

# JOURNAL *of* ETHICS & SOCIAL PHILOSOPHY

VOLUME XII · NUMBER 3

*December 2017*

## ARTICLES

Constitutivism without Normative Thresholds

*Kathryn Lindeman*

Can Objectivists Account for Subjective  
Reasons?

*Daniel Wodak*

Immigration Policy and Identification Across  
Borders

*Matthew Lindauer*

## DISCUSSIONS

Can We Intend the Past?

*Oded Na'aman*

Hybrid Non-Naturalism Does Not Meet the  
Supervenience Challenge

*David Faraci*

Is Agent-Neutral Deontology Possible?

*Matthew Hammerton*



JOURNAL *of* ETHICS  
& SOCIAL PHILOSOPHY

VOLUME XII · NUMBER 3

*December 2017*

ARTICLES

- 231 Constitutivism without Normative Thresholds  
*Kathryn Lindeman*
- 259 Can Objectivists Account for Subjective  
Reasons?  
*Daniel Wodak*
- 280 Immigration Policy and Identification Across  
Borders  
*Matthew Lindauer*

DISCUSSIONS

- 304 Can We Intend the Past?  
*Oded Na'aman*
- 312 Hybrid NonNaturalism Does Not Meet the  
Supervenience Challenge  
*David Faraci*
- 319 Is Agent-Neutral Deontology Possible?  
*Matthew Hammerton*

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.

*Editor*

Mark Schroeder

*Associate Editors*

James Dreier

Julia Driver

David Estlund

Andrei Marmor

*Discussion Notes Editor*

Douglas Portmore

*Editorial Board*

Elizabeth Anderson	Philip Pettit
David Brink	Gerald Postema
John Broome	Joseph Raz
Joshua Cohen	Henry Richardson
Jonathan Dancy	Thomas M. Scanlon
John Finnis	Tamar Schapiro
John Gardner	David Schmidtz
Leslie Green	Russ Shafer-Landau
Karen Jones	Tommie Shelby
Frances Kamm	Sarah Stroud
Will Kymlicka	Valerie Tiberius
Matthew Liao	Peter Vallentyne
Kasper Lippert-Rasmussen	Gary Watson
Elinor Mason	Kit Wellman
Stephen Perry	Susan Wolf

*Managing Editor*

Susan Wampler

*Editorial Assistance*

Renee Bolinger

*Typesetting*

Matthew Silverstein



## CONSTITUTIVISM WITHOUT NORMATIVE THRESHOLDS

*Kathryn Lindeman*

Even bad coffee is better than no coffee at all.  
—David Lynch

CONSTITUTIVISTS ABOUT NORMS in metaethics explain the normative standards in a domain by appealing to constitutive features of the members of the domain.<sup>1</sup> So, for example, Korsgaard seeks to explain what it is to be a good person by appeal to what it is to *be* a person. This is an appealing explanatory strategy because, if successful, it offers a general and comprehensive account of how different normative standards govern distinct normative domains. It could, e.g., explain how moral normative standards govern agents, practical normative standards govern purposive behavior, epistemic normative standards govern beliefs and inferences, artifactual normative standards govern artifacts, and biological normative standards govern organs and biological processes.<sup>2</sup>

- 1 Constitutivism is often presented, as with Enoch, as “a family of views that hope to ground normativity in norms, or standards, or motives, or aims that are constitutive of action and agency” (“Shmagency Revisited,” 208). Or, by Tiffany, as “the view that it is possible to derive contentful, normatively binding demands of practical reason and morality from the constitutive features of agency” (“Why Be an Agent?” 223). I prefer a more ecumenical characterization in which the norms need not be those of practical reason and they need not be grounded in what is constitutive of action or agency, but rather can be grounded in any constitutive feature of the individual governed by the norms. An argument for this preference is beyond the scope of this paper, but this view is implicitly presupposed by the targets of my arguments. An even more inclusive characterization is defended by Michael Smith (in “A Constitutivist Theory of Reasons”), who characterizes any account that reduces a normative feature to a constitutive feature of some other thing.
- 2 Constitutivism is often contrasted with constructivism and realism. To understand these contrasts, it is important to specify the *domain* about which one is a constitutivist, realist, or constructivist. It is possible to be a constitutivist about one domain and an error theorist about another, for example, though it is not possible to be an error theorist and constitutivist about the same domain. I discuss constitutivism about norms. For a useful discussion of the relationship between constructivism and constitutivism, see Schafer, “Realism and Constructivism in Kantian Metaethics (1).”

A central challenge for normative constitutivism is to account for the connection between an individual's constitutive nature and its normative standards. Critics worry that this kind of explanation is impossible.<sup>3</sup> One prominent objection to constitutivism is the worry that normative standards governing an individual cannot be grounded in its kind in the way constitutivists claim without ruling out the possibility that kind-members could violate those standards. This objection is sometimes called the Problem of Bad Action.<sup>4</sup>

The main lesson of this paper is that current constitutivist accounts face a serious threat, though not from the well-known Problem of Bad Action. Rather, the way in which they take norms to constitute norm-governed kinds commits them to what I call the Threshold Commitment. This, I argue, leads to serious costs for constitutivists that should lead them to reject this commitment and seek another explanation of this connection.

#### 1. VIOLABILITY AND NAÏVE CONSTITUTIVISM

One naïve constitutivist explanation of norms is that the norms themselves are constitutive of individuals in virtue of being satisfied. This account generates norms that are unable to account for what I call Violability.

*Violability:* That a norm applies to  $x$  and requires  $p$  does not logically or metaphysically imply that  $x$  satisfies  $p$ .

Problematically, such explanations cannot be used to account for the obvious fact that there are better and worse things in the world. Railton raises a concern of this form when he worries how we could, e.g., determine whether someone who fails to aim at truth in forming beliefs is defective. He writes that “to dis-

- 3 In this paper, I focus on the criticism that the constitutive connection between norms and kinds causes insurmountable problems for constitutivists. Recently another criticism has become increasingly popular against constitutivism. David Enoch (see “Agency, Shmagency” and “Shmagency Revisited”) and Kieran Setiya (see “Explaining Action” and *Reasons without Rationalism*) worry that the constitutive starting point is either too anemic to provide a full accounting of the normative landscape or too robust to be normatively uncontroversial. This objection is orthogonal to the problem I discuss in this piece, but I discuss it in “An Explanation of Constitutivist Normativity.” For helpful discussions of the issues remaining for constitutivists in light of these objections, see Rosati, “Agents and ‘Shmagents’”; and Silverstein, “The Shmagency Question.”
- 4 The problem of bad action is so called because it has been raised in response to constitutive accounts of practical norms. However, because the problem is supposed to arise from the form of constitutive explanation, not the nature of the practical domain, there will be analogous problems for all normative domains. The problem is, therefore, more general than the name lets on.

cover that the metal in the sample tray on one's laboratory bench has atomic number 82 is not to discover that it is 'defective gold,' but rather that it is not gold at all. A similar problem confronts all constitutive arguments."<sup>5</sup> Clark raises similar worries specific to Velleman: if any action must have the aim of autonomy, but autonomy is a precondition of action, all action necessarily achieves its aim. If constitutive accounts are the only explanation of the normative features of action (including reasons for action), any action that achieves its aim should, it seems, have no reasons that speak against it, making it incapable of being rationally criticized.<sup>6</sup>

Railton and Clark both highlight specific instances of a more general concern about the role this naïve account gives to standards: if what it is to be a member of a norm-governed kind is to meet a certain standard, then that standard seems incapable of simultaneously serving as a way to differentiate good kind-members from bad kind-members. The naïve view then is a nonstarter for constitutivists for exactly the reason Railton and Clark highlight. Any successful constitutivist explanation must provide a metaphysical account of the constitutive features of normatively evaluable kinds that is consistent with Violability. This account must show that there is some connection between the kind and the norms tight enough to account for the constitutive explanation, but weak enough to allow for defect and other evaluations according to norms.<sup>7</sup>

Constitutivists are clearly aware of the need for non-naïve constitutive explanations. Velleman, for example, recognizes the importance of distinguishing between the conditions on aiming and actually achieving an aim. He writes that "If autonomy were the constitutive aim of action, then every instance of action . . . would turn out to be a success."<sup>8</sup> Korsgaard seems similarly interested in leaving room for norms to be constitutively understood while not satisfied. In the specific case of norms constitutive of action, she writes:

- 5 Railton, "On the Hypothetical and Non-Hypothetical in Reasoning about Belief and Action," 70.
- 6 Clark, "Velleman's Autonomism," 581–82. Clark quotes Velleman, *The Possibility of Practical Reason*, 185.
- 7 For worries of this nature by critics, see Railton, "On the Hypothetical and Non-Hypothetical in Reasoning about Belief and Action"; Clark, "Velleman's Autonomism"; Barandalla and Ridge, "Function and Self-Constitution"; Lavin, "Practical Reason and the Possibility of Error"; Silverstein, "Teleology and Normativity." For worries acknowledged by constitutivists, see the introduction in Velleman, *The Possibility of Practical Reason*; Korsgaard, *Self-Constitution*, especially sec. 8.1; and Katsafanas, *Agency and the Foundations of Ethics*, 61–63.
- 8 Velleman, *The Possibility of Practical Reason*, 30n37. Further discussion of Velleman's view can be found in section 6, below.

[If] it is the essential nature of [a kind] that it have a certain metaphysical property ... but in order to have that metaphysical property it must have a certain normative property ... then this explains why the [individual] must meet the normative standard: *it just isn't [a member of the kind] if it doesn't*. But it also seems as if it explains it rather too well, for it seems to imply that only good [individuals] really [are kind-members], and that there is nothing left for bad [kind-members] to be.<sup>9</sup>

There are less naïve explanations of how normative standards are related to the constitutive features of kind-members that preserve Violability. Constitutivists might, for instance, take normative satisfaction to be a scalar matter, and claim that some degree of normative satisfaction is constitutive. So long as the metaphysical property can be realized by a number of normative properties, or by having some normative property to some degree, this connection need not rule out Violability. So, constitutivist accounts that permit scalar norm satisfaction as a criterion of kind-membership would be perfectly compatible with violated norms and defective kind-members. Indeed, such views seem to be held by many constitutivists. Korsgaard and Velleman, for example, hold that constituting oneself and aiming at agency are activities that agents can do in better and worse ways. Neither accounts for the relationship between individuals and kinds on the naïve model that would make them natural targets for Railton's objection.

So, for example, Korsgaard, in *Self-Constitution*, explains the constitutive feature uniting kinds as a matter of having an internal teleological organization or form, which is a matter of minimally satisfying some constitutive standards or norms. Though you must be in the practice of constituting yourself to be evaluated according to the constituting-oneself standards (i.e., practical norms), there are better and worse ways of doing this, and so better and worse agents.<sup>10</sup> The norms of self-constitution are, therefore, scalar; there are ways of violating the norms and still being an agent, because you are still constituting yourself, just doing so badly.<sup>11</sup> Velleman introduces similar teleological features in his account by making it a criterial feature of agency that one constitutively aim at self-knowledge, where that aiming can be done more or less successfully.<sup>12</sup> Katsafanas has recently defended a Nietzschean constitutivist account of agency

9 Korsgaard, *Self-Constitution*, 160.

10 Korsgaard, *Self-Constitution*, ch. 2.

11 This point is nicely made by Walden, "Laws of Nature, Laws of Freedom, and the Social Construction of Normativity."

12 Velleman, *The Possibility of Practical Reason* and "Replies to Discussion on *The Possibility of Practical Reason*."

in which the constitutive aims of action are coherence and the will to power. On this view, one can be more or less coherent and thus be a better or worse agent.<sup>13</sup> Michael Smith has developed his account of idealized agency into a constitutivist explanation of practical normativity that seems best accounted for on a scalar model. On his view, for example, it is constitutive of the ideally rational agent that they have the desire to help and not harm, but in order to be an agent—and thereby be governed by the norms that the ideally rational agent necessarily satisfies—individual agents need not fully have these desires.<sup>14</sup>

Of course, there are worries about these scalar accounts and whether they can provide genuine normative assessments or standards. Clark, for example, considers a variant of Velleman's view on which some amount of autonomy, short of perfect autonomy, is required for action. Clark worries that this move would not fully account for the standard normative assessment for actions because the goal of autonomy up to a threshold cannot serve as a standard of assessment for action. Setiya objects that the constitutive normative requirements on these accounts do not provide sufficient material to fully account for practical norms or their scope.<sup>15</sup> Though the scalar accounts do not yet amount to responses to these worries, they do make progress over the naïve constitutivist account that falls to Violability concerns. Here, I want to focus on a problem arising precisely from the feature of scalar accounts that allows them to satisfy Violability.

## 2. THRESHOLD CONSTITUTIVISM

Such constitutivists make room for defective kind-members by identifying the metaphysical property essential for kind-hood with some minimal threshold of normative properties, rather than a single bivalent normative property. So, e.g., minimal conformity to the norms of action, rather than perfect conformity, is constitutive of action. This is progress over the naïve constitutivist account, which risks identifying the perfect with the real. However, by giving minimal norm satisfaction a criterial role, they maintain an identification of the real with the minimally good. Such views are unsatisfactory because they involve an implicit endorsement of what I call the Threshold Commitment:

*Threshold Commitment:* For norm-governed kinds, an individual must at

13 Katsafanas, *Agency and the Foundations of Ethics*.

14 Smith, "A Constitutivist Theory of Reasons."

15 See Clark, "Velleman's Autonomism"; Setiya, "Explaining Action," 371–76, and "Akrasia and the Constitution of Agency."

least partially satisfy the constitutive norms of a kind, or partially meet the constitutive aim of a kind, in order to be a member of that kind.<sup>16</sup>

In other words, norm-governed kinds are those that have kind-membership conditions constituted by minimally having the normative properties that are good-making for that kind. This is supposed to explain why individual kind-members must meet normative standards in a way that does not, in Korsgaard's words, explain it a little too well. Normative violators risk nonexistence, though there are many ways of falling short before you cease being a kind-member.<sup>17</sup> So, the Threshold Commitment is another commitment to normative satisfaction itself being a constitutive condition on kind-hood, albeit one that permits satisfying those norms to a greater degree with being better *qua* kind-member.<sup>18</sup>

This commitment is widespread among constitutivists.<sup>19</sup> Korsgaard, in introducing constitutive standards in the introduction to *The Constitution of Agency* writes:

Two things are important to notice about standards of this kind. First of all, constitutive standards are at once normative and descriptive. They are descriptive because an object must meet them, or at least aspire to meet them, in order to be what it is. And they are normative because an object

- 16 Note that the Threshold Commitment does not simply claim that there are thresholds that determine kind-membership; it takes a stand on the nature of those thresholds: they are determined by normative properties. Importantly, the Threshold Commitment takes a normative difference and attributes to it a metaphysical significance. Nothing in this paper should be taken as an argument against the possibility of understanding important or interesting relationships between normative failures and metaphysical status. Here I am simply highlighting a particular commitment that many constitutivists have to normative goodness *itself* being criterial or constitutive of kind-membership, and, correlatively, to there being criterial thresholds constituted by normative conditions.
- 17 It is not essential that there is a scalar norm that serves as the unique criterion for the Threshold Commitment. A similarly workable account might feature a number of binary criteria such that each would make a *K*-member better if it satisfied the criteria but any of which would be sufficient for making an individual *x* such a *K*-member.
- 18 On some readings of Korsgaard, this also leads to an identification of satisfying more *K*-norms with being *more of a K*-member and thus with being *more evaluable* by the constitutive norms of the relevant kind. I think that this is unlikely to be the best version of a Korsgaard-style view, but it is possibly her actual view. For a defense of this interpretation of Korsgaard and an argument that this view faces a related worry, see Silverstein, "Teleology and Normativity."
- 19 Matthew Silverstein has recently argued that endorsement of something like the Threshold Commitment is a problem for Korsgaard's constitutive account. The problems he identifies differ from mine, and his criticisms are limited to Korsgaard, but we substantially agree on the cause of the problem. See Silverstein, "Teleology and Normativity."

to which they apply can fail to meet them, at least to some extent, and is subject to criticism if it does not. This double nature finds expression in the fact that we can criticize such objects by saying that they are poor objects of their kind (“That’s a poor encyclopedia, it isn’t up to date!”) or by saying that they are not such objects at all (“That’s not an encyclopedia: it’s just a compendium of nineteenth-century opinion!”).<sup>20</sup>

Korsgaard holds both that it is a good-making standard of encyclopedias that they are up to date, and that sufficiently out-of-date encyclopedias are not encyclopedias but rather mere compendia of the opinion of a time.

Other constitutivists make similarly clear commitments. Katsafanas writes:

There is a great deal of distance between the standards minimally constitutive of household and the standards of an *excellent* house. There is a great deal of distance between a plywood shack in the forest that will dissolve back into the environment in a few years, and a sturdy stone house that will last for centuries. Both of these count as houses, but it seems natural to say that the stone house is a *better* house. We might describe this by saying that the same standards apply to the stone house and the shack: the standards of household. But there are better and worse ways of fulfilling these standards. At some point, a structure will have met the standards to a sufficient degree to qualify as a house, but will not have met the standards to a sufficient degree to qualify as an excellent house.<sup>21</sup>

Katsafanas, like Korsgaard, is committed to *both* the Threshold Commitment and a connection between norms *and* fundamental evaluative assessment. The normative properties of houses account for both the metaphysical difference between household and non-household, and the evaluative difference between merely adequate and excellent houses. These are the dual roles given to normative standards by those we might call “Threshold Constitutivists.”

As Korsgaard highlights, one appeal of the Threshold Commitment is its support for some (though, of course, not all) of our linguistic practices involving attributing and denying kind-hood to objects of evaluation. Though we might not identify perfection and reality, sometimes we do talk as if we identify extreme defect and nonbeing. I suspect we are less likely to join Korsgaard in accusing out-of-date encyclopedias of being non-encyclopedias; however, we do seem to make claims like (pointing to the liquid in a mug), “This coffee is so bad, it’s not

20 Korsgaard, *The Constitution of Agency*, 8.

21 Katsafanas, *Agency and the Foundations of Ethics*, 64.

even coffee!”<sup>22</sup> If meeting a minimum normative standard of coffee were a metaphysical condition on being coffee, this could be a literally true claim about the metaphysical status of the liquid under discussion rather than a metaphorical expression of disapproval. Implausibly, being *bad* enough at being coffee could constitute (directly and literally) not *being* coffee. If so, Lynch is wrong about some bad coffee; sometimes having bad *enough* coffee just *is* having none at all.

Linguistic appeals aside, according to its proponents the main advantage of understanding threshold normative properties as metaphysically constitutive of kind-membership is the ability to explain the legitimate normativity of those properties and to answer skeptical challenges to their authority. Korsgaard writes that “the *only* way to establish the authority of any purported normative principle is to establish that it is constitutive of something to which the person whom it governs is committed—something that she either is doing or has to do.”<sup>23</sup> According to Korsgaard, what should make you interested in building a good house is the risk that if you do not do it well enough, you will not end up with a house at all. This move requires something like a threshold; there is some *existential* (not merely pragmatic) risk of not satisfying kind-norms. In a similar spirit, Katsafanas writes that “a builder must aim at building a good house, if by ‘good’ [we] mean *a house that is at or above the cut-off point for household.*”<sup>24</sup> The Threshold Commitment is not an *ad hoc* or ancillary feature of these constitutivist views.<sup>25</sup>

### 3. THE COST OF REDUCED EXPLANATORY UNITY

Though Threshold Constitutivism makes progress over naïve constitutivism, it comes with serious costs. Normative Constitutivism has ambitions to be an explanatory strategy for norms in general. If there are norm-governed kinds without normative thresholds, then one cost to Threshold Constitutivists is decreased explanatory scope. Indeed, social kinds like Spouse appear to provide

22 This was once said to me by Luca Ferrero, who, years later, refuses to retract the claim or accept it as a figure of speech. It was, indeed, very bad. I maintain it was still clearly coffee.

23 Korsgaard, *Self-Constitution*, 32.

24 This move worries those suspicious that constitutivism does not require us to care about norms as such, or goodness as such, but only some more minimal standards. For example, see Enoch, “Agency, Shmagency” and “Shmagency Revisited.” The objection I am developing here holds even if the Threshold Constitutivist can address Enoch’s worries.

25 Some constitutivists appeal less to Threshold-type claims. In section 6, below, I argue that versions of constitutive aim accounts do not obviously avoid the Threshold Commitment. In section 7, I sketch a possible constitutive account that I argue does avoid the Threshold Commitment and is worth further development.

ready cases of kinds that lack normative thresholds.<sup>26</sup> Given the importance of explanatory unity, this means that Threshold Constitutivism comes with a cost of reduced explanatory unity.

It is clearly possible for there to be better and worse spouses and for particular spouses to improve or become worse *qua* spouse. We then are committed to thinking that there are ways of being a better and worse spouse that are generally true of spouses that we track when we make these assessments. The norms of spouses are those standards that would, when satisfied, make a spouse better *qua* spouse, and these norms are the subject matter of constitutivist explanation. The constitutivist will want to find some constitutive feature of spouses that explains the norms according to which we assess individual spouses as better or worse.<sup>27</sup> Spouse is, so understood, a goodness-fixing kind, the norms of which come from its constitutive nature.<sup>28</sup>

However, satisfying these norms to some minimal degree does not seem to be a constitutive requirement of being a spouse at all. One does not become a spouse by being a good enough one, and one cannot cease being a spouse *merely in virtue of* being a bad enough one. The metaphysical conditions on spouseness seem to involve multiply realizable social recognition practices. Different social groups can determine different practices for the social recognition, but at least in contemporary practice, we take legal recognition to be at least partially determinant of *becoming* a spouse, and recognized removal of legal recognition to be a sufficient (and, along with the death of one's spouse, exhaustive) condition on *ceasing* to be a spouse.

This is supported by how we view the recent changes in constitutional law in the United States. We recognize these changes to have permitted same-sex couples *to become* spouses. We do not tend to think that they were *already* spouses and that the law needed to be changed to *recognize* this fact.<sup>29</sup> This indicates that we view legal recognition as a condition on having the social kind Spouse,

26 Throughout, I use the singular, capitalized noun to pick out the kind, and lowercase uses to pick out instances in the singular or plural. So, here, I consider the kind Spouse, and later I discuss individual conditions on being a spouse and the norms that spouses are assessed according to.

27 Though it seems pretty clear to me that such an explanation is plausibly in the offing, nothing here turns on there being such an explanation. All we need here is the conclusion that the Threshold Commitment will make this explanation impossible for the kind.

28 See Thomson, *Normativity*, for discussion of goodness-fixing kinds and an argument supporting this move from judgments of being better and worse *qua* *K* to *K* norms.

29 For example, none of the couples in *Obergefell* reported that they wanted to have the state recognize that they were already married; the fight for gay marriage was the fight to recognize the significance of same-sex partnerships by permitting them to be marriages.

rather than as conveying a distinct kind-status that one can gain in addition to the social kind. Though legal recognition is just one way that social groups could choose to determine spousal-recognition conditions, it seems that the way one gains the social kind Spouse *here* is by undergoing certain legally recognized and sanctioned events.<sup>30</sup>

Just as someone does not need to be any good (or even intend to be good) according to the norms of spouses to become one, spouses also cannot lose their spouse-hood in virtue of becoming a bad enough spouse. Divorce and death: those are the ways to cease being a spouse. Divorce is just the term we have for the way in which two spouses dissolve a marriage through social recognition. It is this dissolution of the social recognition, itself multiply realizable, that constitutes loss of kind-membership.<sup>31</sup> Given the need for the retraction of social recognition (or possibly death), extreme defect would only be existentially risky for very bad spouses if extreme defect *was* retraction of social recognition. However, defect just is not the sort of thing that could be retraction of social recognition. Extreme spousal badness could be *grounds* for a marriage to be dissolved, but mistreatment or abandonment it is not *itself* a divorce. Even if we thought that having the intention to divorce your spouse were a way of being a bad spouse, the intention to divorce is insufficient (and given the possibility of dissolution of marriage by death, unnecessary) to cease being a spouse. The completed divorce, not the intention that brings it, dissolves a marriage.<sup>32</sup>

The role that normative thresholds supposedly play in providing reasons is also missing in the spousal case. Among all the reasons you might take yourself to have for being a good spouse, you would never consider that, in being a bad spouse, you might end up automatically divorced. We take spouses to have the right to determine, for themselves, just how bad a spouse they will tolerate. Surely, some people might have the Lynch coffee attitude toward spouses: having a

30 Even in places where common-law spouses are possible, one does not become a spouse *merely* by acting as a spouse does; these cases are possible *because* one way of having social recognition in these places is by recognizing common-law marriages. If acting as a spouse were sufficient to be a spouse, independent of recognition by a larger social body, then many same-sex couples could have been accurately described as having their spousal status *recognized*, rather than constituted, when they first legally married, but this is not what happened.

31 I am tempted to also say that death is a way to stop being a spouse independently of the dissolution of social recognition, but it seems that in some social groups one can still be a spouse when widowed. It might be that in social groups like ours in which death is a way to cease being a spouse, it is a socially recognized way of loss of kind-membership.

32 Even in cases like Catholic marriages in which the conditions for successful recognition include certain intentions, like to raise one's children in the church, if it is later demonstrated that these conditions were not met, this does not dissolve the marriage. Such demonstrations (i.e., annulments) establish that a marriage never occurred in the first place.

very bad spouse might be better than none at all. It seems implausible that such people might end up in a position in which, despite the interests of both to maintain their (very bad) marriage, they find themselves not spouses.

We thus have a kind, Spouse, that is norm governed but whose norms do not constitute threshold conditions on kind-membership. In at least some cases, then, individual persistence conditions do not seem metaphysically tied to normative conditions. Because of cases like this, Threshold Constitutivists are faced with a choice: narrow the scope of constitutivist explanation to exclude the norms of kinds like Spouses, or reject the Threshold Commitment.

The former option should not be taken lightly. One important standard of explanations is how well they unify the various phenomena to be explained and by how well their argumentative patterns can be used to explain a wide variety of phenomena.<sup>33</sup> Thus, a constitutivist explanation that is able to explain the norms of Spouses is thereby better than one that is unable to do so. Moreover, given that one of the promises of constitutivism is its seeming ability to provide a general explanatory account of the norms of any norm-governed kind by appeal to its constitutive features, constitutivists should be averse to any commitment that restricts the project's explanatory scope.

