Errol Lord and Barry Maguire, 3–24. Oxford: Oxford University Press, 2016.

Manne, Kate. "Internalism about Reasons: Sad but True?" *Philosophical Studies* 167, no. 1 (January 2014): 89–117.

Markovits, Julia. "Acting for the Right Reasons." *Philosophical Review* 119, no. 2 (April 2010): 201–42.

———. *Moral Reason.* Oxford: Oxford University Press, 2014.

McDowell, John. "Might There Be External Reasons?" In *World, Mind and Ethics: Essays on the Ethical Philosophy of Bernard Williams*, edited by J. E. J. Altham and Ross Harrison, 68–85. New York: Cambridge University Press, 1995.

McKeever, Sean, and Michael Ridge. "Elusive Reasons." In *Oxford Studies in Metaethics*, vol. 7, edited by Russ Shafer-Landau, 110–37. Oxford: Oxford University Press, 2012.

———. *Principled Ethics: Generalism as a Regulative Ideal.* Oxford: Oxford University Press, 2006.

Nagel, Thomas. *The Possibility of Altruism.* Oxford: Clarendon Press, 1970.

Nair, Shyam. "How Do Reasons Accrue?" In *Weighing Reasons*, edited by Errol Lord and Barry Maguire, 56–73. Oxford: Oxford University Press, 2016.

Paakkunainen, Hille. "Can There Be Government House Reasons for Action?" *Journal of Ethics and Social Philosophy* 12, no. 1 (September 2017): 56–93.

Parfit, Derek. "Rationality and Reasons." In *Exploring Practical Philosophy: From Action to Values*, edited by Dan Egonsson, Jonas Joesefsson, Bjorn Petersson, and Toni Rønnow-Rasmussen, 19–39. Farnham, UK: Ashgate Publishing, 2001.

Scanlon, T. M. *Being Realistic about Reasons.* Oxford: Oxford University Press, 2014.

Schroeder, Mark. *Slaves of the Passions.* New York: Oxford University Press, 2007.

Setiya, Kieran. "Introduction: Internal Reasons." In *Internal Reasons*, edited by Kieran Setiya and Hille Paakkunainen, 1–34. Cambridge, MA: MIT Press, 2011.

———. *Reasons without Rationalism.* Princeton: Princeton University Press, 2007.

Sinclair, Neil. "On the Connection between Normative Reasons and the Possibility of Acting for Those Reasons." *Ethical Theory and Moral Practice* 19 (2016): 1211–23.

Sorenson, Roy. *Blindspots.* New York: Oxford University Press, 1988.

Stroud, Sarah. "Moral Worth and Rationality as Acting for Good Reasons." *Philosophical Studies* 134, no. 3 (June 2007): 449–56.

Van Inwagen, Peter. "Ability and Responsibility." *Philosophical Review* 87, no. 2 (April 1978): 201–24.

Velleman, J. David. "The Possibility of Practical Reason." *Ethics* 106, no. 206 (July 1996): 694–726.

Way, Jonathan. "Creditworthiness and Matching Principles." In *Oxford Studies in Normative Ethics*, vol. 7, edited by Mark C. Timmons, 207–28. Oxford: Oxford University Press, 2017.

———. "Reasons as Premises of Good Reasoning." *Pacific Philosophical Quarterly* 98, no. 2 (June 2017): 1–20.

Way, Jonathan, and Daniel Whiting. "Reasons and Guidance." *Analytic Philosophy* 57, no. 3 (September 2016): 214–35.

Williams, Bernard. "Internal and External Reasons." In *Moral Luck*, 20–39. Cambridge: Cambridge University Press, 1981.

Journal of Ethics and Social Philosophy
VOL. 20, NO. 1 · JULY 2021
https://doi.org/10.26556/jesp.v20i1.954

WHY EXTENDING ACTIONS THROUGH TIME CAN VIOLATE A MORAL RIGHT TO PRIVACY

Björn Lundgren

Recently, Ingmar Persson and Julian Savulescu—in the context of defending their book *Unfit for the Future* and the irrelevance of a moral right to privacy for biomedical moral enhancement—questioned how extending an action in time could "bring a *moral* right into existence."[1] In this brief reply, I will argue that the right to privacy can be violated by temporally extending actions that in themselves do not violate the right to privacy.

