

JOURNAL *of* ETHICS  
& SOCIAL PHILOSOPHY

VOLUME XVIII • NUMBER 1

*July 2020*

ARTICLES

- 1 The Demandingness of Virtue  
*Robert Weston Siscoe*
- 23 Disability as Inability  
*Alex Gregory*
- 49 False Exemplars: Admiration and the  
Ethics of Public Monuments  
*Benjamin Cohen Rossi*
- 85 Kidney Exchange and the Ethics of Giving  
*Philippe van Basshuysen*

DISCUSSION

- 111 Addressed Blame and Hostility  
*Benjamin De Mesel*

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*

Dale Dorsey

James Dreier

Julia Driver

Anca Gheaus

Errol Lord

Tristram McPherson

Colleen Murphy

Hille Paakkunainen

*Discussion Notes Editor*

Kimberley Brownlee

*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*

Jesse Wilson

*Copyeditor*

Susan Wampler

*Typesetting*

Matthew Silverstein



## THE DEMANDINGNESS OF VIRTUE

*Robert Weston Siscoe*

DIOGENES famously held that there are few, if any, virtuous people. He believed that the priorities of the Athenians, pursuits like the acquisition of wealth and popularity, were inimical to living a virtuous life. In a stark demonstration of his protest of Athenian life, he wandered Athens with a lamp in broad daylight, claiming he was looking for an honest man.<sup>1</sup> But were Diogenes's expectations too high? Our society has the same sorts of social pressures, the ever-present pull of greed and social status, yet we know many virtuous people. Even though Diogenes was right that Athenian culture was not a perfect incubator for virtue, surely he was wrong to think that this prevented its citizens from being virtuous altogether.

One way to cash out this opposing view to Diogenes is to say that virtue is a contextual matter. Whether or not someone possesses the virtue of honesty, for instance, is determined not only by the characteristics of the honest person but also by a contextual threshold for what is considered virtuous. On this account, it is possible that Athenians were virtuous so long as they were virtuous enough given their society. "Virtuous" and "honest," after all, are gradable adjectives, and a number of gradable adjectives do function with a contextual threshold. What is considered an expensive handbag may vary by social class, and what is called a long road may vary by county, making it plausible that the truth of gradable virtue adjectives also factors in context. In section 1, I further detail such contextual views of gradable adjectives, showing that they can be used as a response to Diogenes's criticism of the Athenians.

Before taking such contextual proposals on board, however, closer attention is needed to which sorts of gradable adjectives function with a contextual threshold. Contemporary work on gradable adjectives has differentiated be-

1 Diogenes Laertius, *Lives of Eminent Philosophers*, bk. IV, ch. 41. The more accurate translation of what Diogenes inquires as he is traveling through the streets is that "he is looking for a man." The reason that he cannot find one is that he takes man to be essentially rational, but by living apart from virtue, the citizens of Athens are also not living completely rational lives. This then leads to Diogenes's criticism that he cannot find a man as he cannot find anyone completely virtuous and thus rational.

tween absolute and relative gradable adjectives, a distinction whose explanatory power has quickly made it widely accepted within linguistics.<sup>2</sup> Despite the recent uptake of the divergence between absolute and relative gradable adjectives in linguistics, however, the contrast between them has been largely ignored within philosophy.<sup>3</sup> In sections 2 and 3, I introduce the relative/absolute distinction, showing that relative gradable adjectives typically function with a contextual threshold. When terms like “virtuous” and “honest” are also taken to be relative gradable adjectives, the natural result is an account on which their truth conditions differ along with the context.

The contextual threshold view is problematic, however, because “virtuous” and “honest” are absolute gradable adjectives, or so I argue in section 4. Unlike relative gradable adjectives, the denotation of absolute gradable adjectives remains fixed on the maximal element in the underlying scale regardless of context. Absolute gradable adjectives are thus not context sensitive, and so a context-dependent view will be of no use in replying to Diogenes. Terms like “honest” and “just” can only be used to specify something besides the highest point in their scales when they are used imprecisely, but imprecise attributions of “virtuous,” “honest,” and “just” are all literally false. For this reason, I will argue that Diogenes was correct in thinking that a number of the virtues require perfection and that there were perhaps none who were truly virtuous in Athenian society. Diogenes and his followers are not mistaken about the truth-conditions of “honest”; they simply refuse to use “honest” imprecisely.

### 1. VICIOUS CULTURES, VIRTUOUS CITIZENS

Diogenes’s critique, that a corrupt culture risks the virtue of its citizens, has maintained its appeal for contemporary virtue ethicists. Julia Annas argues that, because Stoics in Roman society tolerated slavery, they were not fully virtuous.<sup>4</sup> Lisa Tessman goes a step further, contending that because the majority of societies perpetuate systems of oppression, most people are not virtuous:

- 2 For work in linguistics on the distinction between the two types of adjectives, see Kennedy, “Vagueness and Grammar”; Kennedy and McNally, “Scale Structure and the Semantic Typology of Gradable Predicates”; and Rusiecki, *On Adjectives and Comparisons in English*.
- 3 Notable exceptions within the philosophy literature include Unger, *Ignorance*, and the recent discussion of aesthetic adjectives in Liao and Meskin, “Aesthetic Adjectives”; Liao, McNally, and Meskin, “Aesthetic Adjectives Lack Uniform Behavior”; Phelan, “Gradability and Multidimensionality in Aesthetic Adjectives”; and Stojanovic, “An Empirical Approach to Aesthetic Adjectives.”
- 4 Annas, *Intelligent Virtue*, 63–64.

Given the pervasive injustice of oppression, the high level of participation in maintaining structures of oppression, and the difficulty of unlearning traits associated with domination even for those who become critical, I see unjust and other vicious people as fairly ordinary.<sup>5</sup>

Tessman thinks that Annas's hunch is correct—people who fall short in some way are the typical case. The person who accepts the unjust inequalities at play in their society harbors some vice, an issue faced by the Roman Stoics as well as citizens of contemporary societies, and unless they actively resist such evils they compromise their own characters. Thus, it was just as Diogenes thought—virtuous persons are exceedingly rare.

Like with Diogenes's stance, the charge against Annas's and Tessman's accounts of virtue is that they are too demanding. Even though we live in a society that condones certain oppressive social arrangements, we surely know numerous virtuous people who share our cultural heritage. Annas expresses such a concern about her own account, worrying that the picture she has put forward is too far out of reach:

On this account, does virtue not turn out to be too ideal? While we have seen that thinking in terms of virtue can lead people to act virtuously in ways that go beyond their society and the ways of thinking they have grown up in, it might seem that even this will not render them fully virtuous. For they might be like the Roman Stoics, stuck with a society in which acting effectively in a fully virtuous way was not feasible. . . . Is virtue then an ideal which we can never hope to attain because of the contingency of living in societies which are always in some respect unjust?<sup>6</sup>

Annas is concerned that taking Diogenes's critique seriously undermines the thought that anyone is virtuous. If a person's virtue can be compromised even by participating in an unjust society, then such a charge indicts most, if not all, of the supposedly virtuous people who have lived. This criticism then suggests that no one has been virtuous, with none reaching the ideal that virtue holds out for us to attain.<sup>7</sup>

5 Tessman, *Burdened Virtues*, 56.

6 Annas, *Intelligent Virtue*, 64.

7 By the demandingness objection, I will mean the thought that even a bit of vice is enough to make one vicious. Owen Flanagan argues that there is no such thing as perfectly fulfilling all the virtues because the virtues may come into conflict (*Varieties of Moral Personality*, 33). Even if this point is correct, my thesis applies not only to what is required to possess all the virtues, but a single virtue as well. For a response to Flanagan, see Swanton, "Virtue Ethics and the Problem of Demandingness."