#### 4. THE COST OF COUNTERINTUITIVE PERSISTENCE CONDITIONS

Unlike social kinds like Spouses, physical artifacts make particularly striking examples of cases in which defect seems to lead pretty straightforwardly to destruction. It is not surprising then that they are often used to illustrate the Threshold Commitment. Houses are commonly used to illustrate the metaphysical risk that defect poses. Korsgaard, for instance, uses a discussion of defective houses in one of her clearest discussions of the Threshold Commitment:

On this view, to be an object, and to be teleologically organized, are one and the same thing. Teleological organization is what unifies what would otherwise be a *mere heap* of matter into a particular object or particular kind. . . . At the same time, it is the teleological organization or form of the object that supports normative judgments about it. A house with cracks in the walls is less good at keeping the weather out, less good at sheltering, and therefore a less good house. The ancient metaphysical thesis of the identification of the real with the good follows immediately from this conception, for this kind of badness eventually shades off into *literal* dis-

33 See, e.g., Friedman, "Explanation and Scientific Understanding"; Kitcher, "Explanatory Unification" and "Explanatory Unification and the Causal Structure of the World."

integration. A house with enough cracks in the walls will crumble, and cease to be a house altogether: it will disintegrate back into a *mere heap* of boards and plaster and bricks.<sup>34</sup>

In endorsing what she calls the “ancient metaphysical thesis of the identification of the real with the good,” Korsgaard identifies being a house with minimal house goodness. It is a particularly compelling example: it is certainly true that the house’s very *existence* is increasingly threatened by mounting normative failure. The case is supposed to provide a clear illustration of the Threshold Commitment because in the case of artifacts loss of kind-membership (or “the real”) is a matter of physical destruction. Because badness is supposed to “shade into” physical destruction, being sufficiently bad is a way of becoming non-real, as the Threshold Commitment says.

However, this is a mistake. Though the house case is used to illustrate the Threshold Commitment, this illustration turns on misinterpreting a causal connection as a constitutive one. A house with enough cracks in the walls is very likely to crumble, but it is the crumbling, rather than the badness, that shades into *literal* disintegration. Considering a case of a house at the brink of extreme defect can help us see why artifacts should not be used to support the Threshold Commitment.

*The Very Bad House:* A Very Bad House is vacant and has been left to the elements. By time  $t_1$ , it is on the brink of physical collapse. Let us grant that there is some normative threshold for houses: there is at least one way of becoming so bad according to the norms of houses that any additional normative failure would constitute loss of household. The Very Bad House, by stipulation, sits on the house-side of the threshold at  $t_1$ . If the house were to become any worse, it would pass over the normative threshold and the resulting arrangement of bricks and shingles would not be a house or would be a “house” in name only.

At  $t_2$  two things happen. First, the Very Bad House fails according to yet another house norm (or norms), which constitutes it passing below the normative threshold.<sup>35</sup> However, simultaneously at  $t_2$ , a witch on a

34 Korsgaard, *Self-Constitution*, 28.

35 It is not important to the threshold account whether nonbeing requires normative failure according to one or to many norms. The important thing is that the threshold theorist thinks that normative failure *alone* can make the metaphysical difference. Threshold defenders might claim that the connection between badness and existence is vague, making the threshold vague too. This should not affect the argument. So long as the metaphysical difference is only a matter of normative failure, a case can still be produced by finding some extreme amount of defect that would be sufficient to surpass the vague threshold. For the

conveniently timed walk notices the house's condition. Fearing the worst, she casts a spell causing a strong wind to blow in just the way necessary to keep the house standing.<sup>36</sup> Unlike unsupported structures, which would collapse in the house's condition at  $t_2$ , the Very Bad House has been *physically* preserved.

In order to maintain the identification of the real and the good, the Threshold Constitutivist seems committed to admitting that houses can cease to be real without being physically destroyed. However, it is the connection between physical destruction and reality in the case of artifacts that makes them such compelling examples. To maintain the intuitive force of these cases, the Threshold Constitutivist needs to account for some property that could explain how, contrary to the description in the example, the Very Bad House remains minimally good at  $t_2$  when the wind prevents it from falling. However, I will show that there are no plausibly good-making properties that will ensure the Very Bad House will remain standing at  $t_2$ .

At  $t_2$  the Very Bad House gains the property of being held up by a strong wind.<sup>37</sup> This new property does not seem to be a property that makes a house better *qua* house.<sup>38</sup> More plausibly, the Threshold Constitutivist will think that being held up by the wind preserves some other property of the Very Bad House between  $t_1$  and  $t_2$  that keeps it from becoming worse. Their task is then to show that there is some property of the Very Bad House that (a) the house has at  $t_1$ , (b) is maintained by the wind at  $t_2$ , and (c) is good-making at  $t_2$ . Moreover, in order to preserve the connection between nonbeing and physical destruction, it must (d) be a property that, if lost, would constitute the Very Bad House's physical destruction.

Three properties might meet these conditions: being standing, stability, and having the ability to shelter. The property most obviously suited to (d) is *being standing*. Being standing is clearly related to physical persistence. It is also clearly a property that the wind preserves between  $t_1$  and  $t_2$ . However, it is *too closely* related to physical persistence for it to be good-making for houses. Because the connection between defect and nonbeing is a physical matter in the case of ar-

---

case at hand, all that matters is that the difference between the house at  $t_1$  and  $t_2$  can be explained in terms of normative satisfaction.

36 Or if you prefer, a *very* quick builder on a similar walk erects temporary scaffolding, finishing precisely at  $t_2$ .

37 The Threshold Constitutivist might also think that the house has the property of being bewitched, but the weather is bewitched, not the house.

38 To see this, consider that for a property  $p$  to be a good making property for  $K$ s, it must be that if having  $p$  makes  $X$  a better  $K$ , for any other  $K$ ,  $Y$ , it must make  $Y$  a better  $K$ .

tifacts, remaining standing is a minimal condition on being a house at all. But if it were a norm of houses that they were standing, it would be a consequence of something being a house that it satisfied this norm; i.e., it would run afoul of Violability. So, the fact that the Very Bad House remains standing cannot account for it remaining minimally good at  $t_2$ .<sup>39</sup>

Though remaining standing cannot itself be good-making, it could be that, in addition to remaining standing, the wind keeps the Very Bad House *stable* at  $t_2$ . Unlike remaining standing, stability seems gradable, avoiding an immediate source of conflict with Violability. Even better, stability seems plausibly good-making for houses. However, it is not clear that the house at  $t_2$  does have a relevantly good-making property of stability.

Stability is a dispositional property whose manifestation response is standing. Though the manifestation response cannot be good-making, the disposition itself seems to be. Problematically, however, stability is an unusual disposition, because its manifestation response is not change but stasis.<sup>40</sup> Fragility is the disposition, say, to break when struck—that is, to undergo a certain change under certain stimulus conditions. But stability is a disposition to maintain orientation or structural form given environmental changes—that is, to maintain properties under certain stimulus conditions. Given this and the connection between reality and the manifestation response of this disposition, it is harder to account for how it could be a good-making property the Very Bad House has at  $t_2$ .

To see why, consider a distinction made by Vetter, who notes that possessing some dispositions (threshold dispositions) requires having the manifestation response in some proportion of cases, while possessing other dispositions (permissive dispositions) only requires having the manifestation response in a single case.<sup>41</sup> Breakability is a permissive disposition; in order to be breakable, there need only be one condition in which an object would break. Fragility is a threshold disposition; in order to be fragile, an object must be *easily* breakable—there must be some sufficiently large number of cases in which it would break. So, we can ask, is stability a threshold disposition, like fragility, or a permissive disposition, like breakability?

One way to easily guarantee that the Very Bad House has dispositional stability maintained between  $t_1$  and  $t_2$  is to understand it as a permissive disposition, always had by things that are standing. In keeping the house standing, the wind

39 We might also put the point this way: remaining standing in this case is a binary property that is a criterial condition on kind-membership. It is bound to be ruled out on grounds discussed in section 1, above.

40 See Williams, "Static and Dynamic Dispositions."

41 Vetter, "Dispositions without Conditionals," 144.

preserves the permissive disposition of stability. However, given the connection between physical persistence and reality, all houses are standing and thus all houses are dispositionally stable. So permissive stability runs afoul of Violability like being standing does.

On the other hand, stability could be a threshold disposition. Then there will be some conditions in which objects have the manifestation response of stability (remain standing) without having the disposition (stability). This would allow there to be non-stable houses, which resolves the problem with Violability. Unfortunately, understanding the relevant sense of stability as threshold stability undermines the plausibility that stability is preserved by the wind at  $t_2$ . If there are any cases in which a standing house does not have the threshold disposition of stability, the Very Bad House at  $t_2$  would be one. In fact, the example of the Very Bad House is designed so that it describes an instance of a standing thing that lacks the threshold disposition of stability. Either way of understanding dispositional stability seems unable to serve the Threshold Constitutivist's purposes of being a property that is both good-making and preserved by the wind between  $t_1$  and  $t_2$ .

The final potentially preserved property of the house, the *ability to shelter*, does no better than remaining standing or stability. Whatever this sheltering ability is, if it is to be a good-making property of houses, it cannot be a feature had by any standing house. So there must be some ways of preserving the physical structure of a house while not preserving the good-making ability to shelter. That, by hypothesis, is what is being done by the wind at  $t_2$ . Of course, there might some ability to shelter that all non-destroyed houses possess, but this cannot be a good-making ability. The issue here, as with the other properties, is that defect cannot shade into destruction by shading into *physical* destruction. If Violability is true and being non-destroyed is a requirement on the real, then no property that is had by all non-destroyed  $K$ -members can be good-making for  $K$ s.

The Threshold Constitutivist must instead accept that, despite the fact that it is not physically destroyed, the Very Bad House is not a house at  $t_2$ . There could be some explanation of the kind House that would make this counterintuitive result more palatable. Regardless of whether such an explanation can be provided, the issue remains that the appeal to physical artifacts to support the identification of the real and the good is misguided. Appeals to cases of physical artifacts, like houses, exploit the fact that defect can lead to physical destruction in a mistaken attempt to illustrate that defect can itself constitute nonbeing. But defect does not entail physical destruction, so if defect *does* entail nonbeing, rather than being the most plausible illustrations of the Threshold Commitment,

physical artifacts actually serve as rather counterintuitive cases for proponents to explain.

##### 5. THE COST OF NORMATIVE ANEMIA

In the case of Houses, Threshold Constitutivists hoped that the metaphysical significance of destruction and the relation of defect and destruction could make the Threshold Commitment plausible. But we have seen that preserving the Threshold Commitment requires normative failure *itself*, and not the destruction that it often leads to, to make the metaphysical difference. In the case of Houses, this comes at the cost of accommodating some counterintuitive results. In the case of Spouses, it comes at the cost of restricting the explanatory scope of Threshold Constitutivists' explanatory project. But it comes with another cost: giving norms this metaphysical significance undermines their other normative roles.

For the Threshold Constitutivist, normative standards are not merely the basis for normative evaluation; they also determine the criterial conditions for kind-membership. As Korsgaard puts it, an "object to which they apply can fail to meet them, at least to some extent, and is subject to criticism if it does not."<sup>42</sup> This distinguishes threshold constitutivism from other constitutivist accounts that locate the metaphysical criteria of kind-membership elsewhere. The Threshold Constitutivist thereby holds that there are some events that constitute both a violation of a *K*-norm by some *K*-member, *x*, and the loss of *K*-membership by *x*, and the violation constitutes rather than causes or precipitates the kind-loss.

Constitutivists generally share a commitment to a fundamental role for kind-determined norms in evaluation. According to constitutivists, constitutively understood norms come from the nature of the thing they govern.<sup>43</sup> They thus yield a special sort of internal evaluation that is distinguished from evaluation according to some external purpose. Because of this, constitutively understood norms generate kind-dependent assessments that permit us to, e.g., distinguish the standards that make something good or bad *qua* house from those that make it good or bad *qua* thing in the neighborhood, status symbol, or instance of modern architecture.<sup>44</sup> Constitutively understood norms are thus kind-specific

42 Korsgaard, *The Constitution of Agency*, 8.

43 For a discussion of goodness-fixing kinds and kind-defect/virtue, see Thomson, *Normativity*.

44 For example, Korsgaard distinguishes between being a good or bad *house* in what she calls "the strict sense" and being a house that happens to be a good or bad *thing* "for some external reason." The first strict sense is determined by "constitutive standards," "standards that apply to a thing simply in virtue of its being the kind of thing that it is." What Korsgaard calls "constitutive standards," I am here calling norms. Korsgaard, *Self-Constitution*, 28.

norms that are tied to a particular and important kind of normative status: being good or bad *qua* kind-member. There are two aspects of this evaluative role of norms: first, that defects and virtues are kind-dependent and second, that norm violation and satisfaction by kind-members have kind-relative evaluative consequences.

A noteworthy consequence of the Threshold Commitment is that norm violation of kind-members does not always lead to kind-relative evaluative consequences. The Threshold Constitutivist denies that for any *K*-member, *x*, governed by a *K*-norm, *N*, *x* would be defective if *x* violated *N*. They must, because on their account there will be some *x*'s that, when they violate *N*, do not become a worse *K*, but become a non-*K*. Of course, constitutive norms pick out good- and bad-making features of kind-members, given the nature of the kind in question.<sup>45</sup> If you satisfy a *K*-norm and you are a *K*-member, then the property you have that constitutes your satisfaction of the *K*-norm is good-making. But more than this, according to constitutivists, *K*-norms have particular *force* for *K*-members that they lack for non-*K*-members. This is how Korsgaard claims constitutivists can answer skeptical challenges with ease.<sup>46</sup> When asking why we ought to care about the norms that govern us, the fact that we are at risk of falling apart if we become bad enough is supposed to show that those norms have authoritative force for us; we cannot help but care about them.

But the general normative force of norms cannot turn on the kind-members in question *caring* about the risk of nonexistence. Norms cannot, in general, have force for kind-members they govern because the kind-members care about something else that satisfying the norms is essential for (how, for example, would that account for the force of house norms; houses care for nothing). Norms must have force because of their connection to their kinds, and that is what the constitutive connection is supposed to provide. Somehow the fact that sufficient defect is existentially risky is supposed to give norms their force. On the contrary, however, this existential risk seems to undermine the account of their force. If a particular house being a *house* turns on its being waterproof, if it were at risk of becoming a non-house by becoming leakier, it is hard to see how the demand to be waterproof itself could then have any normative force. The demand loses its normative bite as soon as it is violated.

45 See, for example, Korsgaard, *Self-Constitution*, 33, in which she argues that we need constitutive standards because of the importance of the normative concept of *defect*.

46 "Because it does not make sense to ask whether a house should serve as a shelter, it also does not make sense to ask if the corners should be sealed and the roof should be waterproof and tight ... there is no further room for doubting that the constitutive standard has normative force. For if you fall too far short of the constitutive standard, what you produce will simply not be a house." Korsgaard, *The Constitution of Agency*, 29.

At these thresholds, we can see the weakness, the anemia, of the metaphysical account. Taking minimal normative satisfaction to be criterial for kind-membership makes that kind-membership unsuited to explain the force of those norms. The force of norms for some individual cannot be explained by the fact that they are currently satisfying some subset of them. At the limit case, this would mean that a norm has force because it is being satisfied. This account does not give internal standards any greater authority than the external standards that constitutivists want to distinguish constitutively understood standards from. Taking normative standards to be both criterial and normative undermines the explanation of their normative force. Not only does the Threshold Commitment come with explanatory costs related to scope and intuitive persistence conditions, it also seems to block central explanations of features like normative force.

#### 6. CONSTITUTIVE AIMS DO NOT AVOID THRESHOLDS

Above, I argued that the Threshold Commitment is a flawed account of a constraint on norm-governed kind-membership. Constitutivists should thus recognize a need for reconsideration of the metaphysical basis of constitutivist explanations.<sup>47</sup> However, though many constitutivists like Katsafanas and Korsgaard explicitly endorse something like the Threshold Commitment and it seems implied in other accounts, some constitutivist accounts that explain constitutive features in terms of constitutive aims might already seem to have resources to avoid the Threshold Commitment. To take a brief example, if truth is the constitutive aim of belief, this aim might be used to explain the norms governing beliefs, though no particular belief must be even minimally true to account for it having the constitutive aim. Because *having* a constitutive aim is neither good-making nor scalar, constitutive-aim accounts might already avoid the Threshold Commitment, making further development unnecessary. Despite this appeal, constitutive aims are no more suited than other constitutive properties to avoid the Threshold Commitment, or so I argue in this section.

Constitutive aims will be suitable for this purpose if the account of how kind-members have the aim both avoids requiring minimal kind-goodness and also provides the metaphysical grounds for normative explanations. I first consider whether it is possible to give an account of aiming that any action (or—more broadly—any norm-governed kind) could be understood as having. Then I discuss two views that provide explanations of how kind-members have constitutive aims: Velleman's view that self-knowledge is the constitutive aim of action and Wedgwood's view that truth is the constitutive aim of belief. Nei-

47 Thanks to an anonymous referee for highlighting the need for the discussion in this section.

ther account, I argue, provides a metaphysical account that avoids the Threshold Commitment while providing the resources to ground normative explanations. I conclude that these failures show that having a constitutive aim is not going to provide a satisfactory way to avoid the Threshold Commitment.

Let us start with a commonly discussed example: one might hold that it is constitutive of playing chess to have the aim of checkmating your opponent, though one of course need not minimally succeed at checkmating an opponent, whatever that might be, or be minimally good according to the rules of chess. A constitutivist account of the norms of chess that appealed to this aim would need an explanation of what metaphysical property (or properties) constitutes having this aim. If such an explanation could be given without identifying having the aim with minimal normative properties, and this explanation could then be used to explain chess norms, it would be a successful constitutivist account that avoided thresholds.<sup>48</sup> I think constitutive aims are unlikely to be accounted for by properties that could play this role for constitutivists.<sup>49</sup>

In the case of chess playing, the most straightforward way to account for having its constitutive aim might be by appeal to the intentions of the player. Whatever metaphysical account we might give of *intending* to checkmate might be used to account for having the aim of checkmating. Putting aside the fact that the metaphysics of intentions are controversial, accounting for constitutive aims by appeal to intentions is unlikely to work for constitutivist accounts of the norms of action generally because it does not generalize to the central case of agency. If the constitutive aim of agency requires having an intention, which only agents can have, this introduces problematic bootstrapping: the source of the constitutive feature is the very thing constituted by that feature. The ability to have the constitutive aim seems to presuppose already having it.<sup>50</sup>

Velleman recognizes this bootstrapping problem and attempts to avoid it by locating the source of the constitutive aim of action subagentially. In what follows, I explain how Velleman's account of the constitutive aim of action does implicitly rely on the Threshold Commitment. On his account, the constitutive aim of action is self-knowledge, but an action gets this aim not from an agential intention, but from the agent's subagential desire to know what she is doing. The

48 Importantly, I think there *can* be successful threshold-free constitutivist accounts. My goal here is only to show that constitutive-aim accounts are not the place to look for them.

49 Chess is perhaps not as good a case as constitutivists sometimes assume. It seems unlikely that we will be able to explain the norms of chess by appeal to the aim of chess, regardless of whether we can identify a property that constitutes having the aim of checkmating; see Dreier, "When Do Goals Explain the Norms that Advance Them?"

50 See, for instance, Arruda, "Constitutivism and the Self-Reflection Requirement."

intellect has the aim of knowledge, and a special case of that aim is self-knowledge. Importantly, in acting, we have access to non-observational self-knowledge of what we are up to. What distinguishes action from mere behavior, then, is that in acting we make ourselves the authors of our actions. In order to do this, the intellect takes the self as a subject, and so the constitutive aim of action is “the aim of our intellects as focused on ourselves, the aim to which practical knowledge is the obvious shortcut, the aim of knowing what we are doing.”<sup>51</sup> To explain this, Velleman considers a case of mere behavior: Freud’s account of knocking over an inkstand and realizing, after the fact, that he knocked it off his desk because he desired a new one, and, believing his sister would buy him one as a present, he wished to make room for it. Velleman writes that in order for the mere behavior to have counted as an action Freud “would need to have been actuated not only by the desire and belief mentioned in the story but also by the story itself, serving as his grasp of what he was doing—or, in other words, as his rationale.”<sup>52</sup> So, roughly: the behavior lacked the constitutive aim of action because the behavior was not motivated by a rationale, something that would have provided practical knowledge when acted on.

This role for the rationale introduces thresholds into the account of how sub-agential desires constitute aims. To constitute action, the desire cannot be merely lurking subconsciously; it has to manifest itself in the sort of grasp we have on what we are up to when we act. In order to act, Velleman believes an agent “would need, first, to have been inhibited from acting on his desire and belief until he knew what he was up to; and then guided to act on them once he had adopted this story. He would then have acted autonomously because he would have acted for a reason, having been actuated in part by a rationale.”<sup>53</sup> So, it is essential for action, on Velleman’s constitutive-aiming account, that one must first know what one is up to (or, perhaps, what one will be up to), and then, as behavior is guided by that account as a rationale, one thereby acts for reasons provided by that rationale. The reasons that one has to do something are “considerations in light of which, in doing it, the subject would know what he was doing.”<sup>54</sup> Behavior thus constitutes action when that behavior is guided by reasons, understood as considerations that provide self-knowledge.

But rationales—explanations of what we are up to—can be better or worse at providing self-knowledge. The less self-knowledge they provide, the worse they are as rationales. As Velleman writes, “when an agent selects rationales that

51 Velleman, “Précis of *The Possibility of Practical Reason*,” 236.

52 Velleman, *The Possibility of Practical Reason*, 29.

53 Velleman, *The Possibility of Practical Reason*, 8.

54 Velleman, *The Possibility of Practical Reason*, 26.

are incongruous with who they are, they ... might have wondered 'what am I doing?' That is, they might have been puzzled as to how a person like them, with a makeup like theirs, would come to act on such motives; and so they wouldn't really or fully have known what they were up to."<sup>55</sup> So not all rationales help us accomplish our aim of self-knowledge equally well. Still, on Velleman's view, in order to count as having the constitutive aim of self-knowledge, you already have to partially know what you are up to, i.e., you have to already have some self-knowledge that permits you to know what you are doing. Thus, for Velleman, having a constitutive aim requires minimally satisfying that aim. So, constitutive aiming, at least on Velleman's view, does not avoid the Threshold Commitment at all.

Aims relying on agent intentions do not seem suited to the constitutivist explanatory project, and Velleman's constitutive account relying on subagential aims involves a commitment to Thresholds. One further way we might understand constitutive aims is suggested by discussions of the aims of mental states. For example, many philosophers endorse claims of the form "the constitutive aim of belief is truth," or "belief aims at truth." But these claims are not taken to be in conflict with the possibility of false belief, and they do not seem to think that beliefs have this aim because believers intentionally aim at truth. Perhaps, then, these accounts will be constitutive-aim accounts that already avoid the Threshold Commitment.

Ralph Wedgwood is one prominent defender of the claim that truth is the constitutive aim of belief. Wedgwood explains that when we say belief has the constitutive aim of truth we are not attributing some metaphysical property to all beliefs that accounts for this aim; rather, he endorses a view he calls Normativism, according to which belief is the attitude constituted by having truth as its correctness condition.<sup>56</sup> Similarly we might think that chess has the aim of checkmate in the sense that chess is the game constituted by having checkmate as its success condition. Just as you can actively attempt to believe false things, you can play chess without caring about checkmate or actively throw a game. Maybe you could also act without self-constituting or having self-knowledge.<sup>57</sup> So, according to Normativism, aims are constitutive of some kind when individ-

55 Velleman, *The Possibility of Practical Reason*, 29.

56 See especially Wedgwood, *The Nature of Normativity*.

57 This is also, plausibly, a more charitable understanding of attributions of constitutive aims, because aiming in the literal sense requires knowing beforehand what you are aiming to get, but believers do not first determine what is true and then attempt to believe it, in the way that archers first locate a target and then try to hit it. See Dreier, "When Do Goals Explain the Norms that Advance Them?" 159–60, for this point.

uals are members of that kind in virtue of having that aim as their correctness condition.

Normativism explains what it is for an aim to be constitutive of some kind, but it explains kinds by shared norm-assessability (or shared-correctness conditions), not by shared properties that are used to *explain* their shared norms. So, Normativism does not require any minimal normative satisfaction for constitutive aims and thus avoids the Threshold Commitment. However, it does this by not undertaking the kind of constitutivist explanation we started by motivating: it does not seek to explain the normative features of kinds by appeal to what is constitutive of kind-membership. Instead, it explains what is constitutive of kind-membership by appeal to the correctness conditions shared by kind-members. I do not think this explanatory work is ruled out for Normativists; the view might be paired with a metaphysical account of *why* kind-members had these correctness conditions. But, in this case, the constitutivist explanation of the normative features would not be in terms of constitutive aims.

So, constitutive aims do not seem to help current constitutivists avoid the Threshold Commitment. Constitutive aims that require intentional aimings run into bootstrapping worries, which led Velleman to explain them in terms of sub-agential desires. These are not threshold free, however, as we saw that having behavior rationalized by desire in the way needed for it to have the constitutive aim of self-knowledge required minimal self-knowledge. Alternatively, accounts that explain constitutive aims as correctness conditions avoid normative thresholds, but do not use aims to provide the sort of constitutive explanation under consideration. If the Threshold Commitment is a problem for constitutivism, as I argued above, constitutive aims are not a solution for constitutivists.

#### 7. CONSTITUTIVISM WITHOUT NORMATIVE THRESHOLDS

Though many constitutivists endorse and are committed to the Threshold Commitment, even Threshold Constitutivists do not take it to be an essential feature of constitutivist accounts.<sup>58</sup> The Threshold Commitment is only one account of how norms can be constitutively explained.

58 Some constitutivists have a very narrow view of constitutivism, according to which the only views that count as constitutivist are those that take there to be constitutive aims that generate norms. I suspect that Katsafanas thinks this, given the set-up of his book, in which he classifies both Humeans and Aristotelians who appeal to the nature of individuals to explain their goodness as non-constitutivists. Both Judith Thomson and Michael Smith, constitutivists on my account (and Smith's), are classified by Katsafanas as non-constitutivists. Not all constitutivists are so restrictive, of course. This restriction is fine; it might be that on the narrow use of the term we ought to reject constitutivism on the basis of the arguments

There are a few options open to constitutivists for how norms could be explained in terms of what is constitutive of kind-membership other than making those norms or properties that satisfied them constitutive. We might identify other features as determining constitutive forms or functions (or even aims). Though Korsgaard identifies minimal goodness with form, she also sometimes writes as though performing a function is necessary for having that constitutive function. At times she seems to endorse what we might call a “function-performance” view. For instance, she claims that artifacts must be performing their characteristic activity to count as an artifact. Memorably, she writes that a thing we *call* a vacuum cleaner is really just a heap that “when properly incorporated by you, makes *you* into a vacuum cleaner.”<sup>59</sup> Not only is your vacuum cleaner not really a vacuum cleaner when in the closet, but when you vacuum the floor with your heap non-vacuum *you* become a vacuum cleaner instead of it.