Persson and Savulescu "take a moral right to privacy to be a right against others that they don't *acquire* (and sustain) certain (true) beliefs about us" (35). Furthermore, according to them, a moral right to privacy does not protect against stalking or gawking.[2] This is because, first, they deny any moral right "not [to] be tailed and/or stared at for a short period of time" (36). Second, they argue that if an action *A* does not violate a right *R* then a temporal extension of *A* does not violate *R* either, reasoning that "it would be odd if such a right eventually kicks in if this period is gradually extended: How could the mere passage of time bring a *moral* right into existence?" (36). Hence, if we accept their argument, a moral right to privacy does not protect against stalking or gawking because it does not protect against being briefly tailed/followed or stared/looked at.

Moreover, in defending the idea that a moral right to privacy does not protect against stalking or gawking, Persson and Savulescu attempt to explain why many legal systems may reasonably uphold "a *legal* right not to be systematically followed around." They argue that this legal right cannot be "justified by a moral right to the same effect" because we do not "have a *moral* right not to be tailed and/or stared at for a short period of time"; instead they think that legal protection against stalking or gawking is better defended because it is "highly unlikely that someone takes the trouble of keeping an eye on somebody else for [a]

1 Persson and Savulescu, "The Irrelevance of a Moral Right to Privacy for Biomedical Moral Enhancement," 36, hereafter cited parenthetically.

2 By "stalking" (and "gawking," respectively), I will henceforth mean tailing or following (staring or looking at) a person for an extended period of time.

lengthy period of time, unless he plans to *use* this information for some purpose" (i.e., to violate someone's right). Furthermore, they hold that it is *this* risk that makes people uneasy about being stalked (36).

This raises at least three questions, which I will address in the following sections. First, is it correct—on their own definition of a moral right to privacy—to say that stalking and gawking do *not* violate a moral right to privacy? I will argue that it is not. Moreover, I will show that the same holds true for other conceptions of a moral right to privacy to illustrate a broader conceptual agreement that a moral right to privacy can be violated by stalking and gawking. Second, is there any *other* reason (i.e., other than fear of sinister actions) why people might be uneasy about being stalked or gawked at? I will argue that there is and that these reasons explain why a right to privacy should protect against stalking and gawking. Third, do we more broadly have reasons to think that a moral right to privacy should protect against stalking and gawking? I will argue that we do, because such actions are not part of what we implicitly consent to by entering a public sphere.

I

As previously noted, Persson and Savulescu argue that a moral right to privacy does not protect against stalking or gawking because (1) it does not protect against being followed or being looked at briefly and (2) it does not protect against temporal extensions of actions that it does not protect against.

However, it is easy to see that Persson and Savulescu's claim is false on their own definition of a moral right to privacy. Remember that according to their definition, a moral right to privacy is a "right against others that they don't *acquire* (and sustain) certain (true) beliefs about us" (35). Consider the following example: you are walking in public and you see a person who is suddenly, by accident, nude (e.g., because a powerful gust of wind has blown her clothes away). Because the person's nudity is accidental, it is fair to say that by seeing that person nude (only for a moment), you have not violated that person's right to privacy. Yet, suppose that it is obvious to you that this person does not enjoy being publicly nude. If you were then to temporally extend the action of looking at that nude person, would that be a violation of their right to privacy? It is clear that one consequence of extending the action temporally is that over time you would acquire *more* true beliefs about that person. Hence, on Persson and Savulescu's definition, extending an action in time can violate a moral right to privacy.[3]

3 Strangely, as indicated in previous quotes, Persson and Savulescu seem to recognize that extendedly looking at or following someone implies that one acquires (more) information about that person. An anonymous reviewer pointed out that because they use the concept

More generally, the same type of argument can arguably be applied to any definition of a moral right to privacy that includes a moral right against others' access or control of informational privacy. For example, according to Adam D. Moore, a right to privacy "is a right to limit public access to oneself and to information about oneself."[4] While applying Moore's conception of a right to privacy would require a closer analysis, it arguably follows that we have a right to restrict others' access to information about us (i.e., we have a right that restricts people from, e.g., gawking at our nude bodies), and hence that unwanted stalking or gawking can violate such a right. Furthermore, according to Helen Nissenbaum's account of privacy as contextual integrity, the right to privacy "is a right to live in a world in which our expectations about the flow of personal information are, for the most part, met."[5] On Nissenbaum's account, it arguably follows that gawking and stalking would constitute a violation of our right to privacy because they violate our normal expectations of personal information flow (i.e., gawking and stalking are outside of the appropriate contextual norms, because, e.g., gawking at someone's nude body in the previously given example would violate some such norm).[6]

Thus, it seems that on Persson and Savulescu's, Moore's, and Nissenbaum's

certain information rather than merely *information* in their definition, we should perhaps think that it indicates a special kind of information. However, it is clear that what they have in mind is private information (36).