In response to this criticism, it is possible to lean on a threshold account of virtue. It could well be that whether or not a person is virtuous depends not just on their character but also on the situation in which they find themselves. It is thus possible for a Roman Stoic to be virtuous just so long as they are virtuous enough for their particular context. Christine Swanton provides just such an account, giving the following threshold analysis of virtue:

A virtue is a good quality of character, more specifically a disposition to respond to, or acknowledge, items within its field or fields in an excellent or *good enough way*.<sup>8</sup>

According to Swanton, whether or not the Athenian or the Roman Stoic is virtuous depends not on whether they are fully in accord with virtue, as Annas and Tessman argue, but whether they are doing *well enough* given their situation. Dan Russell follows suit, characterizing virtue as being virtuous enough:

It seems undeniable that *being virtuous enough* is a sufficient condition for being virtuous *tout court*—not perfectly virtuous or even virtuous without qualification, but nonetheless virtuous in a genuine, *bona fide* sense.<sup>9</sup>

Even if the ideal of virtue is unattainable, resting on a contextual account of virtue allows that persons might be virtuous despite the society in which they find themselves. All that is required is that they be virtuous enough for the context in which they find themselves, even if that culture is one that condones slavery or other oppressive social arrangements.

In the remainder of this paper, I will argue that contextual accounts of virtue cannot capture the behavior of all virtue-theoretic adjectives. A number of virtue terms are absolute gradable adjectives, and as such are not context sensitive. Instead, it is only true to call someone virtuous if they instantiate virtue to the maximal degree. This position also resolves a potential ambiguity in Annas's and

8 Swanton, *Virtue Ethics*, 19, emphasis added.

9 Russell, *Practical Intelligence and the Virtues*, 112. Even though Russell and Swanton both invoke a threshold view of virtue, only Swanton makes her view explicitly contextual: "The standards for virtue should reflect the fact that the world is marred by the difficulty of attaining (full) virtue, and the all too frequent occurrence of catastrophe, scarcity, evil, and conflict. The standards for meeting thresholds of virtue are thus plural, in the sense of contextual, and what is salient and important for behaviour to count as displaying virtue will vary" (*Virtue Ethics*, 3). While it is conceptually possible that a variety of gradable adjectives have a non-contextual threshold, this possibility is not currently a serious competitor in the linguistics literature. Because my goal is to bring work on the semantics of gradable adjectives to bear on virtue ethics, I thus only evaluate the possibility that virtue-theoretic adjectives are either relative or absolute.



Tessman's positions. Annas frames her worry in terms of whether the Roman Stoics can be fully virtuous, while Tessman is concerned about whether ordinary citizens are virtuous at all. If we take virtue to be a threshold concept, then virtue can come apart from full virtue, making the worries of Annas and Tessman generate separate concerns. When we take virtue to be full virtue, however, then the criticisms of Annas and Tessman are one and the same. The Roman Stoics not only fall short of full virtue, but thereby fail to be virtuous full stop.<sup>10</sup>

## 2. RELATIVE AND ABSOLUTE GRADABLE ADJECTIVES

### 2.1. Relative Gradable Adjectives

Gradable adjectives come in two forms: relative gradable adjectives and absolute gradable adjectives.<sup>11</sup> A relative gradable adjective (RGA), such as "long," "small," and "short," draws some of its semantic characteristics from the fact that it has an underlying degree scale. Objects are placed along the underlying scale, making use of comparatives true or false.<sup>12</sup> For example, 1 is true just in case the rod has a greater degree of length than the stick:

1. The rod is longer than the stick.

The most distinctive characteristic of RGAs, however, is their use of a contextual

10 In Russell's words, virtuous *tout court*. In this paper, I will not be taking on a substantial commitment to any particular understanding of virtue, Aristotelian or otherwise, but it is possible that the results of this paper will bear on such proposals. Just as the view I will argue for undercuts satisficing accounts of the virtues, it may be that arguing that some virtue terms are AGAs has other theoretical implications as well. Thank you to an anonymous reviewer for encouraging me to specify how I understand the arguments of this paper to apply to more detailed accounts of the virtues.

11 Kennedy, "Vagueness and Grammar"; Kennedy and McNally, "Scale Structure and the Semantic Typology of Gradable Predicates"; Rotstein and Winter, "Total Adjectives vs. Partial Adjectives"; Rusiecki, *On Adjectives and Comparisons in English*; and Unger, *Ignorance*.

12 Though the scale approach of Cresswell ("The Semantics of Degree"), Heim ("Degree Operators and Scope"), Kennedy ("Vagueness and Grammar"), and von Stechow ("Comparing Semantic Theories of Comparison") has been the most influential, the primary competitor is a view on which the extension of a gradable adjective displays contextual shifts with the basis for comparatives being quantifications over possible precisifications of the adjective's extension, a view whose development runs through Fine, "Vagueness, Truth, and Logic"; Kamp, "Two Theories of Adjectives"; Klein, "A Semantics for Positive and Comparative Adjectives"; Larson, "Scope and Comparatives"; and Pinkal, *Logic and Lexicon*. The scale view holds a distinct advantage though in explaining the distinction between relative and absolute adjectives (Kennedy, "Vagueness and Grammar").

threshold. The ordering on the underlying scale is not all that goes into whether 2 is true:

2. The rod is long.

Whether 2 is true or not requires not only an ordering on the underlying scale but also a threshold. If the rod's length is clearly greater than the threshold, then 2 is true, and if the degree of its length is clearly lower, then 2 is false.<sup>13</sup> This threshold is also contextually determined. For example, the length of hair that an infant must possess in order to have long hair is less than the length of hair that an adult must have in order to be truthfully described as having long hair. The truth of 2, then, requires not only an underlying scale but also a contextually adjustable threshold.

Since RGAS have a context-sensitive threshold, it is possible to use this threshold to distinguish between two objects that differ on the underlying scale.<sup>14</sup> As specified before, the degree to which the object possesses the property must "stand out" relative to the threshold.<sup>15</sup> Assuming that this condition is satisfied, in a selection task with two pieces of metal pipe, 3 is felicitous when one pipe is clearly longer than the other even if the pieces are not particularly long:

3. Point to the long one.

Statement 3 is a felicitous request due to the contextually variable threshold on the underlying scale. This threshold, however, is not endlessly accommodat-

13 For more on determining the truth conditions of relative gradable adjectives, see Rotstein and Winter, "Total Adjectives vs. Partial Adjectives"; Kennedy, "Vagueness and Grammar" and "Vagueness and Comparison"; and Kennedy and McNally, "Scale Structure and the Semantic Typology of Gradable Predicates." The core idea is that, in order to accommodate cases of vagueness, the objects that fall within the extension of an RGA must "stand out" relative to the contextual threshold, a view that has a lengthy history. See Barker, "The Dynamics of Vagueness"; Bartsch and Vennemann, "The Grammar of Relative Adjectives and Comparison"; Bierwisch, "The Semantics of Gradation"; Cresswell, "The Semantics of Degree"; Fine, "Vagueness, Truth, and Logic"; Kamp, "Two Theories of Adjectives"; Klein, "A Semantics for Positive and Comparative Adjectives"; Lewis, "General Semantics"; Ludlow, "Implicit Comparison Classes"; Pinkal, *Logic and Lexicon*; Sapir, "Grading"; von Stechow, "Comparing Semantic Theories of Comparison"; and Wheeler, "Attributives and Their Modifiers."

14 See Kennedy, "Vagueness and Grammar"; Kyburg and Morreau, "Fitting Words"; Sedivy et al., "Achieving Incremental Semantic Interpretation through Contextual Representations"; Syrett et al., "Shifting Standards"; and Syrett, Kennedy, and Lidz, "Meaning and Context in Children's Understanding of Gradable Adjectives."

15 See Rotstein and Winter, "Total Adjectives vs. Partial Adjectives"; Kennedy, "Vagueness and Grammar" and "Vagueness and Comparison"; and Kennedy and McNally, "Scale Structure and the Semantic Typology of Gradable Predicates."

ing—at the extreme parts of the scale, 3 is not always felicitous. Consider, for example, two roads, one that is seven thousand miles long and the other that is eight thousand miles long. The infelicity of 3 in such a case demonstrates that the contextual threshold at issue with RGAS does not have an unbounded ability to distinguish between objects that differ on the underlying scale.