The Threshold Constitutivist requires that the very same properties that make one a good *K*-member are minimally required to be a *K*-member at all. In contrast, a function-performance view holds that minimal performance of the characteristic function of *K*s is required for *K*-membership. If by “minimal performance” we mean “possession of the properties that permit minimal performance,” or if we take performing a function to be good-making, then this is a threshold view, but I take it that it is not always so meant. It can also mean that something only has a functional nature when it actually is performing that function, like the vacuum cleaner is only *actually* a vacuum cleaner when it is vacuuming. If performing the function of vacuuming is what it is to be a vacuum, and you think that the user of the machine rather than the machine is performing the function, you might be tempted, with Korsgaard, to call the human user, rather than the machine, the vacuum. Distinguishing normative satisfaction from functional performance allows us to see two ways we might identify constitutive conditions of kind-membership.

Unfortunately, function-performance views are not going to help the constitutivist, because they mis-account for kind-membership and norm-application.<sup>60</sup> *Pace* Korsgaard, I am not a vacuum, as evidenced by the fact that I am

---

of this paper. At least here, I am using the term in a broader sense. See Katsafanas, *Agency and the Foundation of Ethics*, 30–34. For constitutivists self-identifying outside of the narrow use, see, e.g., Schafer, “Realism and Constructivism in Kantian Metaethics (1)”; and Smith, “A Constitutivist Theory of Reasons.” This identification of constitutive aim accounts with constitutivism is also implicit in some literature critical of constitutivism; see, e.g., Enoch, “Agency, Shmagency.”

59 Korsgaard, *Self-Constitution*, 37.

60 For further arguments to this effect, see Lindeman, “Etiological Functions for Constitutivists.”

not assessable according to the norms of vacuums. What the Korsgaard-esque function-performance view calls a heap *is* properly speaking my vacuum cleaner—and I suspect you and the store that sold it to me agree. Things that are not performing their functions are still members of functionally understood kinds. Additionally, making anything that serves a function a member of a functional kind (as Korsgaard implies by calling *me* the vacuum cleaner when I use it) risks erasing the distinction between being badly suited to perform a function and being *defective*. That a piece of aluminum siding is used to saw down a sapling does not make it a saw, and that it is poorly suited to the task does not make it defective. For constitutivists, kind-membership conditions just cannot involve actually performing a function. Constitutivists thus need another account of the constitutive features of kind-membership that can be used to explain norms.

In closing, I suggest that constitutivists should turn to proper functions to provide this account.<sup>61</sup> In the philosophy of biology, a *proper function* is the kind of function that something is taken to have non-accidentally, as a feature of what that thing is.<sup>62</sup> The question, of course, is how proper functions are constitutive of kind-membership or, put another way, what it is in virtue of which any individual has a proper function. There are several requirements on such an account: it must allow Violability by not entailing that any particular normative standard is satisfied, and also not require minimal normative goodness. It must allow for stable kind-membership and for explanations of internal normative standards as well as kind-relative defect and virtue.

Etiological accounts of proper function—those that appeal to the history of individuals and kinds to account for function—are taken to have made teleology safe for naturalists in the sciences by linking functions to evolutionarily determined kinds.<sup>63</sup> Consider the following quote from Millikan as she explains

61 Proper functions are sometimes contrasted with what are known as systems functions, those that give an account of function of some *x* relative to the causal contribution *x* plays to the broader capacity of a system in which *x* can be understood as being a part. See Cummins, “Functional Analysis,” for discussions of systems functions; see Allen, Bekoff, and Lauder, *Nature’s Purposes*, for an overview of this distinction.

62 Millikan stresses that her coinage of “proper” is meant to mirror its etymological ancestor, the Latin *proprius*, meaning “one’s own”; see Millikan, “Biofunctions,” 116.

63 Though I do not address alternative accounts of function in this chapter, for a nice overview of the differences between etiological and propensity accounts of function, see Mitchell, “Dispositions or Etiologies?” For presentations and criticisms of the propensity view of functions, according to which the function of an individual or trait depends on how it would fare under selection in some specific environment, see, e.g., Bigelow and Pargetter, “Functions.”

how the etiological proper function of pumping blood accounts for what it is to be a heart:

That a heart is a heart certainly has something to do with pumping blood. But what kind of connection with pumping blood must a heart have? Some hearts are diseased and some are malformed in such a way that they are unable to pump blood. Other devices, such as water pumps, are perfectly capable of pumping blood, yet these are not hearts.... It is not then the actual constitution, powers, or dispositions of a thing that make it a member of a certain biological category. My claim will be that it is the “proper function” of a thing that puts it in a biological category, and this has to do not with its powers, but with its history.<sup>64</sup>

In addition to threshold accounts and function-performance accounts, the constitutivist also has available to her etiological accounts. The account of the history or creation of an individual could explain what it is to have a constitutive function. Having an etiological proper function neither entails that the individual is any good at performing the function nor that they actually perform the function, and so it is not a threshold or a function-performance view. It could, however, account for kind-membership and normative standards governing kind-members. Moreover, etiological accounts need not restrict selection accounts to evolutionary selection.<sup>65</sup> For example, in the cases of both House and Spouse, histories of individuals are relevant to their kind-membership. In determining whether a person is a spouse, you look to that person’s history. Spouses have their kind-status because of historically determinant events: marriages or similar social-recognition events. Houses are given the forms they have by builders who select those forms because of their connection to performing the function of sheltering people in ways suited to primary residences. Houses are the things they are (i.e., structures for primary residence) because of their forms being selected and designed for that purpose. In both cases, we see the historical accounts of individual kind-members as simultaneously explaining their functions and their kind-membership.

The task of fully developing and defending a specific account of proper functions congenial to constitutivists is beyond the scope of this paper, but the

64 Millikan, “In Defense of Proper Functions,” 85.

65 In fact, even in biology proper function accounts should not be limited to evolutionary selection. See Garson, “Function, Selection, and Construction in the Brain,” for a fascinating discussion of the need for non-evolutionary differential-retention accounts to account for proper functions of neural structures, allowing for novel functions unaccountable for on an evolutionary scale, like the function of facilitating Tetris play.

substantive lessons here should be heartening to those interested in developing constitutive explanations of norms as well as those who are interested in seeing why current constitutive accounts are unsatisfactory. Specifically, norms cannot play the metaphysical role of criteria for kind-membership that Threshold Constitutivists have given them. Constitutivists who give such a criterial role to norm satisfaction must restrict the scope of their constitutive explanations, are pushed to accept implausible permanence conditions for artifacts, and have restricted evaluative consequences of normative failure that block explanations of normative force.

Rather than this being a blow to the constitutivist project, this result should instead encourage constitutivists to look elsewhere for an account of the constitutivist features of good-making kinds with which to explain their norms. I briefly sketched a promising alternative on which functional kind-membership is based in historical facts. Etiological proper functions are available to constitutivists seeking an account of the constitutive features of good-making kinds from which they can explain norms. Clearly there is work to be done to show that an etiological account of proper function could be given for the kinds of interest to constitutivists in the practical and theoretical domains. Nonetheless, given the dim prospects of the alternatives and the general appeal of constitutivism, constitutivists should welcome the development of a constitutivist account based on etiological proper functions.<sup>66</sup>

*Saint Louis University*  
kathryn.lindeman@slu.edu

#### REFERENCES

- Allen, Colin, Marc Bekoff, and George V. Lauder, eds. *Nature's Purposes: Analyses of Function and Design in Biology*. Cambridge, MA: MIT University Press, 1998.
- Arruda, Caroline T. "Constitutivism and the Self-Reflection Requirement." *Philosophia* 44, no. 4 (December 2016): 1165–83.

66 This paper has benefitted substantially from discussions and helpful feedback from many, many people. Special thanks are due to Sarah Buss, Eric Brown, Bob Brandom, John Brunero, Lindsay Crawford, Jamie Dreier, Julia Driver, Meghan Page, Luca Ferrero, Guy Fletcher, Kim Frost, Charlie Kurth, Eliot Michaelson, Claire Morrissey, Tyke Nunez, Hille Paakunainen, Michael Ridge, Connie Rosati, Lizzie Schecter, Karl Schafer, Mark Schroeder, Tim Schroeder, Kieran Setiya, Daniel Singer, Paulina Sliwa, Matthew Smith, Michael Smith, Julia Staffel, Daniel Star, Brian Talbot, Michael Thompson, Katia Vavova, Eric Wiland, an audience at the Midsummer Philosophy Workshop, and several very helpful referees.

- Barandalla, Ana, and Michael Ridge. "Function and Self-Constitution." *Analysis* 71, no. 2 (April 2011): 364–80.
- Bigelow, John, and Robert Pargetter. "Functions." *Journal of Philosophy* 84, no. 4 (April 1987): 181–96.
- Clark, Philip. "Velleman's Autonomism." *Ethics* 111, no. 3 (April 2001): 580–93.
- Cummins, Robert C. "Functional Analysis." *Journal of Philosophy* 72, no. 20 (November 1975): 741–65.
- Dreier, James. "When Do Goals Explain the Norms that Advance Them?" In *Oxford Studies in Metaethics*, vol. 5, edited by Russ Shafer-Landau, 153–74. Oxford: Oxford University Press, 2010.
- Enoch, David. "Agency, Shmagency: Why Normativity Won't Come from What Is Constitutive of Agency." *Philosophical Review* 115, no. 2 (April 2006): 169–98.
- . "Shmagency Revisited." In *New Waves in Metaethics*, edited by Michael Brady, 208–33. Houndmills, UK: Palgrave Macmillan, 2011.
- Friedman, Michael. "Explanation and Scientific Understanding." *Journal of Philosophy* 71, no. 1 (January 1974): 5–19.
- Garson, Justin. "Function, Selection, and Construction in the Brain." *Synthese* 189, no. 3 (December 2012): 451–81.
- Katsafanas, Paul. *Agency and the Foundations of Ethics: Nietzschean Constitutivism*. Oxford: Oxford University Press, 2013.
- Kitcher, Philip. "Explanatory Unification." *Philosophy of Science* 48, no. 4 (December 1981): 507–31.
- . "Explanatory Unification and the Causal Structure of the World." *Minnesota Studies in the Philosophy of Science* 13 (1989): 410–505.
- Korsgaard, Christine M. *The Constitution of Agency*. Oxford: Oxford University Press, 2008.
- . *Self-Constitution: Agency, Identity, and Integrity*. Oxford: Oxford University Press, 2009.
- Lavin, Douglas. "Practical Reason and the Possibility of Error." *Ethics* 114, no. 3 (April 2004): 424–57.
- Lindeman, Kathryn. "An Explanation of Constitutivist Normativity." Unpublished manuscript.
- . "Etiological Functions for Constitutivists." Unpublished manuscript.
- Millikan, Ruth Garrett. "Biofunctions: Two Paradigms." In *Functions: New Essays in the Philosophy of Psychology and Biology*, edited by André Ariew, Robert Cummins, and Mark Perlman, 113–43. Oxford: Oxford University Press, 2002.
- . "In Defense of Proper Functions." In *White Queen Psychology and Other Essays for Alice*, 13–29. Cambridge, MA: MIT University Press, 1993.

- Mitchell, Sandra D. "Dispositions or Etiologies? A Comment on Bigelow and Pargetter." *Journal of Philosophy* 90, no. 5 (May 1993): 249–59.
- Railton, Peter. "On the Hypothetical and Non-Hypothetical in Reasoning about Belief and Action." In *Ethics and Practical Reason*, edited by Garrett Cullity and Berys Gaut, 53–79. Oxford: Clarendon Press, 1997.
- Rosati, Connie S. "Agents and 'Shmagents.'" In *Oxford Studies in Metaethics*, vol. 11, edited by Russ Shafer-Landau, 182–213. Oxford: Oxford University Press, 2016.
- Schafer, Karl. "Realism and Constructivism in Kantian Metaethics (1): Realism and Constructivism in a Kantian Context." *Philosophy Compass* 10, no. 10 (October 2015): 690–701.
- Setiya, Kieran. "Akrasia and the Constitution of Agency." In *Practical Knowledge: Selected Essays*, 253–71. Oxford: Oxford University Press, 2016.
- . "Explaining Action." *Philosophical Review* 112, no. 3 (2003): 339–93.
- . *Reasons without Rationalism*. Princeton, Princeton University Press, 2007.
- Silverstein, Matthew. "The Shmagency Question." *Philosophical Studies* 172, no. 5 (May 2015): 1127–42.
- . "Teleology and Normativity." In *Oxford Studies in Metaethics*, vol. 11, edited by Russ Shafer Landau, 214–40. Oxford: Oxford University Press, 2016.
- Smith, Michael. "A Constitutivist Theory of Reasons: Its Promise and Parts." *Law, Ethics, and Philosophy* 1 (2013): 9–30.
- Thomson, Judith Jarvis. *Normativity*. Chicago: Open Court Press, 2008.
- Tiffany, Evan. "Why Be an Agent?" *Australasian Journal of Philosophy* 90, no. 2 (June 2012): 223–33.
- Velleman, J. David. *The Possibility of Practical Reason*. Oxford: Oxford University Press, 2000.
- . "Précis of *The Possibility of Practical Reason*." *Philosophical Studies* 121, no. 3 (December 2004): 225–38.
- . "Replies to Discussion on *The Possibility of Practical Reason*." *Philosophical Studies* 121, no. 3 (December 2004): 277–98.
- Vetter, Barbara. "Dispositions without Conditionals." *Mind* 123, no. 489 (January 2014): 129–56.
- Walden, Kenneth. "Laws of Nature, Laws of Freedom, and the Social Construction of Normativity." In *Oxford Studies in Metaethics*, vol. 7, edited by Russ Shafer-Landau, 37–79. Oxford: Oxford University Press, 2012.
- Wedgwood, Ralph. *The Nature of Normativity*. Oxford: Oxford University Press, 2007.
- Williams, Neil E. "Static and Dynamic Dispositions." *Synthese* 146, no. 3 (September 2005): 303–24.

## CAN OBJECTIVISTS ACCOUNT FOR SUBJECTIVE REASONS?

Daniel Wodak

THE DISTINCTION BETWEEN objective and subjective reasons is quite intuitive, in part because the two seem to play different roles in normative thought. If a blue pill would cure Anna's disease, there is an objective reason for her to take it. If Anna believes that only the red pill would cure her, there is a subjective reason for her to take it. And if Anna received misleading evidence that taking the red pill causes nausea, there would be a less weighty subjective reason *not* to take it. The objective reasons bear on what Anna objectively ought to do (take the blue pill). The subjective reasons bear on what she subjectively ought to do (take the red pill). Objective and subjective reasons also seem to play distinct roles from the second-person standpoint: the objective reasons bear on what a fully informed interlocutor should *advise* Anna to do, but do not bear on what that interlocutor should *criticize* her for doing, and vice versa for subjective reasons, which bear on whether a fully informed interlocutor should criticize Anna, but not on what that interlocutor should advise her to do.

If there are objective and subjective reasons, how are they related? Are they species of a genus? If so, what is the differentia? Does one reduce to the other? If so, how does the reduction go? Perhaps the most popular view on this issue is objectivism, which holds that subjective reasons reduce to objective reasons.

My main aim in this paper is to offer two objections to objectivist accounts of subjective reasons. They concern probability and possibility. A secondary aim is to offer a diagnosis of why objectivists face these two objections.

A quick methodological point will be helpful before we begin. I appeal to intuitions about existential claims about subjective and objective reasons. These can be challenged. One way to resolve such challenges—which I apply extensively below—is to consider the distinct roles that objective and subjective reasons play in normative thought. If one contends that  $p$  is a subjective (objective) reason to  $\phi$ , it should play the same kind of roles as other plausible candidates for subjective (objective) reasons to  $\phi$ . For instance, a putative objective reason for Anna to take the blue pill should play the same kind of roles as the fact that

the blue pill will cure her: it should bear on whether she objectively ought to take the pill, and whether a fully informed interlocutor should advise her to do so.<sup>1</sup> I consider this to be a simple litmus test. It might admit of exceptions. But none of the cases I consider below is an exception to the rule. If one wishes to contend otherwise, the onus is on them to defend this.

### 1. OBJECTIVISM

Let us start by homing in on objectivism. Various objectivist accounts of subjective reasons have been developed and defended by Mark Schroeder, Jonathan Way, Derek Parfit, Eric Vogelstein, Daniel Whiting, Kurt Sylvan, and others. The central objectivist commitment is that we should analyze subjective reasons in terms of objective reasons. Most objectivists commit to a fairly narrow version of this view wherein we analyze all subjective reasons to  $\phi$  directly—that is, in terms of objective reasons to  $\phi$ .<sup>2</sup> This will be part of my diagnosis of why extant forms of objectivism face the problems below. But let us not jump ahead.

It is easier to understand objectivism by focusing on a particular account. Many prominent objectivists have been attracted to counterfactual analyses:

For  $R$  to be a subjective reason for  $X$  to do  $A$  is for  $X$  to believe  $R$ , and for it to be the case that  $R$  is the kind of thing, if true, to be an objective reason for  $X$  to do  $A$ .<sup>3</sup>

- 1 That objective and subjective reasons bear distinct relations to objective and subjective “oughts” is widely accepted: “some reasons will be facts of the sort that explain what one *objectively* ought to do, while other reasons will be facts of the sort that explain what one *subjectively* ought to do” (Wedgwood, “The Pitfalls of ‘Reasons,’” 128). The former are objective reasons, the latter subjective reasons. Mark Schroeder appeals to something like this second role as a distinguishing feature of objective reasons throughout “Having Reasons.” He formulates it as a counterfactual test: the question is whether “a fully informed and beneficent bystander would take [the putative objective reason] into account in advising” the agent to perform the relevant act (59). I prefer my formulation, but nothing hangs on this.
- 2 This constraint is evident in the quotation from Schroeder below; I believe it is also accepted by Way and Parfit, along with other objectivists. For instance, here is Vogelstein’s view: “A subjective reason for one to  $\phi$  is a proposition  $p$  such that the members of some consistent subset of one’s  $a$  jointly entail (1) that  $p$  is true, and (2) that the fact that  $p$  is an objective reason for one to  $\phi$ ” (Vogelstein, “Subjective Reasons,” 250).
- 3 Schroeder, *Slaves of the Passions*, 14. See also Schroeder, “Means-End Coherence, Stringency, and Subjective Reasons,” 233, for his similar “subjective reason test”: “ $X$  has a subjective reason to do  $A$  just in case she has some beliefs which have the property, if they are true, of making it the case that  $X$  has an objective reason to do  $A$ ”; see also his converse test at 245.

*Subjective reasons* [are] believed propositions that would be reasons if true.<sup>4</sup>

If we have certain beliefs about the relevant, reason-giving facts, and what we believe would, if it were true, give us some reason, ... such beliefs give us an *apparent* reason.<sup>5</sup>

In what follows I focus primarily on the following version of this view:

C:  $p$  is a subjective reason for  $A$  to  $\phi$  iff and because (a)  $A$  believes  $p$ , and (b) if  $p$  were true,  $p$  would be an objective reason for  $A$  to  $\phi$ .

To illustrate: that the red pill would cure her is a subjective reason for Anna to take the red pill iff and because she believes this proposition, and if it were true that fact would be an objective reason to take the red pill.

Of course, C is not the only game in town. So we will also consider ways that C can be revised, as well as Whiting's and Sylvan's alternatives to C.

## 2. PROBABILITY

Now let's consider how C handles cases involving (subjective) probability. To begin, consider Baqir. He knows that the blue pill would cure his disease. There are two other pills: red and green. Baqir knows that one of them will cure his disease and improve his eyesight, and the other will kill him. In fact, the red pill would kill Baqir and the green pill would cure and improve him. But Baqir does not know this. He justifiably assigns a .5 credence to each possibility.

Counterfactual analyses like C seem to deliver the right results about cases like Baqir's. Baqir believes that the blue pill will cure him. If that proposition were true (which it is) it would be an objective reason to take the blue pill.

There the discussion of probability typically ends. This is unfortunate, as it only captures one class of cases: when there is subjective reason for agents to be cautious instead of risky. What about when there is subjective reason for agents to be risky rather than cautious? Can C get these cases right?

Consider Carla. She knows that if she does not take the purple pill, she will die painlessly. But she does not know what the purple pill will do. She justifiably assigns a .5 credence to two possibilities: it will cure her completely, or it will kill her painfully. In fact, it will kill her painfully. Here are two data points:

*Datum*<sub>1</sub>: There is a subjective reason for Carla to take the pill.

*Datum*<sub>2</sub>: There is no objective reason for Carla to take the pill.

4 Way, "Two Accounts of the Normativity of Rationality," 3.

5 Parfit, *On What Matters*, 1:111.

These existential claims are quite intuitive, though they can be challenged.

For now, let us take them as given and see whether they can be accommodated by C. What is the subjective reason for Carla to take the pill? The most plausible answer, I believe, is a claim about subjective chance that we can call:

*Probability*: There is a .5 chance that the purple pill will cure Carla.

Is this response available to those who accept C? One immediate issue is whether *Probability* is truth-apt. Many take our discourse about subjective chance to be non-factive; they take subjective probabilities to be nothing more than expressions of credal states. This is a problem for C, which holds that *Probability* is a subjective reason for Carla to take the pill only if Carla believes it, and *if it were true*, it would be an objective reason for her to take the pill.<sup>6</sup>

Let us put this issue aside. Grant that *Probability* can be true. If that is the case, presumably it is true. This generates a far more interesting problem for C. If C is true, *Probability* is a subjective reason to take the pill only if it is a counterfactual objective reason to take the pill. But if *Probability* is true, it is a counterfactual objective reason to take the pill only if it is an objective reason to take the pill in the actual world. And *Probability* is not an objective reason to take the pill in our world. To hold otherwise is to deny *Datum*<sub>2</sub>.

In case the middle step in this reasoning is unclear, the basic idea is that counterfactuals tell us what would be the case if the antecedent were true. So the consequent is true if the antecedent is true. And the antecedent of the relevant counterfactual, *Probability*, is true (supposing, again, that it is truth-apt). So if the counterfactual is true, the consequent must be true, too. But the consequent is the claim that *Probability* is an objective reason to take the pill.

A brief aside. It is worth noting something about the broader dialectical import of the specific point above. Arguably, one objectivist alternative to C is the Factoring Account, which was raised and rejected by Schroeder, and has since been defended at length by Errol Lord.<sup>7</sup> The crucial commitment of the Fac-

6 See the discussion of vacuity in section 3, below, for more on why this would pose an issue.

7 See especially Schroeder's "Having Reasons," 58, in which he argues that "the Factoring Account is wrong. In the relevant sense, reasons you have are not things which are, independently of you, reasons, and which moreover, you *have*. There are simply two *reason* relations at stake." Since the Factoring Account does not take there to be a second *reason* relation, it is unclear whether it counts as a form of objectivism. I say that Lord has defended the factoring account, but there is a way of reading his view where it does not aim to analyze the notion of subjective reasons, and so is not objectivist in the sense I am concerned with. This is suggested in, e.g., Lord, "Having Reasons and the Factoring Account," 290, and "The Coherent and the Rational," 156. In personal communication, however, Lord clarified that he intends to analyze subjective reasons in terms of objective reasons that we have or pos-

toring Account is that subjective reasons for  $A$  to  $\phi$  are analyzable in terms of objective reasons for  $A$  to  $\phi$  and *possession*:  $p$  is a subjective reason for  $A$  to  $\phi$  iff  $p$  is an objective reason for  $A$  to  $\phi$  and  $A$  has  $p$ . What led Schroeder to reject the Factoring Account, and defend C instead, is that the former insists that if  $p$  is a subjective reason for  $A$  to  $\phi$ , then  $p$  is an objective reason for  $A$  to  $\phi$ . But C inherits a similar commitment: if  $p$  is a subjective reason for  $A$  to  $\phi$  and  $p$  is true, then  $p$  is an objective reason for  $A$  to  $\phi$ . Objectivists had good grounds for wanting an account that allowed objective and subjective reasons to come apart, but C does not actually deliver that with true propositions.<sup>8</sup>

Now back to our main thread. How should a proponent of C respond to the objection above? I assume she will want to explain *Datum*<sub>1</sub>. But she might challenge *Datum*<sub>2</sub>. Schroeder argued that “negative existential intuitions about reasons are not to be trusted.”<sup>9</sup> Some might take this to mean that they should never be trusted. But that would be a mistake. Schroeder’s explanation for why they should not be trusted turns on pragmatic implicatures about reasons with low weights. He offered a simple test for whether this applies to a given claim:

If I tell you that there is a reason for you to do something that there are only poor reasons for you to do, what I say will sound wrong. But—first prediction—it will sound less wrong if I tell you what the reason is, because doing so will remove the pragmatic reinforcement of the standing presumption that I have only relatively good reasons in mind. And second, if I then tell you that I don’t think it is a particularly weighty reason, I should be able to cancel the presumption, and so the unintuitiveness of what I say should go down a second time.

Let us apply this to the case at hand. That there is an objective reason for Carla to take the pill sounds wrong. (This is why *Datum*<sub>2</sub> seems true.) Does it sound less wrong if we identify the putative objective reason? The claim now is: that the purple pill has a .5 (subjective) probability of curing her is an objective reason to take the pill. This still seems wrong to me. The explanation for why, to be clear, is not because subjective chance is not part of the extra-mental world. Facts about

---

sess, and resists the terminology “subjective” only because it is associated with the notion of “reasons [that] are just a function of one’s perspective, *even when it is seriously deluded*” (Lord, “What You’re Rationally Required to Do and What You Ought to Do (Are the Same Thing!),” 14n25).

- 8 Schroeder argued that there is “the objective reason relation and the subjective reason relation, and the data give us no good reason to suspect that either is a restriction on the other, as the Factoring Account proposes” (“Having Reasons,” 58). I will not discuss the Factoring Account further, but the objections below target it too.
- 9 Schroeder, *Slaves of the Passions*, 92.

subjective mental states can be objective reasons for actions and attitudes. That a man has pathological desires is an objective reason for him to seek therapy, and for us to not vote him into the White House. So my claim is not that *Probability* cannot be an objective reason *per se*.