Instead, Persson and Savulescu seem to rest their argument on the idea that information acquisition does not constitute an action and, hence, that information acquisition is permissible because it is not "under the control of our will" (36). However, while we can think of examples in which that analysis might make sense (their best example is of an omniscient being, which cannot help knowing things about us), what matters here is that their analysis fails for the example considered above.

Lastly, another possible interpretation is that they think of information that we can be certain of, but even if that were the case, it would not affect any of my arguments.

4 Moore, "Defining Privacy," 420. The formulation Moore provides should be read as a right to *control* access. On the following page in his article, he settles for the following formulation: "Definition: A right to privacy is a right to control access to and uses of—places, bodies, and personal information" (421). I am using the above formulation since it allows me to illustrate the plausibility that both control and limited-access conceptions of the right to privacy standardly should result in a similar analysis of the given example.

5 Nissenbaum, *Privacy in Context*, 231.

6 An anonymous reviewer argued that because Nissenbaum's theory is contextual, my argument above may not hold in some contexts, saying that "in the US, there is a strong (often discussed) sense of 'no reasonable expectation of privacy in public.'" Yet, it is clear that Nissenbaum does think that we have some modicum of a right to privacy in public (see *Privacy in Context*, 116, for examples relating to the United States). This is grounded in her criticism of the private–public distinction and the informational power of social-technical systems. In her discussion on privacy in public (113–26), she concludes that "because of these [socio-technical] powers, there are no actors, no spheres, no information that can be assigned

accounts of a moral right to privacy, we should hold that such a right protects against gawking and stalking irrespective of whether it protects against briefly looking or following. Moreover, it seems reasonable that this would extend to many other conceptions of the right to privacy as well.

II

According to Persson and Savulescu, stalking or gawking makes people uneasy because they think that the stalkers or gawkers "are planning to take some actions against" them (36). Yet, this uneasiness can easily be defended on other grounds.

After the #MeToo movement, it should be clear that extended looking (i.e., gawking) *can* make a person uneasy even if that person is not worried that the gawker (or a stalker) has sinister plans (although such worries should be taken even more seriously). That is, the extended looking can itself be problematic. One problem is simply that staring at a person *can* make them feel objectified.[7]

While feeling objectified can be problematic in itself, it can also be problematic for other reasons (i.e., beyond any fear from the gawker/stalker). For example, it has been shown that objectification of women can lead to a form of self-objectification, which in turn can increase negative emotions such as anger and shame.[8]

However, although it is certain that objectification due to gawking can constitute a harm, we can question whether this is the kind of harm that a moral right to privacy should protect against. The answer to this question depends, in part,

---

unconditionally to the domain of the public, free of all and any constraints imposed by rights of privacy; none are 'up for grabs'" (126).

Moreover, while there may be sensible disagreement about which norms *should* guide information flows and whether that can be suitably established on Nissenbaum's theory, it is fair to say that there has been an uprising about such norms recently (as I will address in the next section).

7 Note that I write "can" (i.e., temporally extended behavior, such as staring, *can* constitute a harm). That does not mean that extending such kinds of behaviors is *always* harmful. For example, in certain settings a person may look at another person, after which that person looks back in a liking manner. This may implicitly indicate a consent to being extendedly looked at (I will discuss the relevance of consent in section III). Therefore, what we may call "romantic staring" may be an example of an action that can be extended over time without being harmful. While there is something interesting to be said about how we qualify the distinction between actions that become harmful if extended and actions that do not, that would be beyond the purpose of this article. My aim here is merely to show that there are some actions that if extended over time would constitute a harm and would violate a moral right to privacy.