Another feature of RGAS is that they always give rise to Sorites paradoxes. With “long,” the paradox gets going with 4:

4. For whatever length an object is, one centimeter of length does not change whether it is long or not.

The trouble of course is created by the fact that 4 seems true. This can be explained by the fact that RGAS cannot be used to pick out a maximally specific point on the underlying scale. As previously noted, an object falls with the extension of an RGA like “long” if it possesses a degree of height that is *clearly greater* than the relevant threshold, a description meant to accommodate for borderline cases. Even though there are roads that are obviously long and others that are not, for some it is unclear whether they should be classified as long or short. This vagueness makes it such that 4 is intuitively correct, leading to the Sorites.<sup>16</sup>

## 2.2. Total and Partial Absolute Gradable Adjectives

Just like RGAS, absolute gradable adjectives, words that include “closed,” “healthy,” “pure,” and “straight,” can be used comparatively. Based on their position on the degree scale, objects can be compared and contrasted, with a use of 5 being true so long as the road is closer to 180 degrees than the river:

5. The road is straighter than the river.

As with RGAS, however, the underlying scale is not the only thing that accounts for the truth of the unmarked form in 6:

6. The road is straight.

Just like with RGAS, 6 is true only if the road is straight to a particular degree. In contrast to RGAS, however, this degree does not change according to context. Even when one stick is clearly straighter than another stick, if they are both somewhat bent, 7 is infelicitous:

7. #Point to the straight one.

<sup>16</sup> For evaluations of what types of semantic accounts of gradable adjectives can diagnose the existence of borderline cases and the Sorites paradox, see Graff Fara, “Shifting Sands”; Kennedy, “Vagueness and Grammar”; Pinkal, *Logic and Lexicon*; and Rusiecki, *On Adjectives and Comparisons in English*.

Unlike an RGA, then, an absolute gradable adjective (AGA) does not have a threshold that is contextually accommodating.<sup>17</sup>

It might seem that the difficulty with 7 is the same problem that we encountered with the two very long roads. Why not say that there are parts of the degree scale that both unmarked RGAs and AGAs are incapable of distinguishing between? The difference between the two is that, whereas the infelicity of 7 occurs in the middle of the underlying scale, the trouble with 3 happens at the extremes of the scale. The sticks are both slightly bent, whereas the roads are thousands of miles long. If we change the situation though, using the AGA at the extreme end of the scale, the situation shifts. In this case, the AGA will be able to distinguish between the two objects. If we have a perfectly straight stick and one that is slightly bent, then 7 *is* felicitous. Instead of getting the result that the problem with the long roads and 3 and the two bent sticks with 7 is the same issue, these difficulties actually reinforce the contrast between AGAs and RGAs. AGAs pass the “point to” test at the extreme ends of their scales but not in the middle, while RGAs pass the test in the middle of their scales but not at their extremes.

Further distinguishing AGAs from RGAs is that they do not always create Sorites paradoxes. Their lack of contextual threshold allows them to pick out a maximally specific point, making Sorites-driving principles intuitively unappealing. If a stick is bent just one degree, then 8 does not have the same pull as 4:

8. For however bent a stick is, one degree of bend does not change whether it is straight or not.

For a stick that is one degree from being straight, 8 is false—straightening the stick by one degree *does* make the stick straight. This is due to the clear boundary between being straight and not being straight. There is not a similar boundary, however, with “long,” making the Sorites-paradoxical result inevitable for it and other RGAs.<sup>18</sup>

Just as gradable adjectives can be differentiated into relative and absolute, it is possible to distinguish absolute gradable adjectives as total and partial. Total AGAs are distinguished by picking out an end point on the underlying scale. A

17 See Aparicio, Miang, and Kennedy, “Processing Gradable Adjectives in Context”; Kennedy, “Vagueness and Grammar”; Kyburg and Morreau, “Fitting Words”; Qing and Franke, “Gradable Adjectives, Vagueness, and Optimal Language Use”; Sedivy et al., “Achieving Incremental Semantic Interpretation through Contextual Representations”; Syrett et al., “Shifting Standards”; and Syrett, Kennedy, and Lidz, “Meaning and Context in Children’s Understanding of Gradable Adjectives.”

18 Burnett (“A Delineation Solution to the Puzzles of Absolute Adjectives”), Kennedy (“Vagueness and Grammar”), and Pinkal (*Logic and Lexicon*), all point out cases in which absolute gradable adjectives avoid the Sorites.

straight line is not bent to any degree and a closed door is not at all open. Partial AGAS, on the other hand, are used truthfully when an object deviates from the end of the scale. One degree off from straightness counts as bent, just as a door being ever so slightly ajar makes it open.<sup>19</sup> As with the two examples currently in focus—straight/bent and closed/open—AGAS often come in total/partial pairs. This corresponds to whenever AGAS are contradictories, as “bent” is synonymous with “not straight” and “open” with “not closed.” How to distinguish between total and partial AGAS? A common test for whether a gradable adjective has a top-closed scale is whether the modifier “completely” can be used to pick out a maximal element.<sup>20</sup> Such constructions are thus felicitous for total AGAS:

9. The door is completely closed.

10. The stick is completely straight.

11. The gold is completely pure.

RGAS, however, do not accept similar attempts at modification:

12. The road is completely long.

Partial AGAS sometimes do accept the modifier “completely,” but not because they have top-closed scales. One instance in which this occurs is when “completely” functions as synonymous with “very,” as in 13:

13. The stick is completely bent.

The use of “completely” in 13 can be seen to be synonymous with “very” using an entailment test. When “completely” is synonymous with “very,” it is possible that something can possess the partial AGA to an even greater degree, while when “completely” is used to pick out the maximum degree on a scale with a total AGA, this is not possible.<sup>21</sup> Suppose that a mechanic is comparing two exhaust pipes, both so bent that they are no longer roadworthy. In such a case, 14 is felicitous:

19 For more on the distinction between total and partial absolute adjectives, see Burnett, “A Delineation Solution to the Puzzles of Absolute Adjectives”; Cruse, *Lexical Semantics*; Kamp and Rossdeutscher, “DRS-Construction and Lexically Driven Inferences”; Kennedy, “Vagueness and Grammar”; Rotstein and Winter, “Total Adjectives vs. Partial Adjectives”; and Yoon, “Total and Partial Predicates and the Weak and Strong Interpretations.”

20 See Burnett, “A Delineation Solution to the Puzzles of Absolute Adjectives”; Cruse, *Lexical Semantics*; Kennedy, “Vagueness and Grammar”; Kennedy and McNally, “Scale Structure and the Semantic Typology of Gradable Predicates”; and Rotstein and Winter, “Total Adjectives vs. Partial Adjectives.”

21 See Kennedy and McNally, “Scale Structure and the Semantic Typology of Gradable Predi-

14. This pipe is completely bent, but that other exhaust pipe is bent even further.

It is not felicitous, however, to say 15:

15. #That pipe is completely straight, but it could still be straightened just a bit more.

A useful diagnostic for testing for the presence of a closed scale is modifying a gradable adjective with “completely.” The test is not foolproof—we have seen that in some cases “completely” permits of non-maximal interpretations—but if an adjective has a maximum on its underlying scale, then it should accept modification by “completely.”

### 3. VAGUENESS AND IMPRECISION

Even though total AGAS function without a contextually accommodating threshold and pick out the endpoint on the underlying scale, contexts can be created in which approximate uses of total AGAS are acceptable. When the purpose of drawing a straight line is to illustrate a rough line graph, then a line that is off by a few degrees can be called straight. Likewise, there are situations in which it is acceptable to call a door closed even if it has not completely slammed shut. In these types of situations, total AGAS *can* generate a Sorites, for it is difficult to say whether one more degree of bend will make the line less than straight, or whether being an inch more open will change the status of the door from closed to open.

The fact that uses of total AGAS can create issues with the Sorites raises some difficult questions. On what grounds can we say that the contextual variation at issue with total AGAS is not, like that of RGAS, that of a contextual threshold? Does the fact that AGAS permit some variety of context shifting enough to undermine the thought that they are distinct from RGAS? As we have seen, AGAS display a distinct profile when it comes to the “point to” test, so we have good grounds for saying that they function differently from RGAS. Using this as our starting point, it can also be argued that the best way to characterize the contextual variation permitted by AGAS is, unlike RGAS, that of imprecision.<sup>22</sup> With

---

cates,” 354.