Rather, the claim I am making is that *Probability* is not an objective reason for Anna to take the pill. If that were the case, *Probability* should play similar roles to other plausible candidates for objective reasons for and against taking the pill. Objective and subjective normative notions play distinct roles in normative thought. Take Anna's case once more. That the blue pill will cure her is an objective reason for her to take the pill. It bears on whether she objectively ought to take it, and whether, knowing the facts, we should advise her to take it. Subjective reasons play different roles: since Anna believes that the red pill will cure her, we should not criticize Anna for taking the red pill. So which roles does *Probability* play? It does not have any bearing on whether she objectively ought to take the pill. Nor does it have any bearing on whether, knowing the facts, we should advise her to take the pill. These roles are played by the facts about what the pills do. But *Probability* does play some roles: it bears on whether we should criticize Carla if she takes the pill, for instance. So *Probability* is a plausible candidate for a subjective reason to take the pill, but not a plausible candidate for an objective reason for her to take the pill.

At this point, some might apply the second part of Schroeder's test. Maybe the problem is just that saying that *Probability* is an objective reason for Carla to take the pill implicates that it is a weighty objective reason for her to take the pill. So let us cancel that implicature: *Probability* is a very weak objective reason for Carla to take the pill. Surely this sounds better! So is *Datum*<sub>2</sub> false?

I do not think so. If the above reasoning is right, *Probability* is not merely an outweighed objective reason. The point of distinguishing between objective and subjective reasons is that they play distinct roles. If the fact that the pill will kill her competes with *Probability* in determining what Carla objectively ought to do, and how all-knowing advisors ought to advise her to act, the distinction between objective and subjective reasons is a distinction without a difference.

This raises the question of why the claim above sounds better once the implicature is canceled. But that is fairly easy to explain. Mistakes can be more or less egregious. Saying that the fact that someone is your mother is a reason to torture her sounds very, very wrong. It sounds less wrong if you cancel the implicature that this is a weighty reason to torture her. But surely that does not mean that the implicature was the sole source of the problem!

So far we have seen that a proponent of C cannot accommodate the data by identifying *Probability* as the subjective reason for Carla to take the pill, and

should not respond to this problem by denying the data. An alternative response is to accommodate the data by finding some other candidate for the subjective reason for Carla to take the pill. Notably, many other initially plausible candidates will raise the exact same problems as *Probability*. For instance, consider the following plausible conjecture: there is some *evidence* that justifies Carla's beliefs and credences; whatever that evidence is, it is the subjective reason for her to act.<sup>10</sup> To fill this proposal out a little, let us imagine that the relevant evidence is testimony—Carla is justified in assigning her credences on the basis of what her reliable friend said. Now we can consider the following proposition:

*Evidence*: The reliable friend said that there is a .5 chance that the purple pill will cure Carla.

Alternatively, consider the plausible conjecture that *Carla's doxastic states* are the subjective reasons for her to act.<sup>11</sup> Now we can consider the proposition:

*Belief*: Carla believes that there is a .5 chance that the purple pill will cure her.

Can a proponent of C say that *Evidence* or *Belief* is the subjective reason for Carla to take the pill? No: both initially plausible conjectures raise the exact same problems as *Probability*. Let us start with *Evidence*. If C is true, *Evidence* is a subjective reason to take the pill only if it is a counterfactual objective reason to take the pill. And since *Evidence* is true, that means it is a counterfactual objective reason to take the pill only if it is an objective reason to take the pill in the actual world. But according to *Datum*<sub>2</sub>, there is no objective reason to take the pill (in the actual world). As with *Probability*, one can try to challenge this datum by appealing to Schroeder's test, but we do not get better results when we apply this test to *Evidence*. It does not play the right roles to be an objective reason for her to take the pill. It has no bearing on whether she objectively ought to take the pill, or on whether, knowing the facts, we should advise her to take the pill.<sup>12</sup> The same reasoning applies to *Belief* with even greater force: it is also true, and clearly fails to play the right roles to be an objective reason for her to take the pill. If merely believing that pills will cure you is an objective reason for you to take them, the distinction between objective and subjective reasons seems to be a distinction without a difference.

This shows that, to accommodate the data, the objectivist needs to find a very

10 I am grateful to an anonymous referee for suggesting this response.

11 I am grateful to an anonymous referee for suggesting this response. It has affinities with Dale Dorsey's view in "Objective Morality, Subjective Morality, and the Explanatory Question."

12 If one doubts this, see the discussion of a similar proposal, *Testimony*, in section 3, below.

different candidate for the subjective reason for Carla to take the pill. It should be a proposition such that Carla believes it and if it were true it would be an objective reason to take the pill; but it should not be a proposition that is actually true, lest we go through another round of the reasoning above. The best remaining candidates, I believe, will be propositions like the following:

*Prejacent*: The purple pill will cure Carla.<sup>13</sup>

Why will the best remaining candidates be propositions like *Prejacent*? Because it is clearly truth-apt, it is false (so it does not face the same problems as *Probability*, *Evidence*, and *Belief*), and if it were true it would be an objective reason to take the pill. The obvious bug is that Carla does not believe *Prejacent*, and C insists that *p* is a subjective reason for Carla only if Carla believes that *p*. This necessary condition is explicitly embraced by Schroeder, Way, and Parfit in the passages quoted above.<sup>14</sup> Plausibly, any alternatives to *Prejacent* that share its virtues will also share this vice; there does not seem to be any false proposition that Carla actually *believes* such that if it were true it would be an objective reason for her to take the pill.

As has been previously noted, other views about reasons have a similar bug in cases like Carla's, insofar as those views also require reasons to be believed (or known, or what have you).<sup>15</sup> I am not sure whether the bug poses a serious problem for these other views, which are not framed as objectivist accounts of subjective reasons. It may be open to those views to identify propositions like

- 13 In case this is unclear, I am calling this *Prejacent* to make a more general point. We have seen the difficulties objectivists face if they treat propositions like  $\diamond p$  as a subjective reason to  $\phi$ , where  $\diamond$  is an epistemic modal that scopes over what is called the prejacent: *p*. The most plausible move remaining is to treat the prejacent itself, *p*, as the subjective reason to  $\phi$ .
- 14 To be clear, the issue here is not that these passages were clumsily formulated. Lord notes that on Schroeder's view "a necessary condition for something to be a subjective reason is that one must believe the proposition that constitutes the subjective reason" (Lord, "Having Reasons and the Factoring Account," 291). See also Schroeder, "Means-End Coherence, Stringency, and Subjective Reasons," 245, for a clear, specific case in which this necessary condition does serious work for Schroeder ("since Wynn does not have a belief about that, it does not figure among her subjective reasons"). In other work, Schroeder adopts a weaker necessary condition: a *presentational attitude* that *p* is all that is required ("What Does It Take to 'Have' a Reason?"). I believe this revised position is still vulnerable to the objection pressed here, for the same reasons discussed by Wedgwood in "The Pitfalls of 'Reasons,'" 135.
- 15 See, in particular, Wedgwood, who objects that common views about the relation between *normative* and *motivating* reasons systematically overestimate "the centrality of outright belief" ("The Pitfalls of 'Reasons,'" 134); and see Schiffer, who objects that the widely held view that "one should act only on what one knows" is problematic in cases where an agent is "justified in acting on a partial belief" ("Interest-Relative Invariantism," 189–90).

*Probability* as agents' believed (or known, or ...) reason for action.<sup>16</sup> But this response is not available to views about subjective reasons such as C.

It seems, then, that we must weaken the doxastic component of C. What if we allow that *p* is a subjective reason for Carla only if Carla has a credence above *n* in *p*, where *n* is specified at some value? What value would that be? It is easy to show that the relevant value for *n* will have to be very low. There can be a subjective reason to act even when one only has a negligible credence in the relevant proposition. Just make Carla almost certain that the purple pill is a placebo: say she justifiably assigns a credence of .99 (or .999, or ...) to this proposition, and assigns a .005 (or .0005, or ...) credence each to the propositions that it will cure her and that it will kill her painfully. Now the revised version of C will not explain why there is a subjective reason for her to take the purple pill unless we allow that any positive credence will suffice.

The objectivist who most clearly embraces such a view is Vogelstein.<sup>17</sup> It allows objectivists to take *Prejacent* to be a subjective reason for Carla to take the pill, and does not imply that *Prejacent* is an objective reason for her to take the pill (because *Prejacent* is false), so it accommodates our data points.

But this move comes with three serious costs. Vogelstein recognizes the first: it makes subjective reasons maximally proliferate. The consequences of acts are contingent. We should have credences between 0 and 1 in contingent truths. So for almost any idiotic action, agents should have some positive credence that it will cure Carla, or bring about world peace, or what have you. So as Vogelstein says, "we have subjective reasons to do almost anything."<sup>18</sup> This does not worry Vogelstein, who insists that such subjective reasons have very low weights, and wields Schroeder's claims about intuitions about negative existential claims. I do not think that Vogelstein's position here seems plausible if we apply the methodological point with which we started. But I do not want to pursue this point as I believe that this form of objectivism faces more damning problems.

The second serious cost is that this form of objectivism misidentifies the rel-

16 For instance, that is how Hawthorne and Stanley respond to Schiffer in "Knowledge and Action," 136, arguing that *contra* Schiffer (and, in effect, Wedgwood in "The Pitfalls of 'Reasons,'" 136), agents can believe or know *Probability* without having an implausible degree of conceptual sophistication. Neither Schiffer nor Wedgwood considers the objection that *Probability* is not an *objective* reason for action.

17 Dorsey endorses a similar view of subjective reasons in "Objective Morality, Subjective Morality, and the Explanatory Question." Dorsey argues that "disbelieved propositions might themselves constitute subjective reasons" (11), and recognizes that his account proliferates subjective reasons (12). Many of my objections to Vogelstein's view apply equally to Dorsey's more complicated position.

18 Vogelstein, "Subjective Reasons," 247–48. See note 2, above.

evant subjective reasons. Even if the analysis gets the result that there is a subjective reason in these cases (*Datum*<sub>1</sub>), it gets the wrong result about which proposition is a subjective reason for Carla to take the pill. The subjective reason is that there is a .5 probability that the pill will cure her, not that the pill will cure her. It is *Probability*, not *Prejacent*. After all, when we compare Anna and Carla, it is intuitive that they have quite different subjective reasons to act because Anna believes that the pill *will* cure her while Carla only believes that the pill *might* cure her. This shows up clearly when we consider what proposition would play the relevant role in normative thought: in explaining why they should not be criticized if they take the pill, we would appeal to something like *Prejacent* in Anna's case and something like *Probability* in Carla's case.<sup>19</sup>

The final serious cost is that C now commits us to an atomic rather than a holistic view. To bring this into focus, consider one final variant on Carla's case. Say she was justifiably almost certain that the pill would kill her painfully, but still had some positive credence that it would cure her. Is there still a subjective reason for her to take the pill? On atomic views, there is: all that matters is her credence in the atomic proposition. On holist views, there is not: what matters is the overall distribution of probabilities over possible outcomes, not the assignment of a probability to a single possible outcome. In other words, holism holds that propositions like "the purple pill might cure Carla" can be a reason to take the pill in one context (in which the subjective chance that the pill will kill her painfully is low) but not be a subjective reason to perform the same action in another context (in which the subjective chance that the pill will kill her painfully is high); the atomist denies that this is possible.<sup>20</sup> I am not sure whether atomism or holism is true. But it strikes me that we should have an account of *what it is* to be a subjective reason that allows for both views to be coherently stated. Counterfactual analyses like Vogelstein's do not have this feature. Chaos ensues once we plug distributions of probabilities over (logically inconsistent) possible outcomes into the antecedents of counterfactuals.<sup>21</sup>

So far I have argued that C faces serious problems with probability. This point is at least somewhat important on its own. But it becomes all the more important once we note that objectivists who eschew counterfactual analyses face the same problem—or, at least, a problem with the same structure.

19 By "something like *Probability*," I mean to include *Evidence* or *Belief*, since these propositions can plausibly also play the same roles as subjective reasons for Carla to take the pill.

20 I use the term *holism* because this view dovetails nicely with Jonathan Dancy's version of value holism, according to which, "For any *x* that has value in one context, *x* may have a different value or none at all in other contexts" ("The Particularist's Progress," 13).

21 For this reason, Vogelstein's view appeals to a *consistent subset of one's credences*.

Consider Sylvan's view, according to which  $p$  is a subjective reason for  $A$  to  $\phi$  iff and because  $p$  is an "apparent fact" that  $A$  is competently attracted to treating like an objective reason to  $\phi$ .<sup>22</sup> I am attracted to this view. But how does it apply to Carla's case? What is the subjective reason for her to act?

Sylvan might say: the answer is *Probability*. We supposed that *Probability* is truth-apt. In a similar vein, let us suppose that it is an "apparent fact." For Sylvan, *Probability* is a subjective reason to take the pill only if Carla is competently attracted to treating it like an objective reason to take the pill. We saw above that, for  $C$ , *Probability* is a subjective reason to take the pill only if it is an objective reason to take the pill, which is problematic because it does not play the right kind of roles to plausibly be an objective reason to take the pill. On that same basis, we can say that Carla is not competently attracted to treat *Probability* like an objective reason to take the pill. (The same holds for propositions like *Probability*, such as *Evidence* and *Belief*.) So for Sylvan, as for  $C$ , *Probability* cannot be Carla's subjective reason to take the pill.

What about *Prejacent*? This would be an objective reason if it were true. But Sylvan's view faces the same problem as  $C$  here: Carla can have a subjective reason to take the pill even when her credence in *Prejacent* is .5, or 0.00005, or what have you. (Keep in mind the variations on the case above.) It is hard to see how *Prejacent* can be an "apparent fact" to Carla when she believes that it is almost certainly false. So for Sylvan, *Prejacent* is not a subjective reason.

In sum, counterfactual analyses and other forms of objectivism struggle to explain the data about subjective reasons in cases involving probability. Why has the problem been neglected? The whole point of introducing a distinction between objective and subjective reasons is that the facts can come apart from our perspective on the facts—subjective reasons, in Whiting's wonderful phrase, are introduced to "keep things in perspective." However, objectivists have only focused on how false beliefs cause the two to come apart. They have ignored how our perspectives systematically come apart from the facts due to our uncertainty.

### 3. POSSIBILITY

At one level of description, the first objection has a simple structure. There is a subjective reason to  $\phi$ . That reason is either a complex proposition (*Probability*), or a proposition embedded within it (*Prejacent*). But both of these options raise serious problems for current forms of objectivism. The most interesting part of this objection is the explanation for why, on current forms of objectivism, that complex proposition (*Probability*) cannot be a subjective reason to  $\phi$ .

22 Sylvan, "What Apparent Reasons Appear to Be."

At that level of description, the structure of the second objection is the same. So is the problem raised by the complex proposition. The most interesting, and distinct, part of this objection will be the explanation for why, on current forms of objectivism, the relevant embedded proposition cannot be a subjective reason.

That is a fairly abstract preamble. So let us jump into some cases. Consider Duquan. He is taking an exam that he must pass, but that requires him to answer every question correctly. And he wants to pass at all costs. He is stuck on one question: “Name one true philosophical theory.” Duquan skipped his philosophy seminars, but his reliable friend told him that modal realism is true. So he answers, “Modal realism: all possible worlds are real.” Unfortunately, modal realism is false. Indeed, it is *necessarily* false. So he fails the exam.

A slight variation on this case will ultimately be more revealing. Consider Emiliano. He is in the same position as Duquan, except that (a) his reliable friend said that error theory is true, and (b) Emiliano is one of those wonderful students who cares most about getting things right—he wants to assert truths, and to pass by offering correct answers. So Emiliano answers, “Error theory: there are no objective reasons.” Error theory is necessarily false. So he fails, too.

Let us introduce two data points that are quite intuitive, albeit challengeable:

*Datum*<sub>3</sub>: There was a subjective reason for Duquan/Emiliano to answer “modal realism”/“error theory” in the exam.

*Datum*<sub>4</sub>: There was no objective reason for Duquan/Emiliano to answer “modal realism”/“error theory” in the exam.

For now, let us take these data points as given and ask whether they can be accommodated by C. What was the relevant subjective reason? As in section 2, the best candidates are a complex proposition (in this case, the testimony) and a proposition embedded within it (in this case, the content of that testimony):

*Testimony*: The reliable friend said that modal realism/error theory is true.

*Content*: Modal realism is true/error theory is true.

I am agnostic about which proposition is the better candidate for the relevant subjective reason. But that does not matter, as both generate problems.

Let us start with *Testimony*, as the problem here will be familiar. In fact, it will have exactly the same structure as the problem with *Possibility*, above. After all, *Testimony* is true. If *Testimony* is both true and a subjective reason to  $\phi$ , then according to C it must be an objective reason to  $\phi$  in the actual world. This is counterintuitive. It makes *Datum*<sub>3</sub> true only if *Datum*<sub>4</sub> is false.

These intuitions about negative existential claims can be challenged. But Schroeder's test applies the same way to this case as it did to Carla's case above. Intuitively, *Testimony* does not seem to be an objective reason for Duquan and Emiliano to answer "modal realism"/"error theory" and thereby fail the exam, even if we cancel the implicature that it is a weighty objective reason. And as before, this claim can be bolstered by noting that *Testimony* does not seem to play the roles that objective reasons for such actions play: it does not bear on whether Duquan and Emiliano did what they objectively ought to do, or bear on what informed bystanders ought to advise them to do. If Duquan whispered to an all-knowing invigilator, "Should I answer 'modal realism'?" the fact that modal realism is false would clearly be relevant to whether he should answer no, but the fact that *Testimony* is true would clearly be irrelevant. *Testimony* is not merely outweighed by objective reasons (such as the fact that modal realism is false, which is a decisive objective reason not to answer "modal realism"); rather, *Testimony* fails to even compete with objective reasons at all.

Still, some may doubt that *Testimony* fails to compete with the objective reasons to answer "modal realism"/"error theory." And, indeed, they might doubt the similar verdicts about *Probability*, *Evidence*, and *Belief* in section 2. Without relying on contentious claims about the nature of objective and subjective reasons, what more can be said to bolster the claim that these facts do not play the right roles to be objective reasons for Carla, Duquan, and Emiliano to act?

One way to bolster these verdicts is to carefully distinguish the negative existential claims in question from nearby claims that are either irrelevant or implausible. The relevant negative existential claim is that *Testimony* is not an objective reason for Emiliano to answer "error theory" (and *mutatis mutandis* for Duquan). This should be distinguished from other plausible but irrelevant negative existential claims in the vicinity, such as the claim that it is not the case that Emiliano objectively ought to answer "error theory" (and *mutatis mutandis* for Duquan). These verdicts at the all-things-considered or summative level are not what is at stake here. The relevant negative existential claim is that *Testimony* has no bearing on whether Emiliano objectively ought to answer "error theory" (and *mutatis mutandis* for Duquan). These verdicts about the contributory level are relevant, and are also plausible. Moreover, I am not making the implausible claim that facts about testimony can never be objective reasons for anything; I am making the claim that *Testimony* is not an objective reason to answer "modal realism"/"error theory." Nor am I making the claim that in no sense is *Testimony* a reason to answer "modal realism"/"error theory." After all, *Testimony* does bear on whether Duquan and Emiliano did what they subjectively ought to do, and on whether fully informed interlocutors should (a) recognize this, and/or

(b) criticize them for failing the exam. So *Testimony* is plausibly a subjective reason for these students to perform these actions. It just is not plausibly a (counterfactual) objective reason for them to do so.<sup>23</sup>

Some may still doubt these verdicts about *Testimony* et al., so I discuss this more in section 4. For now, let us put such doubts aside and consider the second (and perhaps more interesting and distinct) part of this objection to objectivism. This is the explanation for why, on current forms of objectivism, the relevant embedded proposition—*Content*—cannot be a subjective reason for Duquan and Emiliano to answer as they did. As I said before, I am not sure whether this proposition is the better candidate for the relevant subjective reason. I just think that there should be a good candidate (in order to explain *Datum*<sub>3</sub>), and if we have ruled out true, complex propositions like *Testimony*, the natural remaining candidates will be false, simple propositions like *Content*.

Recall that, for C, *Content* is a subjective reason for Duquan and Emiliano to answer as they did only if the following holds: if *Content* were true, it would be an objective reason for Duquan to write “modal realism”/Emiliano to write “error theory.” However, *Content* cannot be true. So the relevant counterfactual will have a necessarily false antecedent. This poses two problems for C. The first is not devastating, but it is a good way to warm up to the second.

The first problem is fairly obvious and immediate.<sup>24</sup> Say we accept any theory that holds that a counterfactual is vacuously true if its antecedent is necessarily false. For instance, say we take a counterfactual to be true iff its consequent is true in the nearest possible worlds in which its antecedent is true. If there are no possible worlds in which the antecedent is true, the counterfactual is vacuously true. So for C, *Content* is a subjective reason to do anything.<sup>25</sup>

I do not want to focus on this first problem as I do not think it is devastating. Many balk at the idea that counterfactuals with necessarily false antecedents are vacuously true, and propose that we should fix this bug by considering what is true at the nearest possible or impossible worlds in which the antecedent is true.<sup>26</sup> This fix avoids getting the wrong results in Duquan’s case.

But notice the implications that this fix has when we turn to Emiliano’s case.

23 I am grateful to an anonymous referee for pushing me to further clarify these matters.

24 The following point was Whiting’s main objection to C in “Keep Things in Perspective.”

25 To be clear, the problem here would not be that C does not get the right result that *Content* is a subjective reason for Duquan to answer as he did. The problem would be that we get every single wrong result. *Content* is a subjective reason for Duquan to do literally anything, because any counterfactual with *Content* as the antecedent is vacuously true.

26 For general discussion, see Nolan, “Impossible Worlds.”

This is where we get to the second problem, which I do think is devastating.<sup>27</sup> In the impossible world in which error theory is true, there are no objective reasons whatsoever. In that impossible world, there is no objective reason for Emiliano to answer “error theory,” or indeed do anything at all, ever. So we have a subjective reason that could not be an objective reason in any possible or impossible world. What could be worse for C than the result that some subjective reasons for action are not (im)possible objective reasons for action?

What can objectivists say in response? Some might claim that the objection relies on a false dilemma. Perhaps we should identify subjective reasons by appealing to the agent’s desires or goals.<sup>28</sup> If this is so, then in some cases we will not need to appeal to *Testimony* or *Content*. Duquan, for instance, is a bit of a grade-grubber, so the subjective reason in his case might be as follows:

*Pass*: By answering “modal realism” I will pass the exam.

I doubt that *Pass* poses problems for objectivists. But note that the corresponding move is not available in Emiliano’s case because he is not a grade-grubber. Appealing to the content of his desires or goals leads us right back to *Content*. So far from being supported by the view that we should identify subjective reasons by appealing to the agent’s desires or goals, C is at odds with that picture when we consider students like Emiliano—or, indeed, professional philosophers who assert that error theory is true and only want to assert the truth. Indeed, C seems to be at odds with any view on which *p* can be our subjective reason to assert that *p*, which is a surprising and unwelcome result.

Alternatively, some might respond that the objection above is devastating for C, but leaves other forms of objectivism unscathed. You might think that Whiting’s view is especially well placed to address this objection, since it is largely motivated by appealing to problems concerning possibility and vacuity.<sup>29</sup>

Interestingly, however, Whiting’s view faces the same objection—or at least, an objection with the same structure. For Whiting, *p* is a subjective reason for A to  $\phi$  iff and because *p* is an epistemically necessary objective reason for A to

27 It might be said that this second problem has the form of the conditional fallacy (Shope, “The Conditional Fallacy in Contemporary Philosophy”), though that is not terribly informative given Shope’s “elaborate but somewhat obscure generic formulation of the conditional fallacy” (Bonevac, Dever, and Sosa, “The Conditional Fallacy,” 275). If one is tempted by the thought that conditional fallacy-style counterexamples will obviously arise for a counterfactual analysis like C, it is worth reading on to note that the same problem undermines a prominent alternative to counterfactual analyses, namely Whiting’s view.

28 I am grateful to John Brunero for suggesting this.

29 See note 24, above. Whiting neglects the objections below in part because he focuses on counterfactuals involving necessary falsehoods where the truth is knowable *a posteriori*.

$\phi$ . More perspicuously:  $p$  is a subjective reason for  $A$  to  $\phi$  iff and because it is *a priori* that, if the facts of the situation are as they appear to  $A$ ,  $p$  is an objective reason for  $A$  to  $\phi$ . Can this view accommodate *Datum*<sub>3</sub> and *Datum*<sub>4</sub>?

Whiting cannot appeal to *Testimony*. We saw above that it does not play the right roles to be an objective reason to answer “modal realism”/“error theory.” On that same basis, we can say that it is not *a priori* that if *Testimony* is true—which it is, and which it appears to be for Duquan and Emiliano—it is an objective reason to answer “modal realism”/“error theory.”

More interestingly, Whiting cannot take *Content* to be a subjective reason for Duquan and Emiliano to act. Both cases pose difficulties. In Duquan’s case, there are technical issues about how to handle necessary truths that are knowable *a priori*. It is knowable *a priori* that modal realism is false. On some views, that means that it is epistemically necessary that *Content* is false, and so epistemically necessary that *Content* is not an objective reason to do anything.

The more interesting problem is posed by Emiliano’s case. Even if it is not epistemically necessary that error theory is false, it is clearly *a priori* that if things are as they appear to Emiliano—if *Content* is true—there is no objective reason for him to answer “error theory,” or indeed do anything. So for intuitive, non-technical reasons, *Content* cannot be Emiliano’s reason to answer “error theory.” And, more generally, the content of what one asserts cannot be even part of a subjective reason for one to assert that error theory is true.

#### 4. OUTWEIGHED REASONS

In section 2 I argued that complex propositions like *Probability*, *Evidence*, and *Belief* are not objective reasons for Carla to take the pill. Similarly, in section 3 I argued that the complex proposition *Testimony* is not an objective reason for Duquan and Emiliano to answer as they did. These verdicts generate problems for current forms of objectivism, as it makes it tough for current objectivist views to explain how these propositions could be subjective reasons for Carla to take the pill or Duquan and Emiliano to answer as they did. I offered one way to bolster these verdicts by carefully distinguishing them from nearby claims that are either irrelevant or implausible. But given their central role in the objections to current forms of objectivism, more needs to be said to assuage any lingering doubts about these verdicts. In this section I attempt to remedy that.