8 Koval et al., "How Does It Feel to Be Treated Like an Object?" 894.

on whether such a harm is a privacy harm. That is, if the harm is a privacy harm, then there is reason to think that a moral right to privacy should protect against such a harm. Consider, again, the previous example about the person that accidentally becomes nude in public (keep in mind that the person obviously does not enjoy being publicly nude). In such a situation, it is fair to say that the nude person is harmed by others' acquisition of *personal* or *privacy-sensitive information* about her (here I suppose that information about a person's naked body is a standard example of what can be privacy sensitive). Because this acquisition is a privacy harm, we should hold that a moral right to privacy should protect against it (while we cannot be blamed for purely accidentally seeing someone nude, we can be blamed for extending such a look).[9]

It is important to note that while being exposed in the nude may also be harmful for other reasons (e.g., because the gawker has sinister plans), that does not affect the point that I am making here, since there *can* be a pure privacy harm without such other factors.[10] Hence, if we agree that there can be such a harm, then because this harm is rooted in the acquisition of personal or privacy-sensitive information, we should hold that this harm is specifically what a moral right to privacy should protect against.[11]

Although the arguments I have presented should suffice to show how extending an action of looking (or following) can violate a moral right to privacy, in the third and final section, I will address this issue from the perspective of consent and privacy in public.

III

Consider the example from the previous sections (i.e., you see a person in public accidentally becoming nude). If Persson and Savulescu's argument is correct, you would not violate any right (to privacy or otherwise) of that person by *con-*

9 Of course, whether an action is a violation of a right to privacy depends on various other conditions. I take it here that the examples I will use either are compatible with or can be adapted to such requirements.

10 See Rachels, "Why Privacy Is Important," for arguments about the nonderivative value of privacy.

11 An anonymous reviewer suggested that the arguments in this section seem to indicate that Persson and Savulescu are wrong to restrict the right to privacy to an information-based account. Although the above issues can be analyzed in terms of a purely information-based account of the right to privacy (i.e., because the harm can always be explained by a process of information transfer), I agree that it would arguably be both easier and more analytically sensible to analyze them in terms of a conception of the right to privacy that also covers, for example, bodily, decisional, or spatial privacy.

tinuing to look at her (even if she is obviously not comfortable with the current situation). However, as I have argued, this is wrong on the grounds that the nude person is obviously harmed (and that this is a privacy harm), because, for example, we—in accordance with Persson and Savulescu's information-based account of the right to privacy—acquire more (private) information (knowledge) about the individual.

The consequence of accepting an information-based account of the moral right to privacy—such as that of Persson and Savulescu—can be further supported by an argument that it matters whether there is (implicit) consent. This "argument from consent" can also be used to address another related issue: privacy in public.[12]

Whether we have a right to privacy in public is debated, in part, because we allow others to acquire information about us when we enter a public sphere. However, in many situations, these information acquisitions are implicitly consensual and hence unproblematic vis-à-vis a moral right to privacy. That is, by entering a public sphere, you standardly (implicitly) consent to being seen, but—as I have argued—not to being stared at. Why is that? To explain this, let us start by considering that a moral right to privacy protects against being looked at in various situations. For example, when alone in your home, a moral right to privacy morally protects you against a Peeping Tom. Yet, alone in your home you also have a right to look at your surroundings. If both these rights were to apply fully—to everyone—and in public, we would have conflicting rights. People cannot both have a right not to be looked at and a right to look at others.

Some might think that this (supposed) conflict can be resolved because a moral right to privacy does not apply in the public sphere (or that it is limited somehow). While my aim here is not to fully defend the notion of privacy in public, I believe that those that hold that we cannot have privacy in public must defend a notion of privacy as a bivalent concept (e.g., a state, or condition, that you are either in or not). I, however, hold that we can have at least a modicum of privacy in public. That is, if privacy is something that one can have less and more of, then it is reasonable to think that one can have less or more privacy in public (e.g., the person who accidentally becomes nude is in a state of less privacy than she was before). Likewise, it is reasonable to think that one can have a right to privacy in public, a right that provides some protection, in some situations, but perhaps not complete protection in all situations.

My suggestion is, thus, that this *prima facie* conflict of rights can be resolved, in part, by the fact that by entering the public sphere, we (implicitly) consent to

12 In arguing for this, I do not presume that the concept of (implicit) consent can be used in *every* situation to determine whether there is a violation of a moral right to privacy.