22 For work on imprecision as it relates to absolute gradable adjectives, see Pinkal, *Logic and Lexicon*; and Kennedy, “Vagueness and Grammar.” For more on imprecision more generally, see Krifka, “Be Brief and Vague!” and “Approximate Interpretations of Number Words”; Lakoff, “Hedges”; Sadock, “Truth and Approximations”; and Sauerland and Stateva, “Scalar vs. Epistemic Vagueness” and “Two Types of Vagueness.”

imprecise terms, a particular point plays a privileged role in its meaning. Consider, for instance, saying that a rod is ten inches long. The exact length of ten inches plays a critical role in the meaning of “ten inches” even though there are situations in which it is acceptable to use this locution imprecisely. If a stick is nine inches long, no one will raise much of a fuss if it is described as being ten inches, but this does not tell against the fact that the precise length of ten inches is central to its meaning. With total AGAS, we can think of the endpoint of the scale as taking on the role of being exactly ten inches long. Even though there are cases in which it can be acceptable to use “ten inches” and “straight” imprecisely, this does not undermine the thought that an exact point captures the meaning of the terms. Thus, the contextual variability permitted by AGAS is not due to a contextual threshold, differentiating AGAS from RGAS.

Further corroborating the thought that AGAS differ from RGAS in allowing imprecise use, AGAS allow what is known as natural precisification. Vague terms can only be made more precise by stipulation, whereas imprecise terms can be precisified simply by particular uses in natural-language contexts. If a construction worker provides a plank that is eleven inches long, the contractor can reject the board saying:

16. The length of the board needs to be ten inches, but this one is only nine inches so it will not do.

With a use of 16, the foreman eliminates imprecise uses of “ten inches” from consideration. Such precisification cannot occur with RGAS. Suppose, for example, that the foreman wants a long steel girder, over one hundred feet, to serve as one of the supports of a skyscraper. He cannot precisify the meaning of “long” to that of a precise value with a use of 17:

17. I need a long girder, but this one is only ninety feet so it will not do.

Even though the foreman needs a steel girder that is over one hundred feet long, he cannot communicate this by precisifying the meaning of “long” using 17 like the contractor can precisify the meaning of “ten inches” with 16. Instances of imprecision can be revealed, then, by the possibility of natural precisification.

What about total AGAS? Do they permit natural precisification? Consider: a satellite needs a perfectly straight antenna in order to make proper readings. If an antenna is slightly curved, an engineer can reject it with 18:

18. The antenna needs to be straight, but this one is slightly curved so it will not do.

Statement 18 can be used to disallow any imprecise applications of “straight.”

“Straight” is far from the only total AGA that allows natural precisification. If a parent wants to prevent intruders from entering their home during the night, they can chastise their children with 19 even if the door was just barely cracked open:

19. We are trying to make sure that no one comes in while we are sleeping,  
so next time the door needs to be closed.

The use of “closed” in 19 will not allow any imprecision—only being completely closed will be acceptable with this precisified meaning of “closed.” Thus, it appears that total AGAs do accept natural precisification, confirming that they can be used imprecisely.

Our discussion thus far has revealed some important features of acceptable imprecision. Imprecision is permitted whenever it does not interfere with the practical purposes at hand. Sometimes a board that is approximately ten inches is all that is needed, and in such cases, imprecise uses of “ten inches” will be acceptable. Likewise, rough lines and slightly open doors can sometimes do the job, and it is in these situations that imprecise uses of “straight” and “closed” are permissible. This realization, that the practical circumstances regulate acceptable imprecise attributions, can also be used to respond to critics. Assaf Toledo and Galit Sassoon criticize the orthodox view of AGAs, arguing that the account does not stand up to natural-language data.<sup>23</sup> In particular, they present the following, supposedly problematic, examples:

20. This kitchen knife is clean.  
21. This surgical instrument is clean.  
22. This child’s shirt is dirty.  
23. This tuxedo is dirty.

If total AGAs are used to pick out the maximal element on the underlying scale, then why are uses of 20 and 22 acceptable in cases of cleanliness and dirtiness that are less stringent than is required for uses of 21 and 23? According to Toledo and Sassoon, such cases show that the interpretation of AGAs depends on more than just an end point in a scale. They claim that other features come into play as well, like the object type to which the total AGA is being applied, undermining the standard account of the semantics of AGAs.

An obvious rejoinder to Toledo and Sassoon is to take what we have articulated about permissible imprecision and apply it to 20–23. Imprecision is al-

<sup>23</sup> Toledo and Sassoon, “Absolute vs. Relative Gradable Adjectives.”



lowed whenever the purposes at hand can be satisfied, and in the case of kitchen knives and surgical instruments, it is worth noting that they are typically used with quite distinct purposes in mind. Surgical instruments must be cleaned of all microscopic bacteria, whereas kitchen utensils only need to be cleansed to be rid of food scraps that are visible to the naked eye. Thus, the standards for an attribution of 21 are much more stringent than a use of 20, explaining how they can come apart appealing only to the mechanisms of imprecision. Similarly, tuxedos are worn in circumstances where being spotlessly clean is the norm, so acceptable uses of 23 require less dirt than those of 22. For this reason, the objections of Toledo and Sassoon can be accommodated by the previous discussion of allowable imprecision. Imprecise use is countenanced whenever it does not interfere with the purposes at hand, predicting that acceptable uses of 20 and 22 will come apart from acceptable applications of 21 and 23.

#### 4. VIRTUE TERMS AS ABSOLUTE GRADABLE ADJECTIVES

##### 4.1. *The Virtue-Theoretic AGAS*

All virtue-related adjectives are obviously gradable, as they can be used both with degree modifiers and comparatively:

24. Susan is more virtuous/honest/just than Mark.

25. Susan is very virtuous/honest/just, the most virtuous/honest/just person I know.

The gradability of virtue terms, however, is not what is at issue between the contextual threshold view and the absolute account. What matters is the behavior of the unmarked form of virtue-theoretic adjectives, the sorts of tests that distinguish between RGAS and AGAS.

Before getting into these tests, it is important to note a couple of characteristics of virtue terms. To begin with, a number of object types accept modification by virtue-related adjectives. Since our focus is on virtue ethics, the examples I consider will employ actions and persons, the object types relevant to virtue ethics. Another difficulty to flag is that being honest or just, or being virtuous more generally, is not simply a product of concrete actions. Virtues are dispositional, and as such are vulnerable to masking and otherwise failing to manifest. A dishonest person can have told no lies, even though they are so predisposed, simply by having no opportunities to lie. Thus, we will not only look at examples of occurrent actions, but merely possible actions as well. On then to the tests.

At both the middle and the extremes of their scales, a number of virtue-the-

oretic adjectives, including “honest,” “just,” and “virtuous,” function like AGAS. Let us start with honest. Suppose that Susan and Mark are both relatively middling when it comes to honesty—not perfectly honest, but also not egregious liars: Mark has told four white lies this week and Susan only two. Even though 24 is true, the request in 26 is infelicitous:

26. #Point to the honest/virtuous person.

It thus appears that “honest” is an AGA, a thought confirmed by the fact that if Susan has told no lies and Mark has told two then 26 *is* acceptable. When we have someone who is perfectly honest, we have arrived at the end of the scale, and it is here that the “point to” test is felicitous for AGAS.

Because the scale of honesty has an end point, there are also cases in which it does not permit the Sorites. For example, 27 is clearly false when considering the difference between telling one lie or none at all:

27. For however dishonest a person is, telling one less lie will not make them honest.

Of course, we noted earlier that it is not merely the telling of lies that makes one dishonest, but also the disposition to tell lies. To account for this, we can alter 27 as follows:

28. For however dishonest a person is, the disposition to tell one less lie will not make them honest.

With 28, we are not simply altering the manifestations of the person’s dispositions, but whether they have the disposition altogether. Changing a person from being disposed to tell a lie in a particular scenario to no longer being so disposed does in fact change whether or not they are honest, preventing the Sorites from getting started.

Both of these results also obtain with the virtue of justice. Consider: Susan and Mark are both business owners. For the same task, Susan pays one of her workers twenty dollars and another fifteen dollars. Mark, though, is even worse. He pays one employee twenty dollars and the other five dollars even though they perform the same amount of labor. Statement 24 is applicable in this case, as Susan is more just than Mark, but nevertheless, both the commands in 29 and 30 are unacceptable:

29. #Point to the just/virtuous person.