I suspect that lingering doubts about these verdicts are most likely to turn on a difficulty discussed above—namely, the difficulty of eliciting clear and probative intuitions about whether  $p$  is (a) a massively outweighed objective reason for  $A$  to  $\phi$ , or (b) not an objective reason for  $A$  to  $\phi$ . What more can be said to

bolster my verdicts that in the relevant cases  $p$  is not an (outweighed) objective reason?

Here is a way to bolster these verdicts. Do not compare  $p$  to paradigm cases of strong or decisive objective reasons for  $A$  to  $\phi$ . Instead, compare  $p$  to paradigm cases of massively outweighed objective reasons for  $A$  to  $\phi$ . If this elicits clear and probative intuitions that  $p$  does not play the same roles as paradigm cases of massively outweighed objective reasons, that strongly suggests that  $p$  is not a massively outweighed objective reason after all. This will be my strategy.

I want to execute this strategy using relatively simple cases and ecumenical resources. So let us return to Anna's case, in which the fact that the red pill will cure her is a decisive objective reason for Anna to take the pill. Holding this fixed, we can elicit helpful intuitive data by comparing two versions of this case:

*Mild Nausea*: The red pill will also make Anna mildly nauseous.

*Misleading Evidence*: Anna receives evidence that the red pill will make her mildly nauseous, but this evidence is misleading—it will do no such thing.

In the first version of the case, I take it that the fact that the red pill will make Anna mildly nauseous is a paradigm instance of a massively outweighed objective reason for her not to take the red pill. My intuitions about this case seem clear and probative: this fact is relevant to what she objectively ought to do and what a fully informed interlocutor should advise her to do. Now compare this to *Misleading Evidence*. The fact that she receives misleading evidence does not play the same roles as our paradigmatic massively outweighed objective reason: it bears on what she *subjectively* ought to do, but it does not bear on what she *objectively* ought to do, nor does it bear on whether or not a fully informed interlocutor should advise her to take the red pill.

Since this comparison is between two putative massively outweighed objective reasons, it is hard to see how issues regarding weight could lead us astray here. And, if one wished to contend otherwise, they would face an uphill battle. If there is an objective reason not to take the red pill in *Misleading Evidence*, presumably we could increase its weight by increasing the strength of the evidence or the intensity of the nausea: such changes do not seem to make this misleading evidence relevant in the way that actual nausea would be relevant. Strong misleading evidence of excruciating nausea still does not seem to bear on what Anna objectively ought to do, or on what a fully informed interlocutor should advise her to do, whereas mild actual nausea does bear on such matters.

If objectivists accept this judgment about misleading evidence in Anna's case,

surely they should accept similar judgments about misleading evidence—like *Testimony* and *Evidence*—in cases like Carla’s, Duquan’s, and Emiliano’s.<sup>30</sup> Objectivists should want to preserve a distinction between actual countervailing considerations and misleading evidence of countervailing considerations. This suggests that objectivists cannot easily answer the two central objections above by insisting that *Testimony* et al. are objective reasons for Carla et al. to act.

##### 5. CONCLUSION AND DIAGNOSIS

Why do objectivists face these two objections? As was foreshadowed above, my diagnosis concerns an additional commitment that objectivists have taken on without argument: that we must analyze all subjective reasons *directly*. What does that mean? It means that we must analyze subjective reasons for  $A$  to  $\phi$  in terms of corresponding objective reasons for that agent to perform that same action. For instance, according to the objectivist views we have considered,  $p$  is a subjective reason for Carla to take the pill only if  $p$  is counterfactually an objective reason for Carla to take the pill (counterfactual analyses); Carla is competently attracted to treating  $p$  like an objective reason to take the pill (Sylvan’s view); or  $p$  is an epistemically necessary objective reason for Carla to take the pill (Whiting’s view). By taking on this commitment, objectivists have occupied a fairly narrow region of logical space. And the two objections above reveal that it is also a fairly unappealing region of logical space.

Consider the first objection. We saw that *Probability* is the most plausible candidate for a subjective reason for Carla to take the pill. The problem was not that it cannot be an objective reason *per se*. It was that *Probability* does not play the right roles to be a counterfactual (and hence actual) objective reason for Carla to take the pill, and Carla cannot competently treat it as an objective reason to take the pill. This problem is compounded because on these objectivist views no other proposition is a good candidate for being the subjective reason for Carla to take the pill, so objectivists struggle to explain intuitive data points.

Now consider the second objection. We saw that *Testimony* is a plausible candidate for a subjective reason for Emiliano to answer “error theory.” Again, the problem is not that this proposition cannot be an objective reason *per se*. It is that this proposition is not a plausible candidate for a counterfactual or epis-

30 It is worth noting here that Julia Markovits has defended a similar view: “If expert testimony gives us most reason to believe some act would be best then that testimony *is the reason* we ought to perform that act,” even if that testimony is misleading; but as she argues, these “moral reasons for us to perform some action are subjective” and not objective reasons (Markovits, “Saints, Heroes, Sages, and Villains,” 306–7).

temically necessary objective reason for Emiliano to answer “error theory.” This problem is compounded because objectivists cannot take the false content of that testimony to be a subjective reason for Emiliano to answer as he did, so once again objectivists struggle to accommodate intuitive data points.

This does not show that we cannot analyze subjective reasons in terms of objective reasons. It suggests that we cannot analyze all subjective reasons for  $A$  to  $\phi$  directly, in terms of objective reasons for  $A$  to  $\phi$ . So the problem is not that objectivism is false, but that it has been construed narrowly, and naïvely.

Interestingly, this diagnosis dovetails nicely with a lesson Frank Jackson drew a long time ago about the relation between objective and subjective normative notions. Jackson noticed that, in some cases involving uncertainty, we subjectively ought to  $\phi$  even though we know that it is definitely not the case that we objectively ought to  $\phi$ . Jackson took this to show that naïve views that derived subjective oughts directly from objective oughts were problematic; instead, we should adopt a “decision-theoretic” view about the subjective ought.<sup>31</sup>

Some details of Jackson’s own view are not compatible with objectivism. But they do not concern us. The interesting point is that objectivists could offer a decision-theoretic explanation of the subjective reasons for Carla and Emiliano to  $\phi$  in terms of objective reasons for credences and preferences.

That explanation would require many moving parts that warrant independent motivation and extensive discussion. I will not delve into those details here. All that I hoped to do here is motivate the exploration of neglected areas of logical space: objectivists can and should seek to analyze some subjective reasons for action *indirectly*, in terms of objective reasons for attitudes.<sup>32</sup>

Virginia Tech  
dwodak@vt.edu

#### REFERENCES

- Bonevac, Daniel, Josh Dever, and David Sosa. “The Conditional Fallacy.” *Philosophical Review* 115, no. 3 (July 2006): 273–316.
- Dancy, Jonathan. “The Particularist’s Progress.” In *Moral Particularism*, edited by

31 Jackson, “Decision-Theoretic Consequentialism and the Nearest and Dearest Objection.”

32 Many thanks to John Brunero, Pete Graham, Frank Jackson, Nicholas Laskowski, Errol Lord, Philip Pettit, Douglas Portmore, Pamela Robinson, Gideon Rosen, Michael Smith, several anonymous reviewers, and audiences at the University of Southampton and at the 2016 American Philosophical Association Central Division Meeting.

- Brad Hooker and Margaret Little, 130–56. Oxford: Oxford University Press, 2000.
- Dorsey, Dale. “Objective Morality, Subjective Morality, and the Explanatory Question.” *Journal of Ethics and Social Philosophy* 6, no. 3 (2012): 1–24.
- Hawthorne, John, and Jason Stanley. “Knowledge and Action.” *Journal of Philosophy* 105, no. 10 (October 2008): 571–90.
- Jackson, Frank. “Decision-Theoretic Consequentialism and the Nearest and Dearest Objection.” *Ethics* 101, no. 3 (April 1991): 461–82.
- Markovits, Julia. “Saints, Heroes, Sages, and Villains.” *Philosophical Studies* 158, no. 2 (March 2012): 289–311.
- Lord, Errol. “The Coherent and the Rational.” *Analytic Philosophy* 55, no. 2 (June 2014): 151–75.
- . “Having Reasons and the Factoring Account.” *Philosophical Studies* 149, no. 3 (July 2010): 283–96.
- . “What You’re Rationally Required to Do and What You Ought to Do (Are the Same Thing!).” *Mind* (forthcoming).
- Nolan, Daniel. “Impossible Worlds.” *Philosophy Compass* 8, no. 4 (April 2013): 360–72.
- Parfit, Derek. *On What Matters*. 2 vols. Oxford: Oxford University Press, 2011.
- Schiffer, Stephen. “Interest-Relative Invariantism.” *Philosophy and Phenomenological Research* 75, no. 1 (July 2007): 188–95.
- Schroeder, Mark. “Having Reasons.” *Philosophical Studies* 139, no. 1 (April 2008): 57–71.
- . “Means-End Coherence, Stringency, and Subjective Reasons.” *Philosophical Studies* 143, no. 2 (March 2009): 223–48.
- . *Slaves of the Passions*. Oxford: Oxford University Press, 2007.
- . “What Does It Take to ‘Have’ a Reason?” In *Reasons for Belief*, edited by Andrew Reisner and Asbjorn Steglich-Petersen, 201–22. Cambridge: Cambridge University Press, 2011.
- Shope, Robert. “The Conditional Fallacy in Contemporary Philosophy.” *Journal of Philosophy* 75, no. 8 (August 1978): 397–413.
- Sylvan, Kurt. “What Apparent Reasons Appear to Be.” *Philosophical Studies* 172, no. 3 (March 2015): 587–606.
- Vogelstein, Eric. “Subjective Reasons.” *Ethical Theory and Moral Practice* 15, no. 2 (April 2012): 239–57.
- Way, Jonathan. “Two Accounts of the Normativity of Rationality.” *Journal of Ethics and Social Philosophy* (December 2009).
- Wedgwood, Ralph. “The Pitfalls of ‘Reasons.’” *Philosophical Issues* 25, no. 1 (October 2015): 123–43.

Whiting, Daniel. "Keep Things in Perspective: Reasons, Rationality and the A Priori." *Journal of Ethics and Social Philosophy* 8, no. 1 (2014): 1–22.

## IMMIGRATION POLICY AND IDENTIFICATION ACROSS BORDERS

*Matthew Lindauer*

IMMIGRATION POLICIES can express disrespect for members of society, nonmembers, or both. Proponents of the traditional state sovereignty view on immigration have generally held that only policies in the first and third categories could be moral wrongs—it is morally regrettable, perhaps, but not morally impermissible for a state to implement immigration policies that express disrespect for outsiders.

One problem with the state sovereignty view, I argue, is that it is insensitive to the ways in which members and nonmembers relate to one another. The “external relationships” of members, relationships they stand in with nonmembers, make it the case that the treatment of nonmembers can affect or express attitudes about members themselves. Some of these relationships are what we might call “relationships of closeness,” such as family and romantic relationships. In this paper, I focus instead on “relationships of likeness,” or more specifically, relationships between members and nonmembers that hold due to a shared quality or set of qualities on the basis of which members identify with nonmembers. These relationships of likeness make it the case that immigration policies that discriminate against nonmembers also often discriminate against members, and while this point has been recognized to some extent, its full implications have not been appreciated.

Some theorists who defend the state sovereignty view, in fact, have tried to curtail the permissive implications of the view for policies that are racially, ethnically, or otherwise discriminatory by appealing to discrimination against members. One argument of this kind that Christopher Heath Wellman has used to defend the state sovereignty view was originally given by Michael Blake.<sup>1</sup> Blake argues that, because immigration policies that discriminate against nonmem-

<sup>1</sup> Blake, “Immigration,” 233–34. See Wellman, “Immigration and Freedom of Association,” 139–41. For different critical responses to Wellman, see Fine, “Freedom of Association Is Not the Answer”; Blake, “Immigration, Association, and Antidiscrimination”; Cavallero, “Association and Asylum”; and Wilcox, “Do Duties to Outsiders Entail Open Borders?”

bers typically also make invidious comparisons between members, discriminatory immigration policies can be morally impermissible even by the lights of the state sovereignty view. Blake did not intend for this argument to undermine the state sovereignty view, nor did Wellman think it did. However, I argue that the fact that states generally cannot implement discriminatory immigration policies without expressing disrespect for their own members, *contra* Blake and Wellman, is a serious problem for the state sovereignty view.

As Wellman acknowledges in later work, Blake's argument may be unsatisfying for at least two reasons.<sup>2</sup> First, it seems odd to think that the White Australia policy, the Chinese Exclusion Act, and Donald Trump's recent executive orders on immigration have only been wrong in virtue of discriminating against insiders. Second, the argument is silent about the use of discriminatory immigration policies when the groups that these policies discriminate against are not already present. If there are no people of Mexican descent in a given society, that society can discriminate against Mexicans seeking admission, for all the argument says. For these reasons, Wellman states that, while he is most drawn to this strategy for explaining why discriminatory immigration policies are morally wrong, he is not fully satisfied with it.<sup>3</sup>

In this paper, I argue that the domestic implications of discriminatory immigration policies are far-reaching and undermine, rather than support, the state sovereignty view. Once we grasp the full extent to which immigration policies are constrained by the principle of equal respect for members, a principle that contemporary forms of the view are committed to, we will see that the view cannot hold on to one of its main distinguishing features—the wide latitude it ascribes to societies in determining and implementing their immigration policies. On considerations of domestic justice alone, my argument shows that the state sovereignty view cannot serve as a satisfactory framework for the normative assessment of immigration policies. Notably, this argument differs from those offered by critics of the view that favor open borders, who have often challenged its partiality toward members. This article also gives a unified explanation of how historical and hypothetical immigration policies discussed in the ethics of immigration literature could express disrespect for members of society. This explanation draws on the existence of the external relationships of members that are grounded in identification with nonmembers on the basis of a shared quality or set of qualities.<sup>4</sup> In conclusion, I suggest that the existence of these external

2 Cole and Wellman, *Debating the Ethics of Immigration*.

3 Cole and Wellman, *Debating the Ethics of Immigration*, 150.

4 The title of this article may be taken to suggest that nonmembers will always be “across borders” from members, or outside of the polity. In actuality, of course, societies often contain

relationships has additional implications for the ethics of immigration that have yet to be fully explored, and that examining this terrain will be essential in developing a fully satisfactory framework for the normative assessment of immigration policies.

### 1. THE STATE SOVEREIGNTY VIEW

The state sovereignty view regarding immigration policy accords a great deal of latitude to societies to exclude or give less than equal treatment to some nonmembers who wish to enter and become full members. In this paper, I at times refer to these kinds of policies as “discriminatory” immigration policies, since they involve discrimination against certain nonmembers. I will not discuss restrictive immigration policies that would not, at least in any obvious way, discriminate against a particular group of immigrants, such as a policy of excluding all potential newcomers, or a policy of capping immigration at a desired number. The view that I am concerned with here is committed to the sovereignty of states to determine immigration policies as they see fit, with very few exceptions. This view continues to be endorsed by prominent theorists working on the ethics of immigration and border control.<sup>5</sup> As we will see, state sovereignty has been thought to entail the moral discretion of states to discriminate against immigrants on the basis of features such as race and ethnicity.

Two of the most prominent contemporary defenses of the state sovereignty view are due to Michael Walzer and Christopher Heath Wellman.<sup>6</sup> According to both theorists, members of society, understood as a political entity or state, are given priority in the determination of that society’s immigration policies. Wal-

---

people who are not full members, such as temporary migrants. For the purposes of this paper, my focus is on nonmembers who are not yet present within a society and policies concerning whether they will be admitted and given membership. There are many interesting questions pertaining to how societies should respond to people who are already present within their borders but who are not full members; I cannot address them all here. I also refer to members and nonmembers throughout the paper, rather than citizens and noncitizens, for ease of exposition. These two categories are treated, for my purposes, as mutually exclusive and exhaustive, and nothing in my argument turns on the existence of other categories of membership that fall between them.

- 5 My argument does not conclude that it is morally impermissible for states to restrict immigration in general. More moderate views that justify only a highly circumscribed right to restrict immigration are not my target here. For three such views, see Miller, “Immigrants, Nations, and Citizenship”; Pevnick, *Immigration and the Constraints of Justice*; and Blake, “Immigration, Jurisdiction, and Exclusion.”
- 6 For one of the earliest philosophical attempts to defend the state sovereignty view, see Sidgwick, *The Elements of Politics*.

zer's piece on immigration policy is the *locus classicus* of the state sovereignty view in the ethics of immigration literature.<sup>7</sup> He argues that if we think that individuals have a right to form self-determining societies we must believe that the members of these societies have a collective right to control and restrain the influx of immigrants. The right of a society to determine its own immigration policy, in other words, is necessary for the sovereignty and security of that society as a distinct political community. Admission and exclusion "suggest the deepest meaning of self-determination."<sup>8</sup> All else being equal, then, a society is permitted to implement discriminatory immigration policies. According to Walzer, as long as Australia gave up some of its unused territory to persons seeking to set up a society where non-whites could enter and become members, the White Australia policy would have been morally permissible to keep in place. This was a policy of excluding non-whites as candidates for immigration to Australia, instituted at the time of Australian federation in 1901 and gradually dismantled between 1949 and 1973. Walzer does not commend White Australia, of course, but believes that "White Australia could survive only as Little Australia."<sup>9</sup> Morally regrettable as it might have been, this racially discriminatory immigration policy did not in itself wrong anyone. For Walzer, societies do have to take in at least some refugees and perhaps certain family members of insiders, but policies that discriminate against immigrants on the basis of race or ethnicity are not morally wrong *per se*.<sup>10</sup>

Wellman has recently offered a prominent defense of the state sovereignty view on grounds of freedom of association.<sup>11</sup> He also emphasizes the rights of individuals to form self-determining societies, but his argument centers on freedom of association as an integral component of self-determination. As in the case of individuals, a society's freedom of association entitles it to associate or not associate with persons as it sees fit. Legitimate societies, ones that respect the human rights of their constituents and all other persons, are entitled to self-determination and hence freedom of association for Wellman. Taking seriously freedom of association at the level of societies entails that societies are entitled to wide latitude in setting and administering their immigration policies. In contrast with Walzer, Wellman holds that a state is not required to admit

7 Walzer, *Spheres of Justice*, ch. 2.

8 Walzer, *Spheres of Justice*, 62.

9 Walzer, *Spheres of Justice*, 47.

10 Walzer, *Spheres of Justice*, 48–51.

11 Wellman, "Immigration and Freedom of Association"; Cole and Wellman, *Debating the Ethics of Immigration*.

refugees so long as there are other ways of helping them.<sup>12</sup> He regards family members and romantic partners of insiders as perhaps the only exception to the nearly unqualified right to exclude.<sup>13</sup> Most importantly for my purposes, as I noted above, he is inclined toward Blake's insider-focused explanation of why discriminatory immigration policies are morally wrong, but not fully satisfied that this explanation works.

As I noted above, both defenses of the state sovereignty view ultimately rely on the rights of members as the source of a society's prerogatives to discriminate against immigrants.<sup>14</sup> The rights of members to form and maintain self-determining societies generally trump any rights that nonmembers might have to enter or receive equal treatment. We should also note that both defenses are intended to work within liberal egalitarian political theory, which is committed to the equality of members of society.<sup>15</sup> Walzer and Wellman deny that, as critics have asserted, the rights of nonmembers limit the moral discretion of societies in determining their immigration policies.<sup>16</sup> I will avoid this impasse regarding the relative weight of the rights of members and nonmembers in matters of immigration policy. I will instead grant Walzer and Wellman the premise that the rights of members trump the rights of nonmembers for the sake of argument, and argue that a core principle of liberal egalitarianism, the principle of equal respect for members, is part of the explanation of a deep internal problem with the state sovereignty view. The argument relies only on grounds of domestic justice, and concludes that even granting this premise, it can be demonstrated that the view is not a satisfactory framework for the normative assessment of immigration policies.

12 Wellman, "Immigration and Freedom of Association," 109–41; Cole and Wellman, *Debating the Ethics of Immigration*, 117–24.

13 Cole and Wellman, *Debating the Ethics of Immigration*, 92.

14 David Miller also defends a state's sovereignty regarding immigration in general. However, Miller regards it as unjust to discriminate against immigrants on the basis of race or ethnicity, although he denies that there is a human right against such treatment. See Miller, "Border Regimes and Human Rights," 19.

15 On Walzer's liberal credentials, it is worth noting that he regards his criticisms of certain liberal theories of justice, sometimes referred to as "communitarian" objections (Walzer has sought to distance himself from this label), as being "available for incorporation within liberal (or social democratic) politics." See Walzer, *Thinking Politically*, 97. In *Spheres of Justice*, Walzer offers a social democratic model of justice that, as we see from the above quotation, he does not regard as at odds with liberalism. Even if we leave these points aside, his view is also meant to apply to the immigration policies of liberal societies.

16 See, e.g., Carens, "Aliens and Citizens: The Case for Open Borders"; Abizadeh, "Democratic Theory and Border Coercion: No Right to Unilaterally Control Your Own Borders."

## 2. EXTERNAL RELATIONSHIPS AND IDENTIFICATION WITH NONMEMBERS

I will use the term “external relationship” to capture the notion of a morally significant relationship that exists between at least one member of a society and at least one nonmember. For my purposes, it will suffice to characterize a relationship as a tie or connection between two or more parties that exists due to a quality or qualities that some or all of these parties possess. External relationships, then, are ties or connections between members and nonmembers that exist due to qualities that they possess and that are morally significant. Some members of modern democratic societies participate in relationships with nonmembers—they are tied to or connected with these nonmembers—in virtue of accepting shared cultural or religious norms. For instance, many Mexican-Americans identify with Mexican nationals on the basis of treating shared cultural experiences and history as valuable, and many Jewish-Americans take themselves to have a special tie to the Israeli people on the basis of valuing a shared ethnic and religious tradition. These relationships have wholly domestic analogues; there are similar kinds of relationships in which the parties are all members of the same society. We can think of these other wholly domestic relationships as “internal relationships.”

The kinds of external relationships that I focus on here, like the ones mentioned above, are relationships with nonmembers that members seek to enter or remain in. Further, this paper only addresses external relationships that members seek to enter or remain in that involve identification with nonmembers. Other kinds of external relationships exist, and some are clearly morally significant. In particular, I have in mind what I referred to above as “relationships of closeness,” including family relationships, romantic relationships, friendships, and perhaps other forms of partnership.<sup>17</sup> Because Walzer and Wellman are both willing to make concessions for certain relationships of closeness, I focus in this article instead on relationships of likeness, and in particular external relationships that involve identification with nonmembers.

My thought is that a member’s sense that they themselves are being respected,

17 These external relationships also constrain a society’s immigration policies by giving rise to demands of domestic justice. See Matthew Lister’s discussions of the importance of family reunification (“Immigration, Association, and the Family”) and the demands of equality for same-sex couples seeking marriage-based admissions (“A Rawlsian Argument for Extending Family-Based Immigration Benefits to Same-Sex Couples”). See also Luara Ferracioli’s argument that liberal societies must extend similar immigration benefits to friends and creative partners of members if they extend them to family members and spouses (“Family Migration Schemes and Liberal Neutrality: A Dilemma”). Examining the case for these arguments is beyond the scope of the present paper.

disrespected, helped, or harmed can depend in part on how nonmembers whom they identify with are being treated. Identification of this kind involves seeing oneself as sharing a quality with other people such that when they are treated in a certain way on the basis of possessing this quality this treatment is taken to express certain attitudes about oneself. For instance, if John is a Catholic and a shop is known to have a policy of refusing to hire Catholics, John may regard the policy as a personal affront, even if he has no intention of seeking employment at the shop. The shop's policy is most naturally interpreted as expressing the attitude that Catholics in particular are not wanted around the shop and John, in identifying with other Catholics, would justifiably feel that the policy expressed disrespect for him for this reason.

To be sure, there are senses of "identifying with" others that may not give rise to this sort of indirect expression of attitudes. Anne may be a Protestant who identifies with the Catholics who are being discriminated against by the shop's employment policy in the sense that she sees their struggle as one that she is personally invested in. Nonetheless, the shop's policy does not express disrespect for Anne, or at least not in the same way that it expresses disrespect for John. To give a rough characterization, two conditions are jointly sufficient for the treatment of others to potentially express attitudes about oneself. First, there is *identification as* possessing a certain quality possessed by the other person or persons, and second, there is *identification with* the other person or persons on the basis of their possessing this quality.<sup>18</sup> Both conditions are factive with respect to the possession of the quality on the basis of which other persons are being discriminated against. In the example just given, John identifies as a Catholic and also identifies with other Catholics on the basis of this fact. Anne, by contrast, does not identify as a Catholic, and cannot be said to identify with Catholics except in an attenuated sense of the term. My claim is that when both conditions obtain for a person, all else being equal, discrimination against other persons on the basis of the relevant quality will also express disrespect for that person. For ease of exposition, I will refer to this way of relating to others as "identifying with" them. It is this sense of the term that more plausibly captures the notion of group iden-

18 This is compatible, of course, with discrimination toward insiders being a result of discrimination toward outsiders even in cases in which insiders do not identify as members of the targeted group. My claim is not that either *identification as* or *identification with* are necessary conditions for the occurrence of indirect discrimination of this kind. In cases such as that of one of the plaintiffs in *Romer*, a person may be discriminated against for a quality that they are thought by others to possess (i.e., being gay) even when they do not in fact possess that quality. *Romer v. Evans*, 517 U.S. 620 (1996). There is no reason why immigration laws could not similarly target some members of society who are merely thought to possess a given quality. I am grateful to an anonymous reviewer for urging me to clarify this point.

tification, even if we should allow that there are other senses of identifying with others that do not involve identification as sharing the same quality or qualities nor identification with them on this basis.<sup>19</sup>

The phenomenon of identifying with others helps explain why statements of the form “I regard you as an exception, a good one” when the speaker is addressing a member of some group that the speaker has a prejudice against are rarely comforting to the addressee. The fact of not being the direct object of discrimination does not cancel the fact that discrimination against other members of one’s group expresses disrespect for oneself when one identifies with them. Additionally, while these two conditions are jointly sufficient for indirect expressive disrespect of the kind just discussed to be possible, discrimination against others who share a quality that one possesses may precipitate identifying with these other persons, or doing so more strongly. There are many interesting questions that one can take up regarding these notions that I do not have the space to engage at length here. I will focus on cases in which identification with others is relatively stable and, as noted earlier, the relationship of identification with nonmembers already exists. Before giving my argument in more detail, I must first say a bit more about how I am understanding respect and disrespect in this context.