being seen. Moreover, by entering the public sphere, we standardly prepare ourselves for the possibility of being seen and take precautions to avoid exposing ourselves too much. For example, most people cover parts of their body to keep others from looking at those parts. Consider, again, the example of the person who accidentally becomes nude in front of us. Granted that she used clothes to partly protect against her body being seen, extended looking at her nude body would not be something that she (implicitly) consented to when entering the public sphere. Similarly, a person that enters the public sphere normally consents to being looked at, but not to being gawked at. There are various illustrative examples that further support this way of analyzing these situations—for example, there are people, such as street performers, who arguably implicitly consent to being extendedly looked at (see note 7 for another illustrative example). Thus, if my arguments are correct, then a moral right to privacy can also be said to morally protect against gawking and stalking because we normally do not (implicitly) consent to such activities.[13]

Some may ask why our moral right to privacy would protect us in this way. That is, why would our moral right to privacy protect against being gawked at in many ordinary public situations? This is where my argument reaches full circle and refers to the previous arguments. The first two arguments that I have presented stand on their own. To some extent, this is true for the argument from consent as well (i.e., we can use this kind of reasoning to make sense of, e.g., why people normally implicitly consent to being looked at in public, but not to being gawked at). However, for full justification of the argument from consent, we must rely on "the argument from harm." This is not circular; instead, it shows a strong coherence between different ways of reasoning about privacy.[14]

To briefly summate: an action @ by an agent *A* that does not violate a moral right to privacy of an agent *B can*, if extended temporally, violate *B*'s moral right

13 Given that people sometimes have a right not to be looked at, it follows that people occasionally have a duty to look away. However, such a duty would arguably be *pro tanto* (i.e., a duty that can be overridden). The duty can, for example, be overridden in case the person (who has a right not to be extendedly looked at) needs some sort of help and you can only do so by looking at that person, or it may be overridden if you have a need to look in a direction that inadvertently includes extending your gaze at this person (e.g., if you are driving a vehicle and need to look at the road). The precise limits of such a duty are beyond the scope of this paper.

14 Of course, there are also other arguments that could be used to further extend the support for my position—for example, by reference to *respect for persons* and what is a *private affair* (see, e.g., Benn, "Privacy, Freedom, and Respect for Persons"). An anonymous reviewer pointed out that this type of literature may, in particular, be relevant for the discussion in section 11. However, because of the desire to keep the discussion short, I could not engage with it in any detail.

to privacy—for example, because by extending @, *A* can acquire more information about *B*, something that *B* has not explicitly or implicitly consented to. Furthermore, *B* has a right to withhold such a consent, because such actions can harm *B* relative to her privacy.[15]

Umeå University, Institute for Futures Studies, and Stockholm University
bjorn.lundgren@umu.se

REFERENCES

Benn, Stanly I. "Privacy, Freedom, and Respect for Persons." In *Philosophical Dimensions of Privacy: An Anthology*, edited by Ferdinand D. Schoeman, 223–44. Cambridge: Cambridge University Press, 1984.

Koval, Peter, Elise Holland, Michael J. Zyphur, Michelle Stratemeyer, Jennifer Makovec Knight, Natasha Bailen, Renee J. Thompson, Tomi-Ann Roberts, and Nick Haslam. "How Does It Feel to Be Treated Like an Object? Direct and Indirect Effects of Exposure to Sexual Objectification on Women's Emotions in Daily Life." *Journal of Personality and Social Psychology* 116, no. 6 (June 2019): 885–98.

Moore, Adam. "Defining Privacy." *Journal of Social Philosophy* 39, no. 3 (Fall 2008): 411–28.

Nagel, Thomas. "Concealment and Exposure." *Philosophy and Public Affairs* 27, no. 1 (January 1998): 3–30.

Nissenbaum, Helen. *Privacy in Context: Technology, Policy, and the Integrity of Social Life*. Stanford, CA: Stanford Law Books, 2010.

Persson, Ingmar, and Julian Savulescu. "The Irrelevance of a Moral Right to Privacy for Biomedical Moral Enhancement." *Neuroethics* 12, no. 1 (April 2019): 35–37.

———. *Unfit for the Future: The Need for Moral Enhancement*. Oxford: Oxford University Press, 2012.

Rachels, James. "Why Privacy Is Important." *Philosophy and Public Affairs* 4, no. 4 (Summer 1975): 323–33.

15 Many thanks to three anonymous reviewers for the *Journal of Ethics and Social Philosophy*. Moreover, I want to acknowledge that this work was partially supported by the Wallenberg AI, Autonomous Systems and Software Program—Humanities and Society (WASP-HS), funded by the Marianne and Marcus Wallenberg Foundation and the Marcus and Amalia Wallenberg Foundation (grant number: MMW 2018.0116).