30. #Point to the owner who justly paid their employees.

Just as the “point to” test fails in the middle of the scale, though, it succeeds at the end of the scale. If Susan pays both of her staff members twenty dollars while Mark pays one twenty dollars and the other nineteen, we can differentiate between them with the unmarked form of just in 29 and 30. This result holds up in relation to the Sorites as well. At the end of the scale, 31 is false, depriving the paradox of its crucial premise:

31. For however unjust a person is, paying their employees a dollar more will not make them just.

The case is very clear. Both “honest” and “just” behave as AGAS with the Sorites as well as the “point to” test.

Interestingly enough, the arguments for the absolute accounts of honesty and justice are also arguments that being virtuous requires being perfectly virtuous. When 24 is infelicitous for “honest,” it is also infelicitous for “virtuous,” and likewise, whenever 29 is unacceptable for “just,” it is unacceptable for “virtuous.” The same is true of natural precisification. Virtue terms are sometimes used to rule out those that are quite vicious. If a job candidate lies several times on an application, a manager could refuse to consider them with a use of 32:

32. We need someone who is honest/virtuous.

On other occasions though, virtue language can be used to rule out those who are vicious even to a small degree. Suppose that an intelligence agency is hiring a spy and, in the course of their interview, even though they do not lie, they nevertheless fail to divulge one of their previous marriages. Intelligence staff may move on from the candidate with a use of 32 even though the candidate is very honest, a much stronger standard than someone who outright lies on their application. Likewise, suppose that a police officer is being considered for a junior-level promotion, but it is found that he arrested someone purely on the basis of their race. The committee should dismiss them with a use of 33:

33. We need someone who is just/virtuous.

The same locution can be used, however, for even a minute breach of justice. If it comes to light that the would-be sheriff passed over one of his friends for their routine traffic violation, 33 can be given as a justification for preventing him from assuming his new post. Terms for the individual virtues, as well as the more general “virtuous,” are thus subject to natural-language precisification depending on the practical task at hand, providing compelling evidence that they are all AGAS.

#### 4.2. Are All Virtue-Related Adjectives AGAs?

It is clear that a number of virtue terms are AGAs. As we have seen, “virtuous,” “honest,” and “just,” all pass the “point to” and precisification tests for AGAs, making it clear that they are distinct from RGAs. A further question, however, is whether *all* virtue-theoretic adjectives are AGAs. Take, for example, another test we considered in section 2, modification by “completely.” All of the three virtue terms discussed thus far pass the “completely” test:

34. Susan is completely virtuous/honest/just.

Each of “virtuous,” “honest,” and “just” has no issue being modified by “completely,” exactly what we should expect from AGAs. There are other virtue terms, however, that present difficulties. Consider, for example, “courageous” and “generous”:

35. #Susan is completely generous/courageous.

Statement 35 is not as obviously felicitous as are all the variations of 34, calling into question both whether all of the virtues have maximal-degree interpretations as well as whether all virtue-related adjectives are AGAs.

How should we make sense of the infelicity of 35? One possibility is that the “completely” test does not totally bifurcate RGAs from AGAs. “Courageous,” for example, functions in the “point to” test as an AGA in the middle of its scale. If Mark hides in his foxhole while Susan wills herself out of the trench but then is overcome with cowardice at the gates of the enemy’s fortress, 36 is infelicitous:

36. #Point to the person whose war effort was courageous.

Had Susan continued to infiltrate the enemy stronghold, 36 would have been an acceptable request, and so there is at least one data point that supports an AGA interpretation of “courageous.” It also makes sense to think of courage as the absence of any cowardice, making “courageous” and “cowardly” a total/partial pair. One verdict to draw in this case then is that modification by “completely” is not an infallible test for diagnosing AGAs.

Another way to interpret the infelicity of 35 is to say that some virtue terms, like “generous” and “courageous,” are not AGAs like the other virtue-theoretic adjectives we have surveyed. The difficulty with going this route is that it makes it challenging to understand what maximal virtue would be like. “Virtuous” is a multidimensional adjective in that there are multiple factors that determine whether it applies. One way of understanding perfect virtue, then, is to understand it as perfection in all of its underlying dimensions, a conjunctive, multidimensional

mensional adjective that only applies when all of its more basic elements apply. “Vicious,” then, is a disjunctive multidimensional adjective in that any element of vice is enough to make one vicious.<sup>24</sup> If generosity and courage, however, do not have *maxima*, how should we understand “virtuous”? It could be that, of virtue’s underlying dimensions, some require perfection while others only call for jumping over a contextual threshold.

Such a picture necessarily complicates the picture of what complete virtue is like, raising the question of why open-ended underlying virtues fail to also make “virtuous” an RGA. Suppose, for example, that “generous” is an RGA and generosity has an open scale. If two people are both above the threshold for “generous” on the underlying scale, then even if one is more generous than the other, this may not make one more virtuous than the other. If they are both completely virtuous, that is, if all the virtue terms truly apply to them, then the fact that “virtuous” is an AGA predicts that it will be false to say that one person is more virtuous than the other. This would be a surprising result, that the RGA nature of some of the underlying virtues does not “bubble up” to make “virtuous” itself an RGA. Thus, taking virtue terms to be a mixed bag of RGAs and AGAs also creates difficulties.

### 4.3. *Virtue and Context*

Despite the issues raised by 35, the argument of this paper is that a number of the virtues require perfection, and this only necessitates that some of the virtues be AGAs. As we have seen, this is true of “honest,” “just,” and perhaps “virtuous,” a point that is significant even if it does not generalize to all virtue-related adjectives. Whether or not all virtue-theoretic adjectives are in fact AGAs is beyond the scope of this paper, so at this juncture I would like to turn to the lessons that can be learned from the virtue terms that *are* AGAs.

Because some virtue terms are AGAs, their meanings are fixed on their maximal degrees. In order for it to be true that an Athenian is honest, they must be completely honest, and for it to be true that a Roman Stoic is just, they must be completely just. It is possible to speak, as we often do, of honest parents, honest teachers, and honest friends, but all of these attributions are instances of imprecision and ultimately false. This is not to say that imprecise uses of virtue-theoretic adjectives do not have their place; such uses can have practical value. A theory of virtue, though, does not have to answer to imprecise applications of the concept. The fact that many virtue terms are AGAs also means that Swanton’s and

<sup>24</sup> For more on the typology of multidimensional adjectives, see Sassoon, “A Typology of Multidimensional Adjectives.” Thank you to an anonymous reviewer for suggesting that I note the multidimensional nature of “virtuous.”

Russell's threshold views cannot account for all of the behavior of virtue terms. For some virtues, it is not enough to be virtuous to simply be good enough for one's circumstances. Strictly speaking, it is true that a person is honest, just, or virtuous only if they are perfectly honest, just, or virtuous.

Diogenes was right to think that there were very few, if any, virtuous Athenians. Given that they lived in a corrupting society, it was very difficult, if not impossible, to be completely virtuous. Similarly, Annas and Tessman do not need to worry that their accounts seem particularly demanding.<sup>25</sup> It turns out that the nature of virtue is demanding—it is only true to say that someone is virtuous if they are perfectly virtuous—and so it is no surprise that participating in a society of systematic oppression compromises the virtue of its citizens.<sup>26</sup>

The arguments of this paper may have wider application. Just as they have shown that it is almost impossible to be virtuous in an unjust society, they could also be used to show that it is very challenging to be virtuous under a just political order. The problem of unjust societies, however, is a particularly poignant case, one that calls to mind quite readily the possibility that virtue terms are context sensitive. The rejection of such an account in favor of the view that a number of virtue-theoretic adjectives are absolute has far-reaching implications for theorizing about virtue, implications that may indeed reach far beyond the problem of unjust societies.<sup>27</sup>

University of Arizona  
rwsiscoe@email.arizona.edu

- 25 Even though the current account vindicates Annas's view on virtue in unjust societies, it does not support her understanding of virtue in all cases. Annas, for instance, says "Being truly (or perfectly) virtuous is indeed an ideal which none of us can exemplify. But it does not follow that none of us are brave, loyal, or generous" (*Intelligent Virtue*, 65). On the present view, if any of "brave," "loyal," or "generous" are AGAS, then any cases in which it is acceptable to say that we are brave, loyal, or generous when we are not completely so are strictly speaking false virtue attributions.
- 26 The view that some virtues are AGAS is compatible with either view in the debate over whether external goods are merely instrumental to virtue or partially constitutive of virtue. The view that one takes on the external goods question may well have consequences for what full virtue requires, but the absolute view can be made consistent with either view on the role of external goods in becoming virtuous.
- 27 For helpful discussion and input on this project, I am indebted to Maria Altepeter, Julia Annas, Stew Cohen, Juan Comesana, Christopher Kennedy, Daniel Nolan, Nathan Oakes, Jeremy Reid, Tristan Rogers, Jackie Sideris, Joshua Stuchlik, Bjorn Wastvedt, Jonathan Weinberg, Sean Whitton, and two anonymous reviewers from this journal.