### 3. THE PRINCIPLE OF EQUAL RESPECT FOR MEMBERS

The principle of equal respect for members is a core principle of liberal egalitarianism, and gives rise to important demands of domestic justice.<sup>20</sup> I will not try to say too much in defense of the principle or any particular version of liberalism. But I will say a bit to specify how I understand the principle. In doing so, I am not committed to the view that other, slightly different versions of the principle would not work equally well for my purposes. The principle is a central commitment of liberal egalitarianism but admits of different specifications. Any plausible version or interpretation of the principle will rule out discriminatory immigration policies as expressing disrespect for members in instances where members are also discriminated against by the policies.

19 On the identification-based elements of group membership, see, e.g., Appiah, *The Ethics of Identity*, and Kymlicka, *Multicultural Citizenship*.

20 Elizabeth Anderson holds that expressing equal respect and concern for all citizens is “the most fundamental test any egalitarian theory must meet” (“What Is the Point of Equality?” 289). Ronald Dworkin also argues that a right to equal concern and respect is the most fundamental right of persons. See Dworkin, *Taking Rights Seriously*. I focus in this paper on equal respect. It is an open question what further conditions on immigration policies the requirements of equal concern could justify.

The principle of equal respect for members holds that society's political institutions, policies, and members in their official capacities must treat all members of society with equal respect and, when applicable, must express equal respect for them. I tend to view these requirements in terms of recognition respect.<sup>21</sup> Respecting or expressing respect for someone in this sense requires giving appropriate consideration or recognition to the fact that they are a person and constraining one's conduct in ways that are morally required by this fact.<sup>22</sup> What this concretely entails will depend on the particular actions or policies at issue and the circumstances in which they are executed. In this article, I focus primarily on immigration policies and their expressive significance, the attitudes that the policies are most naturally interpreted as expressing about members. Immigration policies can express attitudes not only about potential immigrants: that they are wanted or not wanted by the society, that they are on par with others seeking to enter or less valued, to take some examples. They can also express similar attitudes about members of society when members identify with nonmembers on the basis of qualities that they share and these qualities are picked out for discriminatory treatment by the policies.

If equal respect for members is indeed an important principle of liberal egalitarianism, it follows that proponents of the state sovereignty view, strictly speaking, have two choices. One is to give up their commitment to this principle, endangering the liberal credentials of their view. The other is to show that my argument can be answered. The second option is the obvious one for any liberal theorist to take, but it is worth flagging what is at stake in considering the first option. Because I do not regard the first option as viable for contemporary proponents of the state sovereignty view, who claim that their view is consistent with liberal principles, I will not discuss it further.

#### 4. IDENTIFICATION WITH NONMEMBERS AND EQUAL RESPECT FOR MEMBERS

With this theoretical background in place, we can demonstrate the significance of external relationships in which members identify with nonmembers for the normative assessment of immigration policies. Walzer and Wellman defend the state sovereignty view and rely on the premise that the rights of members ultimately justify the wide latitude that the view attributes to societies over how they treat nonmembers seeking entrance. Yet appeal to the standing of members as political equals supports a powerful argument against discriminatory immi-

21 The distinction between recognition respect and appraisal respect is due to Stephen Darwall. See Darwall, "Two Kinds of Respect."

22 Darwall, "Two Kinds of Respect," 45.

gration policies to an extent that renders the state sovereignty view untenable. Immigration policies that would discriminate against certain nonmembers on the basis of their race, ethnicity, nationality, religion, gender, gender identity, sexual orientation, or disability can be morally impermissible to implement on grounds of domestic justice alone. This will be the case when members of society identify with these nonmembers on the basis of the criteria that the policy would use to discriminate against the nonmembers.

First, we consider policies of outright exclusion on the basis of these criteria. We can imagine a society in which 80 percent of the population wishes to implement a policy of refusing entrance for a certain group of nonmembers, picked out for their race, ethnicity, nationality, religion, gender, gender identity, sexual orientation, or disability. On the other hand, 20 percent of the population is composed of members who identify with and are known to identify with these nonmembers on the basis of sharing the quality that they would be excluded for possessing if the policy were implemented.<sup>23</sup> Even if we bracket the claims of the nonmembers in this case for the sake of argument, it would still be impermissible for the majority to impose the policy of refusing entrance for this group of nonmembers, all else being equal. The principle of equal respect for members justifies this judgment. The policy proposal is most naturally interpreted as expressing the attitude that nonmembers in this group are not wanted as potential members of society, and additionally that members who identify with them are not valued as members of society. This second attitude, the one that is directly relevant as a matter of domestic justice, expresses disrespect for members, violating the expressive requirement of the principle of equal respect for members.

In slightly different terms, policies of outright exclusion can express attitudes about persons who constitute a distinct group or a “we” that members of society also see themselves as part of. To paraphrase Michael Dummett’s critical encapsulation, “keep them out, but treat them decently if they are already here”<sup>24</sup> is a discriminatory attitude regarding a group or a “we” consisting both of persons to be kept out, nonmembers, and persons who are already present in society, members in this case. Discrimination against members or “negative discrimination” within a society is morally impermissible, and the principle of equal respect for members justifies this judgment. Policies of outright exclusion can therefore be

23 Of course, these percentages are implausible for some of these qualities in most actual societies, but the point that I am making is not affected by this fact.

24 Dummett’s phrase is “keep them out, but treat them decently if they do get in.” See Dummett, *On Immigration and Refugees*, 111. Dummett is here discussing what he refers to as the “familiar principle” behind the second Race Relations Act passed by the United Kingdom’s Parliament in 1976.

impermissible to implement on grounds of domestic justice. The implication that members' presence in society is regrettable and that if only they were not already present they should be refused just like the others with whom they identify is deeply disrespectful of them.

External relationships significantly limit the moral discretion that societies have over their immigration policies, then, on grounds that even proponents of the state sovereignty view are committed to. The relationships that I have focused on involve identification with nonmembers on the basis of sharing the same race, ethnicity, nationality, religion, gender, gender identity, sexual orientation, or disability. To exclude nonmembers based on these qualities is also to express disrespect for members who possess them.

What about policies that would give preferential treatment to certain nonmembers? Considering policies of outright exclusion may help us to see that immigration policy is constrained by the principle of equal respect for members most clearly. However, it is important to establish that a similar line of argument shows that external relationships grounded in identification and the principle of equal respect for members also render "positive discrimination," discrimination in favor of certain immigrants over others, impermissible in many cases. To use a case that parallels the previous one involving a policy of outright exclusion, imagine that an 80 percent majority aims to implement an immigration policy that would give preferential treatment to nonmembers whom they identify with and are known to identify with. However, a 20 percent minority identifies with and is known to identify with another group of nonmembers. The policy that the majority seeks to implement in this case, all else being equal, would also violate the principle of equal respect for members. Drawing on the previous discussion, the expression of an attitude regarding certain nonmembers coincides with the expression of an attitude about members who identify with and are known to identify with them. By expressing the attitude that certain nonmembers are the kinds of persons that the society wishes to admit more of than others, the policy also expresses the attitude that the members who identify with the preferred nonmembers are more valuable members of society. "We should have more members like us than members like you" is naturally interpreted by the minority as the kind of attitude behind the policy being proposed by the majority. Positive discrimination in this case also implies a negative comparative judgment about the value of members of society in the minority.

I have again deliberately left the basis on which members identify with particular groups of nonmembers open in this case. As suggested earlier, identification on the basis of sharing the same race, ethnicity, nationality, religion, gender, gender identity, sexual orientation, or disability gives rise to possible violations

of the principle of equal respect for members by immigration policies. Of course, there are likely to be important differences between these criteria in terms of the expressive significance of their use in immigration policies. Yet policies that would employ these criteria as reasons for excluding or giving less than equal treatment to nonmembers share the important feature of potentially violating the principle of equal respect for members. If we accept the principle of equal respect for members, we recognize that societies have a duty not to implement policies that express disrespect for members. Policies of giving less than equal treatment to certain nonmembers can express the attitude that some members of society are more or less valuable than others as members.

The state sovereignty view, then, contains a deep inconsistency, at least in its contemporary forms. The justification for the wide latitude over immigration policy that it attributes to societies purports to rest on giving members their due morally. Yet even if we grant for the sake of argument that the rights of members generally trump the rights of nonmembers in matters of immigration policy, we should deny the claim that societies have wide latitude in this policy area. Indeed, societies are greatly morally constrained when setting their immigration policies by the requirements of the principle of equal respect for members. I have argued, in particular, that societies may not discriminate against nonmembers—exclude them or give them less than equal treatment—on the basis of their race, ethnicity, nationality, religion, gender, gender identity, sexual orientation, or disability when doing so would express disrespect for persons who are already members of society. Given the racial, ethnic, national origin, and religious diversity in modern democracies, and the presence of gender diversity, trans people, sexual minorities, and disabled persons in every society, these moral reasons nearly always come into play. Notably, they are reasons that proponents of the state sovereignty view are committed to recognizing in general, and they undermine the commitment that makes the view distinctive. If societies do not have the moral discretion to exclude or give less than equal treatment on the basis of this range of criteria, they are greatly constrained by the demands of morality in setting their immigration policies. Of course, for those of us who are not committed to the state sovereignty view, these may seem to be merely additional reasons against discriminatory immigration policies. Yet it is important to recognize that the state sovereignty view can be shown to be an unsatisfactory framework for the normative assessment of immigration policies from within, by an argument that meets its proponents on their own terms. Additionally, I will suggest that the focus on external relationships and the principle of equal respect for members has implications that are of independent interest beyond this internal critique of the state sovereignty view.

## 5. OBJECTIONS AND REPLIES

I will now consider and respond to some of the objections to my argument that are likely to be raised. Along the way, I will also discuss the scope of the argument and some of the considerations that can render exclusion or less than equal treatment of certain nonmembers permissible. Lastly, I will discuss some of the implications that external relationships and the principle of equal respect for members have for the permissibility of immigration policies independently of the argument against the state sovereignty view.

As mentioned earlier, Wellman generally regards attempts to respond to cases of outright exclusion by appealing to the demands of domestic justice as helpful to the state sovereignty view.<sup>25</sup> However, he has also raised objections to such attempts and stated that he is not fully satisfied with them.<sup>26</sup> The discussion in the ethics of immigration literature on whether policies of outright exclusion on the basis of race or ethnicity are morally permissible arose largely in response to Walzer's discussion of the White Australia policy. As mentioned earlier, Walzer holds that this policy of banning non-whites, however morally regrettable, was not morally impermissible to implement in itself. As long as Australia ceded any large areas of unused territory to persons seeking to establish a society that non-whites could enter and become members of, non-whites could be excluded on the basis of their race or ethnicity. Responding to Walzer, Michael Blake was one of the first commentators to argue that, given the fact that Australia was not and is not an ethnically homogenous nation, the White Australia policy was morally problematic domestically.<sup>27</sup> Blake holds that whenever "there are national or ethnic minorities—which is to say, the vast majority of actual cases—to restrict immigration for national or ethnic reasons is to make some citizens politically inferior to others."<sup>28</sup> The goal of eliminating the presence of a group from society through selective immigration is insulting to members of that group already present in society, and thus morally problematic. Blake did not take this argument to undermine the state sovereignty view and Wellman at first accepted it as a way of accommodating the intuition that the White Australia policy was morally wrong.<sup>29</sup> Notably, this shows that Wellman is not merely concerned to argue for the less interesting and ambitious claim that third parties in general should not coerce societies to adopt morally acceptable immigration policies,

25 Wellman, "Immigration and Freedom of Association," 139–41.

26 Cole and Wellman, *Debating the Ethics of Immigration*, 149–50.

27 Blake, "Immigration."

28 Blake, "Immigration," 233.

29 Wellman, "Immigration and Freedom of Association," 139–40.

which even proponents of open borders can endorse. Rather, he is concerned with what is morally wrong, which is a distinct question from what can be prevented or punished by third parties.<sup>30</sup> In later work, however, Wellman states that he does not find Blake's line of argument fully satisfactory because it "justifies too little and too much."<sup>31</sup> I will explain and examine each of these charges separately. Because of the concerns that I have with Blake's argument and the fact that my argument is directed against the state sovereignty view rather than at defending it, I will respond to these objections on behalf of my own argument.

Wellman's objection that the argumentative approach of appealing to the demands of domestic justice "justifies too little" points to the fact that it focuses on existing members of society who belong in some way to the group to be excluded. In a state that is entirely devoid of members who are in the group, Wellman states that the argument does not deliver any verdict.<sup>32</sup> What would be wrong, for instance, with an entirely homogenous white society explicitly excluding all prospective non-white immigrants? Given that most of us would find such a policy morally abhorrent, we would like to have a good argument for the view that racially or ethnically discriminatory policies would be morally impermissible to implement even for a racially or ethnically homogenous society. Yet it does not seem that this approach can provide us with an argument against them.

It is correct to point out that the scope of Blake's argument, if used to defend the state sovereignty view, is limited in important ways. If the goal is to show that immigration policies involving outright exclusion on the basis of race or ethnicity are morally impermissible in general, it cannot be met with Blake's argument. In fact, there are three related but distinct concerns along these lines. First, as Wellman points out, the argument does not establish that discriminatory immigration policies would be morally impermissible for a fully homogenous society to implement.

Second, even ethnically diverse societies would not be prohibited by the argument from discriminating against persons on the basis of race or ethnicity *per se*, so long as persons of the relevant races or ethnicities were not present within the society as members. Wellman's own endorsement of Blake's argument in his 2008 article had a similar tendency to overgeneralize, holding that "because no state is completely without minorities who would be disrespected by an immigration policy which invoked racial/ethnic/religious categories, no state may

30 For an interesting discussion of whether democratic societies could have a moral right to implement wrongful policies and institutions, see Barry and Øverland, "Do Democratic Societies Have a Right to Do Wrong?"

31 Cole and Wellman, *Debating the Ethics of Immigration*, 149.

32 Cole and Wellman, *Debating the Ethics of Immigration*, 149.

exclude potential immigrants on these types of criteria.”<sup>33</sup> It is true that no state is completely without minority groups, but false that any immigration policy invoking racial, ethnic, or religious criteria must use criteria that some members of society fall under.

A third concern, which also highlights the importance of identification with nonmembers, is that even if there are members of society who fall under criteria that an immigration policy uses to discriminate against nonmembers, that policy still may not express disrespect for them. For a variety of reasons, members may not take it as a sign of disrespect toward themselves if nonmembers are discriminated against on the basis of possessing a quality that they also possess. Perhaps the members in question have lived in the new society for some time, and have ceased to regard their ethnicity as an important part of their sense of self. To take a slightly different case, they may have left their prior society during a time of political upheaval, and now wish to distance themselves from other persons of the same ethnicity seeking to enter. Something else is required beyond the mere fact that members share some quality with nonmembers for the exclusion of nonmembers on the basis of this quality to express disrespect for members. This is a third reason why there may not be any members of society who are disrespected by a policy of exclusion on the basis of race or ethnicity. However, one important reason why such members might be present, and perhaps the paradigmatic reason, is if some of them are in external relationships involving identification with nonmembers. When members are in external relationships involving identification with nonmembers, discrimination against nonmembers on the basis of race, ethnicity, nationality, religion, gender, gender identity, sexual orientation, or disability can express disrespect for these members.<sup>34</sup> In contrast to Blake’s approach, the argumentative approach that I have offered here explains the mechanism by which core cases of discrimination toward outsiders in the form of immigration policies manage to also discriminate against insiders.

Notably, the goal of my argument is not to show that racially or ethnically discriminatory policies are always morally impermissible on grounds of domestic justice. I view these policies as morally impermissible in virtue of wronging

33 Wellman, “Immigration and Freedom of Association,” 140.

34 This is not to suggest that discriminatory immigration policies cannot express disrespect for members unless members identify with nonmembers. For instance, if all the members of society who are in favor of a policy of exclusion intend for it to express disrespect for some members, the policy may do so even if these members do not themselves identify with nonmembers. It may be tempting to describe this as a case in which members are “identified with” nonmembers by fellow members but do not identify with them. Given my purposes in this paper, it suffices to flag the fact that these are distinct notions and that it is identification with nonmembers on the part of members that I am interested in.

nonmembers in general, although this is the topic of a separate paper. My interest in the range of policies that considerations of domestic justice render morally impermissible stems instead from the fact that exclusion or less than equal treatment on the basis of race, ethnicity, nationality, religion, gender, gender identity, sexual orientation, or disability can be morally impermissible on grounds of domestic justice. This suffices to show that societies are greatly constrained by the demands of morality when setting their immigration policies, undermining the state sovereignty view's most significant distinguishing feature, the wide latitude that it attributes to societies in this policy area. While its scope is in a sense more limited, then, my argument points to much broader implications of recognizing the demands of domestic justice for immigration policy and constitutes a powerful internal critique of the state sovereignty view. For this reason, Wellman's "justifies too little" objection has no force against my argument, while this criticism does apply to Blake's argument as a way of defending the state sovereignty view. At this point, I will fully leave Blake's argument aside, having demonstrated the differences between it and my own.

Wellman also raises the objection that the argumentative approach of appealing to demands of domestic justice "justifies too much," in the sense that it commits us to regarding some morally permissible immigration policies as morally impermissible. He provides a hypothetical case involving Norway and persons of Pakistani descent living there that is supposed to bring out this judgment.<sup>35</sup> Wellman has us imagine that Norwegians have a national discussion and collectively decide that while one hundred thousand more Pakistanis would be a welcome addition, no more than one hundred thousand should be allowed in. There is an imagined consensus that taking in more than one hundred thousand Pakistanis might give rise to difficult societal issues due to having a large national origin group with its own cultural practices and traditions present within the society's borders. Reflecting on this case, we are supposed to consider what should be done once one hundred thousand additional Pakistanis have entered Norway, and whether it would be permissible for the Norwegians to decide to prevent further immigration from Pakistan. Wellman states that he is not convinced that doing so would be unjust, even if many members of Norwegian society who had immigrated to Norway from Pakistan would be insulted, and understandably so, by the policy. He thinks that something must be wrong with an immigration policy that excludes prospective newcomers on the basis of their nationality.<sup>36</sup> Yet he is not fully satisfied with the view that this type of immigration policy can

35 Cole and Wellman, *Debating the Ethics of Immigration*, 149–50.

36 I am grateful to an anonymous reviewer for urging me to clarify that this is a case of discrimination on the basis of nationality. Separating discrimination on the basis of nationality and

be shown to be morally impermissible to implement by appealing to demands of domestic justice.

The hypothetical case on offer involves many factors and assumptions that make it hard to test what our intuitions regarding it tell us about my argument. For instance, if Norwegians of Pakistani descent were part of a national consensus in favor of the proposed policy, why would some of these persons at a later time find it insulting to continue carrying it out? However, an important fact about the case makes it possible to accept its setup for the sake of argument and still have grounds to reject the intuition that he reports as evidence that the argumentative approach that I take justifies too much. It is that some of the Norwegians of Pakistani descent who would live in Norway when Pakistanis would start to be excluded would not have been members at the time of the policy's determination, but later would be members. Some of them would not have been born at that earlier time, while others would have been children or adults when the policy was put in place but not yet members of Norwegian society. This includes the one hundred thousand people who would be allowed to immigrate under the policy. Are these people to be fully spoken for by the other Norwegians of Pakistani descent whom we are to imagine having been part of a consensus in favor of the policy? Should their say in this case be entirely beholden to what other persons in their national origin group living in Norway agreed to? These questions, which I do not think Wellman or any other liberal theorist can provide an affirmative answer to, arise even if we grant the setup of Wellman's case for the sake of argument. In other words, even if we grant the premise that the persons who agreed to the policy would have no grounds to complain about its expressive significance later on, this does not imply that persons who did not participate in the national dialogue would have no grounds to do so. A policy of excluding further Pakistani immigrants on the basis of their nationality expresses the judgment that Norwegians of Pakistani descent are not valued members of society. The latter have standing, grounded in the principle of equal respect for members, to reject the policy, even if we suppose that the others who agreed to it do not. We can grant the setup of Wellman's case, then, and nonetheless see that the judgment that the policy of excluding Pakistanis after one hundred thousand more have immigrated could violate the principle of equal respect for members, and therefore be morally impermissible on grounds of domestic justice. Hence the "justifies too much" objection against the attempt to show that policies of outright exclusion are morally impermissible on grounds of domestic justice that Wellman has provided does not go through. The argument that I

---

ethnicity, of course, will be difficult in many real-world cases. Addressing these difficulties would go beyond the bounds of the present article.

have provided allows us to give a principled response to this objection, even if we grant the terms of Wellman's case.<sup>37</sup>

To anticipate further objections, I will make two additional points of clarification. First, my argument concludes that societies cannot implement discriminatory immigration policies when these policies would also involve discrimination against members. However, this should not be taken to suggest that I regard all restrictions on immigration as morally impermissible. Even the most steadfast proponents of the open borders position generally hold that threats to national security and social order, such as enemies of the state and persons with serious criminal records in their country of origin, may be excluded. Societies are certainly permitted to exclude terrorists or members of criminal organizations like the Mafia to protect their members from these threats, even when some such persons are also members of society. Similarly for policies that exclude persons with highly morbid and contagious diseases whose presence would genuinely and significantly threaten public safety, even if these diseases afflict some members of society.<sup>38</sup> The attitudes naturally interpreted as lying behind such policies need not express disrespect for any members of society. Of course, we may have reason to be skeptical of the legitimacy of particular uses of these criteria, for instance if a country used them to exclude only persons from a particular ethnic group or religion. Merely cosmetic attempts to hide discriminatory intentions

- 37 An anonymous reviewer has raised the issue of whether, rather than blanket exclusion after a certain number have entered, Norway could introduce an annual quota to control the rate of migration from Pakistan in order to support immigration integration. On my view, if the sincere goal of this policy is to help Pakistanis integrate into Norwegian society, and that goal cannot be accomplished equally well in some other way, such a policy would not in any obvious way express disrespect for Norwegians of Pakistani descent. Indeed, it would partly be designed to help them, as integration benefits both immigrants and the receiving society. Of course, the policy's goal must be genuine and not a cover for discriminatory attitudes, as the National Origins Formula in the United States had been and was abolished under the Immigration and Nationality Act of 1965, Pub. L. No. 89-236, 79 Stat. 911 (1965), also known as the Hart-Celler Act, largely on that basis. For a detailed discussion of this transitional period in American immigration law, see Joppke, *Selecting by Origin*, ch. 2. I am grateful to the reviewer for raising this point.
- 38 In addition to permitting these grounds for exclusion, which even open borders advocates allow, certain forms of preferred treatment in favor of particular groups can be permissible on my view. For instance, it is plausible that a society may give a preferred immigration status to persons who have been victims of injustices that it has perpetrated without expressing the attitude that any members of society are less than equal to one another. This is perhaps part of justification for the Orderly Departure Program, created in 1979 under the United Nations High Commissioner for Refugees, which involved the United States taking the lion's share of Vietnamese refugees after the Vietnam War. I am grateful to two anonymous reviewers for urging me to clarify these points.

are not significantly different from policies that involve upfront discrimination morally. These policies would still violate the principle of equal respect for members, whereas policies genuinely grounded in realistic concerns for national security or social order would not.<sup>39</sup>

Second, my argument is compatible with a plausible commitment to responsibilities of immigrants to integrate into the receiving country's political culture in various ways.<sup>40</sup> Indeed, it is compatible with the view that reasonable adjustment is required both of immigrants and members of the receiving society. This may have been another confounding factor in Wellman's case, where it is presumed that the Norwegians of Pakistani origin must not be allowed to increase their numbers by more than one hundred thousand or else difficult social and political issues might arise. If this were a genuine worry, we might wonder whether members of that community were doing their part to integrate into Norwegian society. If they were not, some immigration policy that would reflect this problem could be justified, but it is far from clear that exclusion of all persons of Pakistani origin would be the morally and practically preferred choice. It is even further from the truth, as far as I can tell, that Norwegians would regard this example as realistic. Indeed, Norway is one of the countries most committed to equality and nondiscrimination, and Pakistani-Norwegians are well-integrated on the whole. Generous immigration policies of societies like Norway are testament to their commitments to these values and the possibility of mutual adjustment in a world in which movement is often necessary. The challenges we face in advancing together as fellow members of liberal societies should be surmounted by policies and approaches that are consonant with our own principles. Exclusion or less than equal treatment on the basis of race, ethnicity, nationality, religion, gender, gender identity, sexual orientation, or disability would conflict with the principles that we strive to secure in the name of liberal values.

Moreover, even if one accepts my argument against the state sovereignty view, one might wonder whether external relationships that involve identification and the principle of equal respect for members have further implications for the ethics of immigration. If someone is inclined to regard any restrictions on entry as impermissible, aside from restrictions that are genuinely required to maintain national security or social order, of course these considerations will not add much more than additional reasons against restrictive policies. However,

39 As Sarah Fine points out, the long history of racial and ethnic criteria being used in the immigration policies of liberal democracies may mean that employing skill-based or other types of criteria in contemporary societies in particular ways will be morally problematic. See Fine, "Immigration and Discrimination."

40 See Miller, "Immigrants, Nations, and Citizenship," for an account of these responsibilities.

I do think that cases can be given where a moderate position would regard certain policies as permissible to implement when only the claims of nonmembers are in view, but impermissible once the argument from disrespect for members is added. Consider a policy of excluding immigrants who have disabilities that would be costly to accommodate in terms of healthcare resources. On the face of it, it may seem reasonable for societies to appeal to the cost of taking certain groups of immigrants in when deciding whom to admit. The policy involves appeal to excessive costs rather than exclusion simply on the basis of having a disability, where the latter may seem morally impermissible to us but the former is not obviously so. When we consider the fact that there are already persons in the society with disabilities that involve similar medical costs to the projected costs that would be used to justify exclusion, the moral situation arguably changes. This is because the policy expresses the attitude that members with these disabilities are a burden on society, contributing less than they receive. Indeed, the Canadian immigration policy of using an excessive demand clause to exclude nonmembers with costly disabilities has been opposed, and I think rightly, by the Council of Canadians with Disabilities partly on the grounds of these kinds of considerations.<sup>41</sup> It may be reasonable for societies to consider costs in terms of societal resources when setting their immigration policies, but excluding persons with costly disabilities expresses disrespect for members, and is not the only way of managing these costs. It is arguably also the case that policies that would exclude immigrants on the basis of lacking financial resources might be permissible, even if ungenerous, when only the claims of outsiders are taken into account, but impermissible in virtue of expressing disrespect for poor persons within society.<sup>42</sup> For similar reasons, it may be impermissible to exclude certain immigrants on the basis of age, for instance in skilled-worker admissions.<sup>43</sup>

41 Immigration and Refugee Protection Act, 2001 S.C., ch. 27, s. 38(c). The Council of Canadians with Disabilities has stated that “the current law devalues Canadians with disabilities and does nothing to recognize the contribution persons with disabilities and their families can and do make to Canadian society.” “Immigration,” accessed August 30, 2014, <http://www.ccdonline.ca/en/socialpolicy/immigration>.

42 Joseph Carens holds that this sort of policy is not unjust if, contrary to his own considered view, states are morally entitled to control their borders. See Carens, *The Ethics of Immigration*, 179. With regard to the use of skill-based criteria in immigration, Ayelet Shachar raises important questions about the risks that the use of skill as a qualification for acquiring citizenship may pose to the stability of political equality and shared notions of societal belonging among citizens. See Shachar, “Selecting By Merit.”

43 Of course, in the case of spousal reunification, policies that set lower limits on age can be morally permissible and need not express disrespect for any members of society. Such policies may have the legitimate aim of protecting people who would be brought in as underage brides. However, the Danish Immigration Service’s “24-year rule,” which set the age restric-

Such policies are typically justified in terms of long-term productivity but have been opposed on the grounds that they involve problematic age discrimination against members of society.<sup>44</sup> I cannot discuss these cases at length here, but they suggest paths that can be explored with the argumentative resources used in this article that go beyond the argument against the state sovereignty view.

Additionally, the option of embracing the view that disrespect for even one member is morally significant is compelling, I hold, in thinking through the implications of equal respect for members. To build on the previous example involving disability, there being only one member of society with a costly disability makes the difference between the policy being morally impermissible and being morally permissible to implement, all else being equal. It may be worse in general to discriminate against larger numbers of members than smaller numbers, but the moral difference between zero and one disabled member of society, for instance, is greater here than the difference between forty-nine and fifty. If we take equal respect for members seriously, then this requires that policies not express disrespect for any members unless there is a very significant moral reason to do so. It would be hard to make the case that the only way of managing the potential costs of allowing nonmembers in is to exclude nonmembers with costly disabilities outright. Again, I must leave the details of this sort of proposal for further work. Still, it is an important consequence that disrespect for even one member of society would be this significant, which should be explored outside of the argument against the state sovereignty view.

## 6. CONCLUSION

I argued in this article that the state sovereignty view in the contemporary ethics of immigration literature is unsatisfactory. This view grants societies wide

---

tion for spousal reunification at twenty-four in Denmark, has come under sharp criticism on the basis that it expresses discriminatory attitudes toward Muslims and is ineffective in achieving its stated purpose of protecting women in forced marriages. As I noted above, the use of age and other criteria can violate the principle of equal respect for members when they are used as a cover for discriminatory attitudes toward particular social groups. I am grateful to an anonymous reviewer for encouraging me to consider the permissibility of different uses of age as a criterion in immigration policy.

44 The Australian organization COTA (formerly the Council on the Ageing), which is intended to represent elderly Australians, has opposed the use of fifty years of age as a cutoff for skilled worker migrants on the grounds that it conflicts with the Age Discrimination Act of 2004, even though the act itself treats migration as an exempt category. See Bolton, "How Old Is Too Old to Become a Migrant?" Opponents of this age restriction also note that similar restrictions are not used in the United States or the European Union. See Biggs, Fredvang, and Haapala, "Not in Australia."

latitude in determining and implementing their immigration policies. The arguments that Walzer and Wellman provide in defense of the state sovereignty view both rely on the premise that the rights of members generally trump the rights of nonmembers in matters of immigration policy. Yet even if we grant them this premise for the sake of argument, it does not follow that societies can refuse or prefer nonmembers at their discretion. If members of a society are parties to external relationships involving identification with nonmembers on the basis of a shared race, ethnicity, nationality, religion, gender, gender identity, sexual orientation, or disability, immigration policies that discriminate against nonmembers on the basis of these qualities may express disrespect for members. Hence societies are greatly constrained by the demands of domestic justice when setting their immigration policies on grounds that are internal to the state sovereignty view, and the view is internally inconsistent. After giving this argument, I discussed and responded to objections that Wellman has raised against the argumentative approach that I take, anticipated further objections, and suggested additional implications of the approach that would go beyond the argument against the state sovereignty view. In the absence of further objections, it seems that we should put aside the state sovereignty view when attempting to determine what types of immigration policies are and are not morally permissible for societies to implement.<sup>45</sup>

Brooklyn College, CUNY  
Australian National University  
matthew.lindauer@anu.edu.au

#### REFERENCES

Abizadeh, Arash. "Democratic Theory and Border Coercion: No Right to Unilaterally Control Your Own Borders." *Political Theory* 36, no. 1 (2008): 37–65.

45 Earlier versions of this paper were presented at the Moral, Social, and Political Theory (MSPT) Seminar at the Australian National University and the Yale Works in Progress Seminar. For comments and discussion, I would like to thank Ray Briggs, Joe Carens, Eric Guindon, Ten-Herng Lai, Holly Lawford-Smith, Christian List, Mark Maxwell, Gerhard Øverland, Jason Stanley, and Ken Winkler. For written comments, I am especially grateful to Christian Barry, Pam Corcoran, Steve Darwall, Luara Ferracioli, Zoltán Gendler Szabó, Shelly Kagan, Serene Khader, Josh Knobe, Seth Lazar, Chris Lebron, Shmuel Nili, Sara Protasi, Daniel Putnam, Julian Reid, Barbara Sattler, Sun-Joo Shin, Jiewuh Song, Laura Valentini, Yuan Yuan, and two anonymous reviewers for this journal.

- Anderson, Elizabeth S. "What Is the Point of Equality?" *Ethics* 109, no. 2 (January 1999): 287–337.
- Appiah, Kwame Anthony. *The Ethics of Identity*. Princeton: Princeton University Press, 2004.
- Barry, Christian, and Gerhard Øverland. "Do Democratic Societies Have a Right to Do Wrong?" *Journal of Social Philosophy* 42, no. 2 (Summer 2011): 111–31.
- Biggs, Simon, Marthe Fredvang, and Irja Haapala. "Not in Australia: Migration, Work, and Age Discrimination." *Australasian Journal on Ageing* 32, no. 2 (June 2013): 125–29.
- Blake, Michael. "Immigration." In *A Companion to Applied Ethics*, edited by R. G. Frey and Christopher Heath Wellman, 224–37. Malden, MA: Blackwell Publishing, 2003.
- . "Immigration, Association, and Antidiscrimination." *Ethics* 122, no. 4 (July 2012): 748–62.
- . "Immigration, Jurisdiction, and Exclusion." *Philosophy and Public Affairs* 41, no. 2 (Spring 2013): 103–30.
- Bolton, Andrew. "How Old Is Too Old to Become a Migrant?" SBS World News Australia Radio. October 24, 2012. Updated August 26, 2013. <http://www.sbs.com.au/news/article/2012/10/24/how-old-too-old-become-migrant>.
- Carens, Joseph H. "Aliens and Citizens: The Case for Open Borders." *Review of Politics* 49, no. 2 (Spring 1987): 251–73.
- . *The Ethics of Immigration*. New York: Oxford University Press, 2013.
- Cavallero, Eric. "Association and Asylum." *Philosophical Studies* 169, no. 1 (May 2014): 133–41.
- Cole, Phillip, and Christopher Heath Wellman. *Debating the Ethics of Immigration: Is There a Right to Exclude?* New York: Oxford University Press, 2011.
- Darwall, Stephen L. "Two Kinds of Respect." *Ethics* 88, no. 1 (October 1977): 36–49.
- Dummett, Michael. *On Immigration and Refugees*. London: Routledge, 2001.
- Dworkin, Ronald. *Taking Rights Seriously*. Cambridge, MA: Harvard University Press, 1977.
- Ferracioli, Luara. "Family Migration Schemes and Liberal Neutrality: A Dilemma." *Journal of Moral Philosophy* 13, no. 5 (2016): 553–75.
- Fine, Sarah. "Freedom of Association Is Not the Answer." *Ethics* 120, no. 2 (January 2010): 338–56.
- . "Immigration and Discrimination." In *Migration in Political Theory: The Ethics of Movement and Membership*, edited by Sarah Fine and Lea Ypi, 125–50. Oxford: Oxford University Press, 2016.

- Joppke, Christian. *Selecting by Origin: Ethnic Migration and the Liberal State*. Cambridge, MA: Harvard University Press, 2005.
- Kymlicka, Will. *Multicultural Citizenship: A Liberal Theory of Minority Rights*. Oxford: Oxford University Press, 1995.
- Lister, Matthew. "Immigration, Association, and the Family." *Law and Philosophy* 29, no. 6 (November 2010): 717–45.
- . "A Rawlsian Argument for Extending Family-Based Immigration Benefits to Same-Sex Couples." *University of Memphis Law Review* 37 (Summer 2007): 745–80.
- Miller, David. "Border Regimes and Human Rights." *Law and Ethics of Human Rights* 7, no. 1 (August 2013): 1–23.
- . "Immigrants, Nations, and Citizenship." *Journal of Political Philosophy* 16, no. 4 (December 2008): 371–90.
- Shachar, Ayelet. "Selecting by Merit: The Brave New World of Stratified Mobility." In *Migration in Political Theory: The Ethics of Movement and Membership*, edited by Sarah Fine and Lea Ypi, 175–201. Oxford: Oxford University Press, 2016.
- Sidgwick, Henry. *The Elements of Politics*, 2nd ed. London: MacMillan, 1897.
- Pevnick, Ryan. *Immigration and the Constraints of Justice: Between Open Borders and Absolute Sovereignty*. Cambridge: Cambridge University Press, 2011.
- Walzer, Michael. *Spheres of Justice: A Defense of Pluralism and Equality*. New York: Basic, 1983.
- . *Thinking Politically: Essays in Political Theory*. Edited by David Miller. New Haven: Yale University Press, 2007.
- Wellman, Christopher Heath. "Immigration and Freedom of Association." *Ethics* 119, no. 1 (October 2008): 123–32.
- Wilcox, Shelley. "Do Duties to Outsiders Entail Open Borders? A Reply to Wellman." *Philosophical Studies* 169, no. 1 (May 2014): 1–10.

## CAN WE INTEND THE PAST?

Oded Na'aman

IN THE BURGEONING LITERATURE on the rationality of regret, Jay Wallace's *The View from Here* has played an important role. It is a book full of provocative and gripping ideas. One such idea concerns the way in which our love and attachment may, and often do, implicate us in the wrongs and evils of the past. In particular, Wallace argues that our love and attachment involve something like a willingness to bring about the necessary conditions for the existence of their objects, even when those conditions involve wrongs and evils that we should not be willing to bring about. Therefore, the persons and things that are most important to us implicate us in their morally dubious genealogy.

Wallace calls the backward unfurling of intention due to one's present attachments the *affirmation dynamic*. Some have found the affirmation dynamic plausible and offered further defense of it.<sup>1</sup> Others have found plausible a qualified version of the affirmation dynamic.<sup>2</sup> I argue that the affirmation dynamic is impossible, at least as Wallace construes it. In particular, I argue that the idea that we may have intention-like attitudes about the past is fundamentally confused. It betrays a misguided conception of retrospection as involving a choice between courses of history. As an alternative to this *timeless conception of retrospection*, I briefly propose that retrospection is *diachronic*: it begins after the moment of choice or action has passed and unfolds over time and in accordance with its own standards of appropriateness. Contrary to Wallace's view, affirming and regretting the past involve no longer viewing it as a matter of choice.

### I

We may regret the past in various ways. However, according to Wallace, *all-in regret* involves an on-balance preference that the object of regret did not occur.<sup>3</sup> Wallace's counterpart to all-in regret is *on-balance affirmation*. On-balance affir-

1 Kolodny, "Dynamics of Affirmation."

2 Nagel, "An Invitation to Hand-Wringing"; Munoz-Dardé, "Puzzles of Regret."

3 Wallace, *The View from Here*, 55.

mation involves an on-balance preference that the object of affirmation did occur. In regretting or affirming (I henceforth drop the qualifications “all-in” and “on-balance”) we settle the question: “Would we, knowing what we now know about how things have since played out, bring it about that things were otherwise in the respect that we are focusing on, if it were in our power to do so?”<sup>4</sup> The attitudes of regret and affirmation are therefore like conditional intentions. They involve a commitment to doing something in certain conditions. These attitudes are only *like* conditional intention because the conditions they invoke will never obtain and therefore these attitudes will never lead to action.<sup>5</sup>

Affirmation of one object, according to Wallace, has a way of spreading to others. The people we love and the projects to which we are attached give us powerful reasons to affirm them.<sup>6</sup> The affirmation that is called for with regard to objects of love and attachment is especially demanding, for it involves being “glad on balance that those objects are part of the history of the world, taking into account the totality of things that they involved.”<sup>7</sup> Crucially, Wallace claims that this *unconditional affirmation* entails a commitment to affirming “the historical conditions that were necessary for the existence of the thing that one affirms.”<sup>8</sup> Such affirmation is incompatible with regret, and therefore our unconditional affirmation of the objects of our attachment precludes regretting any of the necessary conditions (whether causal or constitutive) for their existence. This is the affirmation dynamic.

## II

The affirmation dynamic leads to serious trouble. The causal lineage that made one’s objects of attachment possible is bound to involve some deeply regrettable events somewhere along the way. I, for one, can confidently say that neither my existence nor the existence of most of the people I love would have been possible if it were not for the occurrence of the Holocaust. Similar concerns arise with regard to many of the projects and activities we cherish. For instance, Wallace remarks that the readers of his book are likely to be attached to academic philosophy, itself a “bourgeois pursuit” made possible by serious past (and present) injustice. Once we consider the intricate courses of history that made the objects of our attachments possible, none of us remain unscathed by their sordid past.

4 Wallace, *The View from Here*, 62

5 Wallace, *The View from Here*, 55–57.

6 Wallace, *The View from Here*, 75.

7 Wallace, *The View from Here*, 75.

8 Wallace, *The View from Here*, 75.

We are committed to willing those evils and wrongs without which the objects of our attachments would not have existed.

Thomas Nagel accepts the affirmation dynamic but in light of its disturbing upshots claims it should be contained: "our affirmation of anything . . . is bounded by a statute of limitations on its reach into the past. We can take much about the world that we have not created, good and bad, as simply given, and limit our affirmations and regrets to what is downstream from that."<sup>9</sup> Niko Kolodny defends Wallace's unbounded affirmation dynamic by arguing against Nagel's proposed cutoff point.<sup>10</sup> Why does the affirmation of an object of attachment commit us to affirming *all* the necessary conditions for its existence? Kolodny argues this is due to the principle of instrumental rationality, according to which if one intends an end one is committed to intending the necessary means to it.

Let me briefly rehearse Kolodny's argument. As Wallace argues, and both Kolodny and Nagel accept, by affirming an object of our attachment we intend to bring it about if we could. If a past wrong was necessary for the existence of the object of our attachment, then if we could bring it about this wrong would be a necessary means for the existence of the object of our attachment. Given that we intend to bring about the object of our attachment, the principle of instrumental rationality commits us to intending to bring about the necessary means to it, including the past wrong. It does not matter where on the axis of time the past wrong lies; what matters for our commitment to affirming it is only that it was necessary for our end, that is, for the existence of the object of our attachment. Thus, instrumental rationality leads us from affirming the object of our attachment to affirming *all* of its necessary conditions.<sup>11</sup>

### III

Kolodny's argument is valid, but it relies on a premise that both Nagel and Kolodny accept without question: that regret and affirmation are intention-like attitudes. If this premise is false, and regret and affirmation should not be construed as intention-like attitudes, then it is also false that by failing to affirm the necessary conditions for the existence of one's object of affirmation one fails to intend the necessary means for one's end. The affirmation dynamic would be blocked right from the start. The principle of instrumental rationality does not plausibly apply to wishes, for example, because they do not relate to action as intentions do.

9 Nagel, "An Invitation to Hand-Wringing," 25–26.

10 Kolodny, "Dynamics of Affirmation."

11 Kolodny, "Dynamics of Affirmation," 772.

One crucial element of intention, conditional or not, is that from the agent's point of view it *might* lead to action. In other words, one cannot intend what one takes to be impossible.<sup>12</sup> If I went to sleep early last night then I cannot go to sleep early last night *again*; if I did not go to sleep early last night then I can *never* do so. But if I *believe* that it is impossible for me to go to sleep early last night, then it is impossible for me to *intend* to go to sleep early last night, even though it is perfectly possible for me to *regret* not doing so or to be *glad* I did.<sup>13</sup> Regret and affirmation cannot be like intending a past occurrence since they are possible when the latter is not.

Wallace, however, does not suggest that in regretting or affirming a past occurrence we actually intend to change or repeat that very same occurrence. Rather, he argues that in regretting or affirming a past occurrence we intend to change or repeat it *if we could*. Wallace may therefore grant that if one believes [ $\phi$ -ing in  $c$ ] is impossible then it is impossible that one intends [to  $\phi$  in  $c$ ]. And yet Wallace may still insist that when one believes [ $\phi$ -ing in  $c$ ] is impossible, one may consistently believe that [ $\phi$ -ing if  $c$ ] is possible and therefore intend [to  $\phi$  if  $c$ ]. The reasoning might be the following: if  $c$  is impossible, then it will never be the case that  $c$  obtains and I do not  $\phi$ , so my success at [ $\phi$ -ing if  $c$ ] is guaranteed. The impossibility of the antecedent guarantees the satisfaction of the intended conditional.

This maneuver, however, is too successful. Since [ $\phi$ -ing if  $c$ ] is trivially satisfied due to the impossibility of  $c$ , there is nothing I need to do in order to [ $\phi$  if  $c$ ]. In particular, I do not ever need to intend to  $\phi$ . My conditional intentions are empty when I take them to be trivially realized. Consider my intention to run in the street naked if  $2 + 2 = 5$ . Since I believe the antecedent impossible, I may intend *never* to run in the street naked and still coherently intend to do so if  $2 + 2 = 5$ . Whether because I cannot intend what I take to be impossible or because I believe the condition of my intention is impossible, I do not intend anything at all. As I said, it is crucial for intention that from the agent's point of view it *might* lead to action.<sup>14</sup>

By contrast, it is precisely the impossibility of action that makes room for regret. It is distinctive of regret that it is directed toward settled facts. Even if we can act to mitigate the negative consequences of a regrettable choice or compen-

12 Davidson, "Intending," 100–1. For the purposes of this discussion, I use "believe impossible" and "take to be impossible" interchangeably.

13 Note that I might falsely believe it possible for me to go to sleep early last night, in which case it would be possible for me to intend to do so. We cannot do what is impossible, but we can intend it as long as we believe it possible.

14 Hills ("Hindsight," 11–12) makes a similar point.

sate those harmed by it, our regret is intelligible only on the assumption that we take the choice itself to be irreversible. We can *wish* we did not make it, or *wish* we could go back and choose differently, but we can intend neither. To say that affirmation and regret are different from conditional intentions *only* in that their conditions are taken by the agent to be impossible is to say that affirmation and regret are *nothing like* conditional intentions.

## IV

I said earlier that many of the people I love would not have existed were it not for the occurrence of the Holocaust. Clearly, it would be insane of me to intend the occurrence of the Holocaust for the sake of the individuals whose existence was made possible by it. But why would it be insane? Wallace and Kolodny are impressed by the fact that I cannot justify causing so much evil and suffering for the sake of my few friends and family members. But to stop here would be to overlook a further aspect of my would-be insanity, namely, that by intending to bring about the Holocaust for the sake of the existence of my loved ones I would be treating the occurrence of the Holocaust as *unsettled*. That is, by intending to bring about the Holocaust I would commit myself to the possibility of *actually* bringing about the Holocaust. I would thereby fail to appreciate the fact that the Holocaust is an event that lies in the past and whose occurrence I can do absolutely nothing about. A willingness to bring about *an event like the Holocaust* is morally insane; a willingness to bring about *the Holocaust* is plain insane.

The idea that we can have intentions about the past can seem plausible if we conceive of retrospection as a choice between courses of history. Such a choice would aspire to take into account the totality of occurrences in each possible chain of events to determine which course of history should be chosen on balance.<sup>15</sup> On this *timeless conception of retrospection*, as we might call it, our regret and affirmation reflect where we happen to be in history, but they are justified by a more fundamental choice that is not itself expressive of any course of history. Our location in time should not impact our evaluation of each course of history taken as a whole. Intending a past occurrence does not seem insane since the intention is expressive of one's endorsement of an entire course of history considered from an atemporal point of view. In light of this model of retrospection, our commitment to people with whom we happen to coexist at the expense of those who existed before us seems difficult to justify.<sup>16</sup>

15 Again, Hills makes a similar point ("Hindsight," 19).

16 Setiya ("The Ethics of Existence") claims that we should prefer the existence of those who coexist with us over their nonexistence. However, his claim seems difficult to square with

Wallace is not alone in implicitly assuming such a timeless conception of retrospection. Kieran Setiya recently discussed retrospection as a preference between “world-histories.”<sup>17</sup> And John Rawls relies on this model of retrospection for his view of rational regret. According to Rawls, regret about one’s choices is rational when the choices were not part of one’s rational life plan, understood as the plan “that would be decided upon as the outcome of careful reflection in which the agent reviews, in the light of all the relevant facts, what it would be like to carry out these plans and thereby ascertained the course of action that would best realize his more fundamental desires.”<sup>18</sup> According to Rawls, in making a particular choice a rational person endorses a life plan of which the choice is a component; correspondingly, in regretting a particular choice a rational person disowns a life plan of which the choice is a component. In short, for Rawls, every rational choice is a choice of a whole life plan, which spans backward and forward in time. Therefore, in intending a specific choice a rational agent intends its rational future as well as its rational past.

## v

A focus on the choice- and action-guiding roles of practical reason might have led philosophers to endorse a timeless conception of retrospection: a conception of retrospection as a choice of history. But retrospection should not be construed as a futile attempt to undo or redo history as viewed from a point of view outside of it. Rather, retrospection occurs when the past is already settled and the space for intention is closed; it occurs *in* time and *over* time and may involve various complicated emotions and thoughts.

For example, a diachronic model of retrospection differs from a timeless model of retrospection in making room for non-comparative preferences or wishes. In a choice situation, preferences are naturally interpreted as comparative: a preference against one option tells in favor of the alternative, and vice versa. Accordingly, if we model retrospection on a choice between courses of history, then a preference against one course of history tells in favor of another. That this model is mistaken is demonstrated by the fact that we often have rational non-comparative preferences about the past. When I order the burger but

---

his timeless conception of retrospection, in which our attitudes toward the past are determined by our overall evaluations of possible courses of history. Setiya’s claim about preferences regarding coexists may therefore seem like an *ad hoc* exception made to accommodate intuitions that tell against his timeless conception of retrospection.

17 Setiya, “The Ethics of Existence,” 294, and “Retrospection,” 10.

18 Rawls, *A Theory of Justice*, 415–16.

later wish I had the schnitzel instead, I have a comparative counterfactual wish, but if the burger gave me food poisoning I might simply wish I did not have it without having any replacement in mind, in which case I have a non-comparative counterfactual wish. Preferring or wishing that things were otherwise does not rationally commit us to preferring the most likely alternative, or any specific alternative at all for that matter.

Upon reading the memoirs of Jean Améry, Primo Levi, or Robert Antelme, I am horrified by the events of the Holocaust and wish with all my heart that they did not occur. But I do not thereby commit myself to preferring any specific alternative history, and I certainly do not commit myself to preferring a history in which my loved ones did not exist. Appropriate retrospection might preclude a consistent view of our preferred course of history. This is an important lesson from Wallace's affirmation dynamic: there is probably no causally possible course of history that would satisfy all our most fervent retrospective preferences—no one course of history that we favor over all others given the totality of things it includes. But this is as it should be given that history is not, for us, a matter of choice.<sup>19</sup>

Stanford University  
onaaman@stanford.edu

#### REFERENCES

- Davidson, Donald. "Intending." In *Essays on Actions and Events*, 83–102. Oxford: Oxford University Press, 1980.
- Hills, David. "Hindsight." Unpublished manuscript.
- Kolodny, Niko. "Dynamics of Affirmation." *Philosophy and Phenomenological Research* 92, no. 3 (May 2016): 771–77.
- Munoz-Dardé, Véronique. "Puzzles of Regret." *Philosophy and Phenomenological Research* 92, no. 3. (May 2016): 778–84.
- Nagel, Thomas. "An Invitation to Hand-Wringing." *London Review of Books* 36, no. 7 (April 3, 2014): 25–26.
- Rawls, John. *A Theory of Justice*. Cambridge, MA: Harvard University Press, 1971.
- Setiya, Kieran. "The Ethics of Existence." *Philosophical Perspectives* 28, no. 1 (December 2014): 291–301.
- . "Retrospection." *Philosophers' Imprint* 16, no. 15 (August 2016): 1–15.

19 I wish to thank Johann Frick and Jorah Dannenberg for their valuable comments on previous drafts of this paper.

Wallace, R. Jay. *The View from Here: On Affirmation, Attachment, and the Limits of Regret*. Oxford: Oxford University Press, 2013.

## HYBRID NON-NATURALISM DOES NOT MEET THE SUPERVENIENCE CHALLENGE

David Faraci

IT IS WIDELY AGREED that normative properties supervene on natural properties. Non-naturalists face a distinctive challenge to explain this relation. Unlike other metanormative contenders, non-naturalists take normative supervenience to be a relation between *metaphysically discontinuous* kinds: natural properties and *sui generis* normative properties. As Tristram McPherson forcefully argues, that discontinuity makes it difficult, if not impossible, for non-naturalists to explain supervenience.<sup>1</sup> And it is widely accepted that an inability to explain a necessary relation between distinct kinds—at least insofar as that relation *demand*s explanation—is a significant theoretical cost.