## REFERENCES

- Annas, Julia. *Intelligent Virtue*. Oxford: Oxford University Press, 2011.
- Aparicio, Helena, Ming Xiang, and Christopher Kennedy. "Processing Gradable Adjectives in Context: A Visual World Study." *Proceedings of SALT 25* (2015): 413–32.
- Barker, Chris. "The Dynamics of Vagueness." *Linguistics and Philosophy* 25, no. 1 (February 2002): 1–36.
- Bartsch, Renate, and Theo Venneman. "The Grammar of Relative Adjectives and Comparison." *Linguistische Berichte* 21 (1972): 19–32.
- Bierwisch, Manfred. "The Semantics of Gradation." In *Dimensional Adjectives*, edited by Manfred Bierwisch and Ewald Lang, 71–261. Berlin: Springer-Verlag, 1989.
- Burnett, Heather. "A Delineation Solution to the Puzzles of Absolute Adjectives." *Linguistics and Philosophy* 37, no. 1 (February 2014): 1–39.
- Cresswell, Max. "The Semantics of Degree." In *Montague Grammar*, edited by Barbara Partee, 261–92. New York: Academic Press, 1977.
- Cruse, D. A. *Lexical Semantics*. Cambridge: Cambridge University Press, 1986.
- Fine, Kit. "Vagueness, Truth, and Logic." *Synthese* 30, no. 3/4 (April–May 1975): 265–300.
- Flanagan, Owen. *Varieties of Moral Personality: Ethics and Psychological Realism*. Cambridge, MA: Harvard University Press, 1991.
- Fox, Sue. *Etiquette for Dummies*. Hoboken, NJ: John Wiley and Sons, 2011.
- Graff Fara, Delia. "Shifting Sands: An Interest-Relative Theory of Vagueness." *Philosophical Topics* 28, no. 1 (Spring 2000): 45–81.
- Heim, Irene. "Degree Operators and Scope." *Proceedings of SALT 10* (2000): 40–64.
- Kamp, Hans. "Two Theories of Adjectives." In *Formal Semantics of Natural Language*, edited by Edward Keenan, 123–55. Cambridge: Cambridge University Press, 1975.
- Kamp, Hans, and Antje Rossdeutscher. "DRS-Construction and Lexically Driven Inferences." *Theoretical Linguistics* 20, no. 2–3 (January 1994): 165–235.
- Kennedy Christopher. "Vagueness and Comparison." In *Vagueness and Language Use*, edited by Paul Égré and Nathan Klinedinst, 73–97. London: Palgrave Macmillan, 2011.
- . "Vagueness and Grammar: The Semantics of Relative and Absolute Gradable Adjectives." *Linguistics and Philosophy* 30, no. 1 (February 2007): 1–45.

- Kennedy, Christopher, and Louise McNally. "Scale Structure and the Semantic Typology of Gradable Predicates." *Language* 81, no. 2 (2005): 345–81.
- Klein, Ewan. "A Semantics for Positive and Comparative Adjectives." *Linguistics and Philosophy* 4, no. 1 (March 1980): 1–45.
- Krifka, Manfred. "Approximate Interpretations of Number Words: A Case of Strategic Communication." In *Cognitive Foundations of Interpretation*, edited by Gerlof Bouma, Irene Krämer, and Joost Zwarts, 111–26. Amsterdam: Koninklijke Nederlandse Akademie van Wetenschappen, 2007.
- . "Be Brief and Vague! And How Bidirectional Optimality Theory Allows for Verbosity and Precision." In *Sounds and Systems, Studies in Structure and Change: A Festschrift for Theo Vennemann*, edited by David Restle and Dietmar Zaefferer, 429–48. Berlin: Mouton de Gruyter, 2002.
- Kyburg, Alice, and Michael Morreau. "Fitting Words: Vague Language in Context." *Linguistics and Philosophy* 23, no. 6 (December 2000): 577–97.
- Lakoff, George. "Hedges: A Study in Meaning Criteria and the Logic of Fuzzy Concepts." *Journal of Philosophical Logic* 2, no. 4 (October 1973): 458–508.
- Larson, Richard. "Scope and Comparatives." *Linguistics and Philosophy* 11, no. 1 (February 1988): 1–26.
- Lewis, David. "General Semantics." *Synthese* 22, no. 1–2 (December 1970): 18–67.
- Liao, Shen-yi, Louise McNally, and Aaron Meskin. "Aesthetic Adjectives Lack Uniform Behavior." *Inquiry* 59, no. 6 (2016): 618–31.
- Liao, Shen-yi, and Aaron Meskin. "Aesthetic Adjectives: Experimental Semantics and Context-Sensitivity." *Philosophy and Phenomenological Research* 94, no. 2 (March 2017): 371–98.
- Ludlow, Peter. "Implicit Comparison Classes." *Linguistics and Philosophy* 12, no. 4 (August 1989): 519–33.
- McCarthy, Ed, and Mary Ewing-Mulligan. *Wine for Dummies*. Hoboken, NJ: John Wiley and Sons, 2015.
- McNally, Louise. "The Relative Role of Property Type and Scale Structure in Explaining the Behavior of Gradable Adjectives." In *Vagueness in Communication*, edited by Rick Nouwen, Robert van Rooij, Uli Sauerland, and Hans-Christian Schmitz, 151–68. Berlin: Springer, 2011.
- Phelan, Mark. "Gradability and Multidimensionality in Aesthetic Adjectives." Unpublished manuscript.
- Pinkal, Manfred. *Logic and Lexicon*. Dordrecht: Kluwer, 1995.
- Qing, Ciyang, and Michael Franke. "Gradable Adjectives, Vagueness, and Optimal Language Use: A Speaker-Oriented Model." *Proceedings of SALT* 24 (2014): 23–41.
- Rotstein, Carmen, and Yoad Winter. "Total Adjectives vs. Partial Adjectives:



- Scale Structure and Higher-Order Modifiers." *Natural Language Semantics* 12, no. 3 (September 2004): 259–88.
- Rusiecki, Jan. *On Adjectives and Comparisons in English*. New York: Longman Linguistics Library, 1985.
- Russell, Daniel. *Practical Intelligence and the Virtues*. Oxford: Oxford University Press, 2009.
- Sadock, Jerrold. "Truth and Approximations." In *Proceedings of the Third Annual Meeting of the Berkeley Linguistics Society*, edited by Kenneth Whistler Jr., Robert D. Van Valin Jr., Chris Chiarello, et al., 430–39. Berkeley: Berkeley Linguistics Society, 1977.
- Sapir, Edward. "Grading: A Study in Semantics." *Philosophy of Science* 11, no. 2 (April 1944): 93–116.
- Sassoon, Galit. "A Typology of Multidimensional Adjectives." *Journal of Semantics* 30, no. 3 (August 2013): 335–80.
- Sauerland, Uli, and Penka Stateva. "Scalar vs. Epistemic Vagueness: Evidence from Approximators." *Proceedings of SALT* 17 (2007): 228–45.
- . "Two Types of Vagueness." In *Vagueness and Language Use*, edited by Paul Égré and Nathan Klinedinst, 121–45. London: Palgrave Macmillan, 2011.
- Sedivy, Julie C., Michael K. Tanenhaus, Craig G. Chambers, and Gregory N. Carlson. "Achieving Incremental Semantic Interpretation through Contextual Representations." *Cognition* 71, no. 2 (June 1999): 109–47.
- Stojanovic, Isidora. "An Empirical Approach to Aesthetic Adjectives." In *Advances in Experimental Philosophy of Aesthetics*, edited by Florian Cova and Sébastien Réhault, 221–40. London: Bloomsbury Academic, 2018.
- Swanton, Christine. *Virtue Ethics: A Pluralistic View*. Oxford: Oxford University Press, 2003.
- . "Virtue Ethics and the Problem of Demandingness." In *The Problem of Moral Demandingness: New Philosophical Essays*, edited by Timothy Chappel, 104–22. New York: Palgrave Macmillan, 2009.
- Syrett, Kristen, Evan Bradley, Christopher Kennedy, and Jeffrey Lidz. "Shifting Standards: Children's Understanding of Gradable Adjectives." *Proceedings of the Inaugural Conference on Generative Approaches to Language Acquisition* (2006): 353–64.
- Syrett, Kristen, Christopher Kennedy, and Jeffrey Lidz. "Meaning and Context in Children's Understanding of Gradable Adjectives." *Journal of Semantics* 27, no. 1 (February 2010): 1–35.
- Tessman, Lisa. *Burdened Virtues: Virtue Ethics for Liberatory Struggles*. Oxford: Oxford University Press, 2005.
- Toledo, Assaf, and Galit Sassoon. "Absolute vs. Relative Gradable Adjectives:

- Variance Within vs. Between Individuals.” *Proceedings of SALT 21* (2011): 135–54.
- Unger, Peter. *Ignorance: A Case for Scepticism*. Oxford: Clarendon Press, 1975.
- Von Stechow, Arnim. “Comparing Semantic Theories of Comparison.” *Journal of Semantics* 3, no. 1–2 (1984): 1–77.
- Wheeler, Samuel. “Attributives and Their Modifiers.” *Noûs* 6, no. 4 (November 1972): 310–34.
- Yoon, Youngeun. “Total and Partial Predicates and the Weak and Strong Interpretations.” *Natural Language Semantics* 4, no. 3 (January 1996): 217–36.

## DISABILITY AS INABILITY

*Alex Gregory*

IF WE WERE TO WRITE DOWN all those things that we ordinarily categorize as disabilities, the resulting list might appear to be extremely heterogeneous. Is there really anything in common between motor neuron disease, diabetes, achondroplasia, deafness, a missing lower arm, and cerebral palsy? This apparent heterogeneity gives rise to two related puzzles.

First, if there is nothing all disabilities have in common, this suggests that the concept of disability is not useful when thinking about justice. But this flies wildly against much current practice. For example, many legal systems make reference to disabilities (e.g., non-discrimination law), and many disabled activists and charities perceive themselves as fighting together for a common cause. The concept of disability also figures in various moral concerns that we have, such as when we aim to ensure that disabled people are adequately represented in government and the workplace, or when we worry whether our healthcare policies are fair toward disabled people. Common practices like these seem to presuppose that disabilities are, in some relevant respect, similar.<sup>1</sup>

The second puzzle is that the apparent heterogeneity of disabilities can make it difficult to categorize new or uncertain cases. Should we think of obesity, or disfigurement, or unusual great height, or nut allergies, as disabilities? If we cannot see what disabilities have in common to begin with, it is hard to know how to approach such questions. This issue clearly relates to the first: when we ask whether to count nut allergies as disabilities, part of our interest is surely about the requirements of justice with respect to people with such allergies. But again, without understanding why we sometimes classify disabilities together when

1 One might suggest that we should reject the idea that disabilities must have something in common in order to be relevant from the point of view of justice: disability might be a useful category for thinking about justice only because disabled people—despite having nothing “objective” in common—are often oppressed in similar ways. But to say this is not to reject the reasoning in the text, but instead to endorse it by suggesting a theory of what disabilities have in common: they are oppressed bodily states. I address views of this kind in sections 5 and 6.

thinking about justice in the first place, it is unclear how we should approach questions about what else to put into the category.

In the face of these puzzles, we need a theory of what disabilities have in common. Such a theory promises to clarify and vindicate the thought that disabilities are in some respects relevantly similar from the point of view of justice, and would enable us to think more systematically about how to treat uncertain cases such as nut allergies. In this paper I shall defend the view that a particular kind of inability is just what all disabilities have in common. Some might find this conclusion unsurprising, but the theory has some recent detractors and few detailed philosophical defenses. It is beyond the scope of this paper to defend this theory and then *also* to discuss the practical issues above in the full depth they deserve. But nonetheless, this theory at least puts us in a position to think about those issues more adequately.

Below, in section 1, I provide an attractive version of the inability theory. Then, in sections 2 and 3, I show how the theory can avoid two central objections. Then, in sections 4 and 5, I show that the theory compares favorably with theories that define disability in terms of well-being, and with theories that define disability in terms of social attitudes. Finally, in section 6, I show that the inability theory is consistent with the social model of disability.

But before we get to all that, I should first clarify our overall topic. When we offer a theory of what it is to be disabled, we might reasonably give different answers depending on exactly what it is that we are seeking. The main thing to say is that I am not seeking a definition of disability that secures some particular *predetermined* political conclusion.<sup>2</sup> As I see things, to do this would be to put the cart before the horse. Only by first understanding what different disabilities have in common can we become clearer about what justice requires with respect to disabled people. Our obligations toward disabled people *depend on* facts about what such lives are like, and as such we need some understanding of the latter issue before we can settle the former.

To this extent, I aim to provide an account of disability that picks out something real in the world—to offer a theory that carves nature at the joints. Such a descriptive theory would show what various disabilities have in common and explain what distinguishes them from other superficially similar phenomena. But although such a theory is descriptive rather than normative, it is nonetheless normatively important insofar as clarifying what disabilities have in common will illuminate the features they have that are relevant for justice.

2 For some related discussion, see Barnes, *The Minority Body*, 39–42; Haslanger, “Gender and Race,” 32–35; Howard and Aas, “On Valuing Impairment,” 1115–16; Vehmas, “Philosophy and Science.”

I said above that I would pursue a theory of disability that carves nature at the joints. In fact, we need a theory that carves nature at the joints *and* whose content is tolerably close to our everyday concept of disability so that we can recognize the theory as a more careful and complete version of that everyday concept. For this reason, I should further clarify that I shall assume that a theory of disability is not the same as a theory of unhealth (neither “disease” nor “illness”).<sup>3</sup> In ordinary thought and talk, having chicken pox, a cold, or a graze do not constitute disabilities. So I shall assume that a theory of disability is not simply the negated expression of a theory of health.<sup>4</sup>

This allows us to immediately set aside bolder versions of the *medical model* of disability, according to which disabilities are just particular kinds of bodily malfunction. Even if we grant that disabilities involve bodily malfunction, we would need to supplement this view in order to explain what distinguishes disabilities from other bodily malfunctions such as chicken pox or a cold, and in so doing we would end up resorting to one of the views below. For this reason, I shall not address the medical model as a theory in its own right.

An objector might reply that we can understand disabilities merely as *permanent* bodily malfunctions, in contrast to other temporary kinds of unhealth.<sup>5</sup> But some kinds of bodily malfunction that are not commonly thought of as disabilities—such as cancer—can be perfectly permanent. Vice versa, and more decisively, some disabilities could be temporary: perhaps there is some virus that paralyzes your legs for a few years, but will eventually work its way out of your system. Plausibly, you are disabled during the intervening period. So this objection fails, and the point stands that a theory of disability is not simply a theory of unhealth.