Stephanie Leary argues that non-naturalists can meet this explanatory demand by positing the existence of *hybrid* normative properties.<sup>2</sup> These properties serve as a kind of “double-sided tape,” allowing a natural property to ground a *sui generis* normative property (e.g., goodness) without violating non-naturalism’s commitment to metaphysical discontinuity—i.e., without that grounding relation’s holding in virtue of the nature of either the natural properties or the *sui generis* normative properties themselves.<sup>3</sup> Each hybrid normative property *H* has two key features: (i) it is part of the (constitutive immediate) essence of *H* that some natural property *G* grounds *H*’s instantiation and (ii) it is part of the essence of *H* that *H* grounds the instantiation of some *sui generis* normative property *F*.<sup>4</sup>

- 1 McPherson, “Ethical Non-Naturalism and the Metaphysics of Supervenience.”
- 2 Leary, “Non-Naturalism and Normative Necessities.” Many of the passages from Leary herein are offered in response to McPherson’s charge that proposals like hers fall prey to “bruteness revenge,” that they offer “an explanation of one necessary connection only by covertly relying on a second brute necessary connection” (McPherson, “Ethical Non-Naturalism and the Metaphysics of Supervenience,” 222–23). My arguments can be seen as clarification and extension of McPherson’s objection.
- 3 Leary, “Non-Naturalism and Normative Necessities,” 99.
- 4 Leary follows Fine (“Senses of Essence”) in distinguishing *constitutive immediate* essences from those that are *consequential* or *mediate*. Except where I rely on these distinctions explicitly, all mentions of “essence” should be taken to refer to constitutive immediate essence.

For example, one might claim that *being in pain* is such a property: that it's part of the essence of being in pain that (a) if one's C-fibers are firing, then one is in pain, and (b) that if  $x$  is a painful experience,  $x$  is bad.<sup>5</sup>

Unfortunately, this proposal does not meet the supervenience challenge. To see this, begin with Leary's formulation of supervenience:<sup>6</sup>

*Strong Supervenience:*

$$(\forall F \text{ in } A)(\forall x)[Fx \rightarrow (\exists G \text{ in } B)(Gx \ \& \ \Box_M(\forall y)(Gy \rightarrow Fy))]$$

$A$  is the class of normative properties and  $B$  is the class of natural properties. Thus, according to *Strong Supervenience*, for every normative property  $F$ , each instantiation of  $F$  is metaphysically necessitated by the instantiation of some natural property  $G$ . With  $A_{SG}$  as the class of *sui generis* normative properties, Leary's proposal entails:

*Hybrid Property:*

$$(\forall F \text{ in } A_{SG})(\exists H)(\exists G \text{ in } B)[\Box_M(Gx \rightarrow Hx) \ \& \ \Box_M(Hx \rightarrow Fx)]$$

*Hybrid Property* does not entail *Strong Supervenience*. It entails that  $G$  is sufficient for  $F$ . It does *not* entail that the instantiation of one or more members of  $B$  (i.e., of at least one natural property) is necessary for the instantiation of  $F$ , and thus does not entail that  $F$  supervenes on the set of natural properties. The latter, Leary claims, is explained by her essentialist metaphysics:

The background essentialist metaphysics explains why it's metaphysically necessary that, if  $x$  has some *sui generis* normative property,  $x$  has some natural property that is involved in the essence of whatever hybrid properties ground it. On the essentialist framework, no grounding facts are fundamental—they are all grounded in essences. So, no derivative property  $F$  can be instantiated by  $x$  unless  $x$  has some more fundamental property  $G$  and there is an essential connection between *being F* and *being G*.<sup>7</sup>

Leary makes two key claims about  $F$  in this passage. First, she claims that  $F$  is a "derivative property"—i.e., is always grounded. Second, she claims that  $F$  is always grounded *in some natural property*. But it is important to see that neither *Hybrid Property* nor essentialist metaphysics entails these claims. Even if  $H$  grounds  $F$  in

5 Leary, "Non-Naturalism and Normative Necessities," 98.

6 Leary, "Non-Naturalism and Normative Necessities," 80, adapted from Dreier, "The Supervenience Argument Against Moral Realism."

7 Leary, "Non-Naturalism and Normative Necessities," 102.

one case, *F* might be ungrounded in another.<sup>8</sup> And even if *F* is always grounded, it does not follow that it is always grounded in *some natural property*.

By contrast, in many other cases essentialist metaphysics *does* guarantee necessary conditions for property instantiation. Because it is in the essence of *H* to be grounded in *G*, *G* is necessary (and arguably sufficient) for *H*, and thus we know both that *H* is derivative and that it is always grounded in a natural property. But this is precisely because the relevant grounding relation holds in virtue of the essence of the grounded property. Where the grounding relation holds in virtue of the grounding property, as in the relation between *H* and *F*, it follows only that the base property is sufficient.

Why, then, does Leary claim that, given her essentialist metaphysics, her proposal entails that instantiation of at least one natural property is necessary for instantiation of any given *sui generis* normative property? One possibility is that she is making the mistake of thinking that all essence facts entail some relevant necessity. But a more charitable interpretation is available. Consider the following passage:

Facts about what is essential of what are *autonomous*: they are brute in the sense that they are simply not the sorts of facts that can, in principle, have a metaphysical explanation. So, taking it to be brute that certain hybrid properties exist and others don't is not problematic. The question of what metaphysically explains such facts does not legitimately arise.<sup>9</sup>

This recommends the following reading. The non-naturalist needs to show that there is a possible metaphysical explanation for natural properties' necessitating *sui generis* normative properties. Hybrid properties provide such an explanation. Now suppose there are hybrid properties with one sort of essence: they "tape" *sui generis* normative properties to natural properties. Suppose also that *sui generis* normative properties are *always* grounded and that there are no properties that "tape" *sui generis* normative properties to themselves or to other non-natural properties. Given those suppositions, *Strong Supervenience* holds as a universal generalization—indeed, a *necessary* one, given that it is a generalization from essences. And since essence facts do not admit of explanation, the explanatory buck stops here.

This reading is more charitable, but the argument fails. If we grant the autonomy of essence, it follows that we cannot raise questions about facts of the

8 Leary might argue that if "*x* is *F*" is grounded, then for all *y*, "*y* is *F*" is grounded. But there is little reason to think this when the grounding relation holds in virtue of the essence of the grounding property.

9 Leary, "Non-Naturalism and Normative Necessities," 102.

form “Y is part of the essence of H.” But it does not follow that we cannot raise questions about facts of the form “There are no properties with essence Z.” Crucially, it is a fact of this latter sort we are concerned with, not the former. The non-naturalist is challenged to explain why it is impossible for normative properties to be ungrounded or fully grounded in other non-natural properties. With respect to Leary’s proposal, this prompts the questions: (i) why is it impossible for there to be *sui generis* normative properties that are ungrounded, and (ii) why is it impossible for there to be properties whose essence it is to ground *sui generis* normative properties, but which are themselves ungrounded or fully grounded in other non-natural properties? These questions remain unanswered.

Here, Leary might double down on the above reading, insisting that the autonomy of essence entails that facts about which essences are impossible are also brute. But this is not the case. Consider:

*Prime*: No number has the essence: (i) is greater than 2; (ii) is prime; (iii) is even.

Suppose someone were to claim that any demand to explain *Prime* is both sufficiently met and can only *be* met by listing the prime numbers and noting that none of those greater than 2 are even. This would be doubly mistaken. This response does not meet the explanatory demand in question, for the fact that the prime numbers greater than 2 are all odd is precisely what we are being called to explain. Moreover, an explanation is available: prime numbers are divisible exactly by themselves and 1, but all even numbers greater than 2 are divisible by at least themselves, 1, and 2.

To better understand how this explanation works, it will be useful to mark the distinction between *constitutive* and *consequential* essences. Q is part of the *constitutive* essence of W if Q is essential to W but is “not had in virtue of being a consequence of some more basic essential properties” of W.<sup>10</sup> R is part of the *consequential* essence of W if R is essential to but not constitutive of W. *Prime* claims that a certain constitutive essence is impossible. This turns out to be true because the negation of one of the listed features, evenness, is entailed by the essences of the other two, primeness and being greater than 2—i.e., there are no even prime numbers greater than 2 because oddness is part of the consequential essence of *being a prime number greater than 2*.

In the case of *Prime*, all of this is a matter of logical necessity. In other cases, the explanation is metaphysical. Consider:

10 Fine, “Senses of Essence,” 57.

*Color*: No object has the essence: (i) is red all over; (ii) is green all over.

Like *Prime*, *Color* is a claim about the impossibility of a certain constitutive essence that is true in virtue of the consequential essences of its parts: colors are metaphysically incompatible, such that red metaphysically necessitates not-green and vice versa. Crucially, these explanations do not violate the autonomy of constitutive essence; no explanation is offered for the essences of primeness, oddness, redness, or greenness.

These cases demonstrate that the *impossibility* of at least some constitutive essences can be explained, either logically or metaphysically, without violating the autonomy of constitutive essence. By contrast, Leary asks us to accept that the following neither can be explained nor demands explanation:

*Natural Grounds*: No property has the essence: (i) grounds or is identical to a *sui generis* normative property; (ii) is not grounded in any natural property.<sup>11</sup>

We should reject her request. First, we have good reason to think that *Natural Grounds* in particular is explicable. Return to *Prime* and *Color*. These can be stated as universal generalizations. But our *knowledge* of them does not proceed via generalization; we do not need to know what the prime numbers are, or what the red and green objects are, to know that *Prime* and *Color* are true. The best explanation for this ability is that we (implicitly or explicitly) recognize that they are true in virtue of the nature of the properties they concern. Given this, even if the explanation offered above were not readily apparent, we would have good reason to suspect that such a further explanation is available, that *Prime* and *Color* are not merely true in virtue of which particular numbers are prime and which particular objects are red and green.

The same is true of *Natural Grounds*. We do not know *Natural Grounds* is true by generalization from our knowledge of particular supervenience relations (and definitely not by knowing which hybrid properties exist!). The supervenience of the normative on the natural is a point of much broader agreement than which things are (e.g.) good and bad. And most of us are confident that supervenience holds regardless of our level of uncertainty about which things are good and bad. Again, this seems best explained by our (implicit or explicit) recognition that *Natural Grounds* is true in virtue of the nature of the related properties themselves, of the natural and *sui generis* normative properties. Indeed, this is precisely the sort of explanation that other views in normative meta-

11 The “is identical to” disjunct accommodates the possibility that some *sui generis* normative properties are ungrounded.

physics provide—e.g., for some naturalists, that natural properties feature in the essences of normative properties.<sup>12</sup>

What is more, we have good reason to think that *all* claims of the impossibility of constitutive essences are explicable. This is because it is eminently plausible that the possible elements of essences are infinitely re-combinable, *except* where those elements logically or metaphysically exclude one another. Consider an analogy with concepts—a particularly apt analogy, given that essential necessity is often taken to be analogous to analyticity. In general, we expect conceptual elements to be similarly re-combinable. We accept the impossibility of the concept UNMARRIED MARRIED PERSON because its elements are clearly inconsistent. But suppose some theory holds that the concept UNMARRIED TALL PERSON is impossible. This demands explanation. And it should be clear that the theory cannot sidestep this demand by claiming that conceptual content is autonomous.

This last point shows us why hybrid non-naturalism not only fails to *meet* the supervenience challenge, but should not even be taken to have made progress with respect to it.<sup>13</sup> True, hybrid non-naturalism explains how natural properties could necessitate *sui generis* normative ones. But this leaves *Natural Grounds* untouched. I have suggested that, if *Natural Grounds* is true, it must be so because the combination it denies—grounding or being a *sui generis* normative property, but not being grounded in any natural property—is ruled out by some logical or metaphysical relation between those elements. But precisely what motivated the hybrid view was acceptance of critics' claim that non-naturalism is incompatible with the existence of such logical or metaphysical relations between natural and *sui generis* normative properties.

My personal inclination is to think that this is a problem with non-naturalism itself, in which case hybrid non-naturalism's limited victory here is Pyrrhic, since the overarching challenge cannot be met. But if, instead, the problem is essen-

12 Though for important concerns about this explanation, see Mitchell, "Mixed Up about Mixed Worlds?"

13 See Elliott ("How to Make Progress against the Supervenience Challenge") for arguments that Leary *does* make progress. As will be clear from what follows, I agree with him that she makes progress in one sense: we can separate out *Strong Supervenience* into necessary and sufficient conditions, and Leary's account explains the sufficient condition. But I see this as a Pyrrhic victory, at best, since her proposal seems to make explaining the necessary condition impossible.

tialist metaphysics, hybrid non-naturalism makes progress only inasmuch as it takes one step forward, two steps back.<sup>14</sup>

Georgetown University  
 david.faraci@georgetown.edu

#### REFERENCES

- Dreier, James. "The Supervenience Argument against Moral Realism." *Southern Journal of Philosophy* 30, no. 3 (Fall 1992): 13–38.
- Elliott, Aaron. "How to Make Progress against the Supervenience Challenge." Unpublished manuscript.
- Fine, Kit. "Senses of Essence." In *Modality, Morality, and Belief: Essays in Honor of Ruth Barcan Marcus*, edited by Walter Sinnott-Armstrong, Diana Raffman, and Nicholas Asher, 53–73. Cambridge: Cambridge University Press, 1994.
- Leary, Stephanie. "Non-Naturalism and Normative Necessities." In *Oxford Studies in Metaethics*, vol. 12, edited by Russ Shafer-Landau, 76–105. Oxford: Oxford University Press, 2017.
- McPherson, Tristram. "Ethical Non-Naturalism and the Metaphysics of Supervenience." In *Oxford Studies in Metaethics*, vol. 7, edited by Russ Shafer-Landau, 205–34. Oxford: Oxford University Press, 2012.
- Mitchell, Cole. "Mixed Up about Mixed Worlds? Understanding Blackburn's Supervenience Argument." *Philosophical Studies* 174, no. 12 (December 2017): 2903–25.

<sup>14</sup> Thanks to Aaron Elliott, Marc Lange, and Tristram McPherson for helpful discussion; to participants at the 2017 Junior Metaphysics Workshop at Virginia Tech; and to the editors and an anonymous reviewer at the *Journal of Ethics and Social Philosophy*.

## IS AGENT-NEUTRAL DEONTOLOGY POSSIBLE?

*Matthew Hammerton*

IT IS COMMONLY HELD that all deontological moral theories are agent-relative in the sense that they give each agent a special concern that *she* does not perform acts of a certain type rather than a general concern with the actions of all agents. Recently, Tom Dougherty has challenged this orthodoxy by arguing that agent-neutral deontology is possible.<sup>1</sup> His argument is simple: he posits a moral rule that he claims is both agent-neutral and deontological. In this article I show that the rule Dougherty posits cannot be both agent-neutral and deontological. The problem is that the rule has several possible interpretations and, although on some interpretations it is an agent-neutral rule, and on some interpretations it is a deontological rule, there are no interpretations in which it is both agent-neutral and deontological. I conclude by considering an alternative rule inspired by Dougherty's approach that looks like it might be both agent-neutral and deontological and showing that it too fails because it has core commitments that are incompatible with deontology.

### 1. DOUGHERTY'S ARGUMENT

Dougherty's argument starts from the assumption that deontological moral theories necessarily contain deontic constraints. Deontic constraints are moral rules that prohibit agents from performing acts of a certain type even if doing so is the only way to prevent more acts of that type from being performed by others. Thus, a deontic constraint on killing innocent people prohibits an agent from killing an innocent person even if doing so is the only way to prevent more killings of innocent people by others.

Deontic constraints have typically been thought to be agent-relative moral rules. For example, a deontic constraint on killing might be stated as follows:

KR: Each agent should ensure that she does not kill innocent people.

KR is an agent-relative rule because it gives each agent a special concern with acts

<sup>1</sup> Dougherty, "Agent-Neutral Deontology."

of killing that she might perform, requiring her not to kill an innocent person even if doing so can prevent more killings by others. Following Parfit we can say that KR gives different aims to different agents.<sup>2</sup> For example, it gives me the aim that *I* do not kill and you the aim that *you* do not kill.

Dougherty suggests that although a deontic constraint on killing is normally formulated as an agent-relative rule like KR there is an agent-neutral alternative that could serve deontologists equally well:

DR: Each agent should ensure that no one kills to prevent more killings by others.<sup>3</sup>

Let us call the act of *killing to prevent more killing by others* “preventive killing.” DR appears to be an agent-neutral rule because it gives all agents a general concern with preventive killings regardless of who they are performed by. Following Parfit we can say that it gives the same ultimate aim to all agents—the aim that there are no preventive killings.<sup>4</sup> DR also appears to produce a deontic constraint on killing because it prohibits each agent from killing an innocent person, even when doing so is the only way to prevent more killings by others. On these grounds Dougherty argues that a deontological theory could contain only agent-neutral rules like DR and thus that deontology can be agent-neutral.

Though this argument may appear plausible, I believe it is too fast. The problem is that DR can be given several different readings. To prove that agent-neutral deontology is possible there must be at least one reading of DR in which it is both agent-neutral and a deontic constraint. Let us consider each of the readings in turn.

First, consider a literal reading of DR. Read literally, DR requires each agent to make it the case that no one ever commits a preventive killing, and classes all agents as having violated DR and acted wrongly insofar as some agent, somewhere, kills preventively. To make this literal reading clearer, I will replace the word “ensure” with the word “guarantee”:

*Guarantee*: Each agent should guarantee that no one kills to prevent more killings by others.

*Guarantee* is an extremely demanding moral rule and, on these grounds alone, is implausible. Furthermore, *Guarantee* violates the *ought-implies-can* principle as it requires agents to ensure that there are no preventive killings even when they lack the ability and opportunity to do so. Many take the *ought-implies-can*

2 Parfit, *Reasons and Persons*, 27.

3 Dougherty, “Agent-Neutral Deontology,” 531.

4 Parfit, *Reasons and Persons*, 27.

principle to be a conceptual truth, which would entail that *Guarantee* is not only implausible but also incoherent.<sup>5</sup>

These problems with *Guarantee* suggest that DR is better interpreted as containing a clause limiting it to cases in which an agent has the ability to satisfy it:

*Ability*: Each agent, whenever she has the ability and the opportunity to do so, should ensure that no one kills to prevent more killings by others.<sup>6</sup>

*Ability* does not place unrealistic and excessive demands on agents and thus is much more plausible than *Guarantee*. However, *Ability* faces another problem. Consider the following case:

*Prevention*: Tom and Mary are each about to commit a preventive killing. The only way for you to stop them is by committing one preventive killing yourself (perhaps the horror of witnessing a preventive killing will cause them to abandon their plans).

In *Prevention* you do not have the ability to ensure that no one commits a preventive killing. Thus, *Ability* does not class you as doing the morally wrong thing when you fail to stop all preventive killings. However, you do face a morally significant choice in *Prevention* between allowing two preventive killings to occur or committing one preventive killing yourself. *Ability* gives you no advice on what to do here. Yet a deontic constraint on killing must give you advice in this situation, requiring you not to perform a preventive killing even when doing so is the only way to prevent more preventive killings by others (as KR clearly does). So, although *Ability* appears to be agent-neutral, it is not a deontic constraint and thus fails to satisfy a necessary condition of deontology.

To address this problem we might further modify DR:

*Oppose*: Each agent should, to the best of her ability, ensure that no one performs a preventive killing, making sure that whenever she cannot prevent all such killings, she at least does not perform any herself.

- 5 See Vranas, "I Ought, Therefore I Can," 167–69, for a defense of the "conceptual necessity" interpretation of the *ought-implies-can* principle (OIC). It should also be noted that those who reject OIC usually agree that the principle holds true for much of our moral theorizing but maintain that it fails in some special cases. However, *Guarantee* violates OIC in regular cases and not just in the special cases in which some argue that the principle fails.
- 6 Among the reasons why Dougherty might want to accept *Ability* is the fact that he borrows the term "ensure" from McNaughton and Rawling and apparently intends to use it in the same way that they use it. However, McNaughton and Rawling ("Value and Agent-Relative Reasons," 34) clarify that they interpret ensuring rules as containing a clause similar to that that appears in *Ability*.

Unlike *Ability*, *Oppose* does qualify as a deontic constraint on killing. For *Oppose* requires of each agent that she never kills, and requires this even if she is in a situation like *Prevention* in which, by killing, she can prevent more preventive killings overall. However, *Oppose* is not an agent-neutral rule. For the last clause in *Oppose* requires each agent to give special priority to her own killings, giving her the aim that she does not kill in cases like *Prevention*, over and above any general aim that no one kills. Thus, *Oppose* also fails as an example of agent-neutral deontology.

Finally, let us consider an alternative way of dealing with *Prevention* that preserves the agent-neutrality of DR:

*Minimize*: Each agent should, to the best of her ability, ensure that no one performs a preventive killing, minimizing the total number of such killings whenever she cannot prevent them all.

*Minimize* appears to be agent-neutral as it gives all agents the same ultimate aim—the aim that there are as few preventive killings as possible. However, *Minimize* is not a deontic constraint as, in cases like *Prevention*, it does not, as a deontic constraint must, prohibit you from killing. Thus, *Minimize* also fails as an example of agent-neutral deontology.

I cannot see any other interpretations of DR that have the potential to give us a rule that is both a deontic constraint and agent-neutral. Thus, on the basis of the arguments above, I conclude that Dougherty's rule DR fails to demonstrate that deontology can be agent-neutral.

## 2. AN ALTERNATIVE RULE

The main problem with DR was that it cannot be an agent-neutral rule and yet also give a deontological verdict in cases like *Prevention*. Dougherty might try to fix this problem by postulating an additional agent-neutral rule that requires agents to ensure to the best of their ability that no one performs a second-order preventive killing (i.e., a preventive killing that itself prevents other preventive killings). However, this rule would be subject to a further counterexample, a case involving third-order preventive killings. In fact, for any additional rule concerning  $n$ -order preventive killing there is a *Prevention*-like counterexample involving  $n + 1$ -order preventive killings.

One way to escape this infinite sequence of counterexamples is to construct a rule requiring agents to prioritize stopping any higher-order killings over stopping any lower-order killings it prevents. To make this idea clearer let us say that when the absence of a higher-order killing is part of the causal explanation of why a lower-order killing occurs, then the higher-order killing is "causally up-

stream” of the lower-order killing. With this idea we can formulate the following rule:

NR: Each agent should, to the best of her ability, ensure that outcome O<sub>1</sub> occurs rather than outcome O<sub>2</sub> whenever O<sub>2</sub> contains a killing that is causally upstream of all the killings in O<sub>1</sub>.<sup>7</sup>

NR appears to be an agent-neutral rule as it gives all agents the same aim—that causally upstream killings are not performed. It also appears to give the same deontic verdicts as the agent-relative KR as an agent will always be prohibited from performing a preventive killing of any order because it will be causally upstream of any killings that will occur if she does not perform a preventive killing. Indeed, even killing to prevent several agents from violating NR is prohibited by NR because such a killing is causally upstream of the killings committed by those who are violating NR. Thus, NR appears to avoid the problems faced by DR.

Nonetheless, I think NR fails to capture an essential feature of deontology that agent-relative rules like KR are able to capture. To see this we need to consider the following case. Suppose that you and Ethel are both enemies of Joe and that each of you wants him dead. Furthermore, suppose you know that if Ethel has the opportunity to kill Joe she will and that, given your weak will, if you have the opportunity to kill Joe you will do so as well. Finally, suppose you know that either you or Ethel will eventually get an opportunity to kill Joe. Thus, an outcome in which neither of you kills him is not going to obtain. It follows that one of the following two outcomes will result:

O<sub>1</sub>: You kill Joe.

O<sub>2</sub>: Ethel kills Joe.

Importantly, neither of the killings in these two outcomes is causally upstream of the other. If O<sub>1</sub> obtains, the correct causal explanation will appeal to the fact that you had the opportunity to kill Joe, the temptation to kill Joe, and a weak will unable to resist temptation. If O<sub>2</sub> obtains the correct causal explanation will appeal to the fact that Ethel had the opportunity to kill Joe, the desire to kill Joe, and a morally corrupted will. In neither case will one of you kill Joe in order to prevent the other from killing him.

Now imagine in this scenario that you have the options of either ensuring that you do not kill Joe or ensuring that Ethel does not kill Joe but not the option of ensuring that neither of you kills Joe. For example, perhaps you and Ethel share the use of a car that tomorrow either of you might use to reach Joe and kill him. Tonight you must decide where to put the car keys. If you put them in one

<sup>7</sup> This rule is adapted from a suggestion made by Tom Dougherty in private correspondence.

spot then Ethel will grab them first tomorrow and use the car to kill Joe. If you put them in the only other possible spot then you will have first access to them tomorrow and, given your weak will, you know that you will give in to temptation and take the car, using it to kill Joe. Thus, your choice of where to put the car keys is also a choice about whether to ensure that you do not kill Joe or ensure that Ethel does not kill Joe.

What should you do in such circumstances? I think it is clear what a deontological approach to killing must say here. It must say that in such circumstances your priority lies first and foremost with yourself. You must ensure that you do not kill innocent people. Maybe you are also required to ensure that others do not kill innocent people. However, if you have to choose between ensuring that you do not do this and ensuring that someone else does not do it you must (all else being equal) choose the former.

By requiring each agent to ensure that she does not kill, the agent-relative KR secures the correct deontological verdict in this case. However, NR is not able to secure this verdict. This is because neither O1 nor O2 is causally upstream of the other. Thus NR does not require the agent to ensure that O2 occurs rather than O1 (or vice versa). Instead, it gives the agent no guidance in this case. It appears then that agent-neutral rules like DR and NR are insufficient to produce moral theories that give the correct deontological verdicts across various cases involving deontic constraints. Only agent-relative rules like KR can produce those verdicts and thus the orthodox view that deontology is necessarily agent-relative is preserved.<sup>8</sup>

Singapore Management University  
mhammerton@smu.edu.sg

#### REFERENCES

- Dougherty, Tom. "Agent-Neutral Deontology." *Philosophical Studies* 163, no. 2 (March 2013): 527–37.
- McNaughton, David, and Piers Rawling. "Value and Agent-Relative Reasons." *Utilitas* 7, no. 1 (May 1995): 31–47.
- Parfit, Derek. *Reasons and Persons*. Oxford: Clarendon Press, 1984.
- Vranas, Peter. "I Ought, Therefore I Can." *Philosophical Studies* 136, no. 2 (November 2007): 167–216.

8 I would like to thank Ryan Cox, Tom Dougherty, Daniel Nolan, and Nic Southwood for helpful comments on an earlier draft of this article.



JOURNAL of ETHICS & SOCIAL PHILOSOPHY  
<http://www.jesp.org>  
ISSN 1559-3061

The *Journal of Ethics and Social Philosophy* (JESP) 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. Articles are typically published under the CREATIVE COMMONS ATTRIBUTION-NONCOMMERCIAL-NODERIVATIVES 4.0 license, though authors can request a different Creative Commons license if one is required for funding purposes.



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.