One final clarification: for ease, throughout this paper I will focus just on physical disabilities. Psychological disabilities (of whatever kind: cognitive, intellectual, emotional, etc.) raise distinctive issues of their own—such as about the difference between a lack of ability and a lack of will—and I cannot hope to

- 3 For one influential theory, and the distinction between disease and illness, see Boorse, “On the Distinction between Disease and Illness,” “Health as a Theoretical Concept,” and “A Rebuttal on Health.”
- 4 Moreover, since too much hangs on exactly how we understand “health,” I shall leave open whether we should think of disability as a particular kind of poor health or instead as something altogether distinct from poor health. Of course, even in the latter case, *some* conditions might coincidentally be both disabilities and health conditions. For some related discussion, see Aas, “Disabled—Therefore, Unhealthy?”; Wendell, *The Rejected Body*; Hull, “Defining Disability,” 200; and Silvers, “On the Possibility and Desirability of Constructing a Neutral Conception of Disability,” 478.
- 5 For example, Amundsen, “Against Normal Function,” 34–35.

resolve all such issues here.<sup>6</sup> But I hope it will be clear that the inability theory at least *promises* to be extendable as a theory of such disabilities, even if the details will have to wait for another day.

### 1. THE INABILITY THEORY

With our focus clarified, I shall now defend the *inability theory*. The label indicates the basic idea, which is that disabilities are particular kinds of inability. To specify the view slightly more precisely, I take the inability theory to be the following view:

*The Inability Theory:* To be disabled is to be less able to do something than is typical, where this degree of inability is partly explained by features of your body that are atypical.<sup>7</sup>

I began by asking what there is in common between motor neuron disease, diabetes, achondroplasia, deafness, a missing lower arm, and cerebral palsy. The inability theory gives us our answer: these things are alike in that they each involve (roughly) some restriction of ability. That is the most natural explanation of why we categorize these things together. And this seems to capture what, in general, all disabilities have in common: some degree of inability in some particular domain. It is very difficult to think of any candidate disability that does not involve any restriction of ability in any domain. To this extent, the inability theory looks like a highly attractive theory of the nature of disability, one that makes clear what it is that apparently heterogeneous disabilities nonetheless have in common. The view also understands disability in a manner that explains its practical significance: disabilities are politically relevant just because the range of people's abilities is politically relevant. Given just how natural the inability theory is, it should be our default view about disability unless we are given good reasons to think it cannot be sustained.

The inability theory promises to help us solve our original puzzles. It promises to clarify why the category of disability is useful for thinking about the requirements of justice. The inability theory tells us that when thinking about the

6 For example, I take it that a theory of disability should tell us that homosexuality is not a disability. To ensure that the inability theory gets that result, we would need to show, for example, that homosexual people are in the relevant sense "able" to conceive children through intercourse. I take it that such a reading of "able" is available, but the issues raised are beyond the scope of this paper.

7 Cf. Amundsen, "Disability, Handicap, and the Environment"; Buchanan, Brock, Daniels, and Wikler, *From Chance to Choice*, 286–87; Hull, "Defining Disability"; see also the UK's Equality Act 2010, s6(1); Americans with Disabilities Act of 1990, sec. 12102.

requirements of justice with respect to disabled people, we are really thinking about the requirements of justice with respect to people who lack certain typical abilities: we are asking about our obligations to equalize those levels of ability, and about our obligations insofar as such equalization is not feasible.<sup>8</sup> Inequalities in ability are clearly relevant from the point of view of justice, and to this extent the inability theory promises to do exactly the job needed: to vindicate, and clarify, our obligations toward disabled people.<sup>9</sup> The inability theory also promises to help us decide how to classify uncertain cases such as nut allergies, by focusing our attention on whether such conditions can limit ability in the relevant manner.

I shall now clarify four features of the inability theory. First, the inability theory makes reference to our *abilities*. I am thinking of abilities as capacities of a person to intentionally act in certain ways. So if your body is less able to manufacture some chemical, that is not itself an inability of the relevant kind, though it might produce a disability if it also stops *you* doing things.<sup>10</sup> One might further restrict the view so that an ability is a capacity to perform some relatively coarse-grained *kind* of action, such as playing the piano. In contrast, the view as I develop it permits that capacities to perform relatively fine-grained actions, such as playing a piano *with your fingers*, count as abilities, and in turn that the absence of such capacities might constitute disabilities. The upside of developing the view in this permissive manner is that it frees us from having to develop some threshold of coarseness of abilities. The main downside is that it entails that we have many disabilities insofar as there are bound to be more abilities you lack, when they are understood in this fine-grained manner. But this worry adds little to existing worries that I address in section 2 about minor disabilities.

Second, the inability theory says that disability consists in being *less able* to do something than the typical person.<sup>11</sup> Such degrees of ability are defined in terms of the difficulty or time involved in performing the relevant action. I shall treat a total inability to do something as the limiting case of being “less able” to do something, and so one can be disabled by lacking an ability altogether as well

8 Cf. Daniels, *Just Health Care*.

9 Cf. Schroeder, “Well-Being, Opportunity, and Selecting for Disability.” For one popular perspective on the relevance of abilities for justice, see the capabilities approach in Nussbaum, *Women and Human Development*; Sen, “Equality of What?” and “Capability and Well-Being.” For the capabilities approach as applied to disability, see Nussbaum, *Frontiers of Justice*; Terzi, “Vagaries of the Natural Lottery?”

10 Cf. Amundsen, “Disability, Handicap, and the Environment,” 107–8.

11 To this extent, the view avoids those of Elizabeth Barnes’s objections (*The Minority Body*, 16–17, 20) that treat the view as saying that disabilities consist in *total* inability rather than reduced levels of ability.

as by having the ability to some reduced degree. One might worry that some disabilities—such as those that involve general fatigue—might not prevent you from doing any particular thing.<sup>12</sup> But this is not right: such disabilities might prevent you from, say, working a forty-two-hour week. So long as we remember that the things people can typically do include extended actions, this objection fails: disabilities that involve general fatigue or fluctuations in energy levels plausibly do result in lower levels of ability at various temporally extended tasks.<sup>13</sup>

Third, the inability theory makes reference to what is *typical*. Here “typical” means “statistically most frequent.” We might instead formulate the view with reference to *normality*, which is a partially evaluative notion. I will not dwell on this choice: everything I say in defense of the inability theory could be redeployed to defend the theory defined in that alternative way. I phrase it in terms of typicality because I find that concept clearer and less dubious than the concept of normality. But if you disagree, you should feel free to rephrase my claims accordingly.

The inability theory actually makes *double* reference to typicality: it makes reference to what the typical level of ability is, and it also makes reference to what a typical body is like. Both qualifications are needed. If you have typical levels of ability but an atypical body (e.g., you have ginger hair), you are not thereby disabled. Vice versa, if you have atypical levels of ability but a typical body (e.g., you are in jail), you are also not disabled. Disability requires an atypically low level of ability in virtue of a bodily atypicality.

“Typical” here means “typical for a human being of your sex at your stage of development.” Humans are not disabled in virtue of lacking the ability to fly; men are not disabled in virtue of lacking the ability to breastfeed; and infants are not disabled in virtue of lacking the ability to talk.<sup>14</sup> (In principle, we might relativize further, such as to race. But it is hard to find intuitive cases that support further restrictions like this.) Note that I refer to a stage of development—understood biologically—and not merely age. I take it that eighty-year-old humans are at the same stage of development as thirty-year-old humans—adults—and

12 See Barnes, *The Minority Body*, 17.

13 In note 4 I allowed that some conditions might be both diseases *and* disabilities. The reasoning here favors that thought: some diseases can fatigue us and prevent our completing various extended tasks. This does not undermine the earlier distinction between disability and disease: early-stage cancer, which is not yet impeding your ability to do anything, is a disease but not a disability.

14 Note that I relativize disabilities to what is typical for human beings, not what is typical for human beings *at the present time*. So if some strange event rendered most people deaf, deafness would continue to be a disability: it would continue to be atypical for human beings in general, even if not atypical for human beings then alive.



to that extent both groups are classified as disabled by the same standards. An eighty-year-old with inabilities that are typical for someone their age may nonetheless be disabled if those inabilities are not typical for human adults in general. In turn, the theory rightly entails that many elderly people are disabled.

Further, I take it that typicality requires that your levels of ability, and bodily features, fall within some particular *range*. To be typical, your body need not be *perfectly* typical. Thinking of typicality in this manner is natural and also allows the theory to avoid one of Elizabeth Barnes's main objections to the view. Barnes worries that the inability theory is likely to classify being a "petite woman" as a disability, since petite women have bodies that diminish their abilities in some dimensions.<sup>15</sup> But while petite women might have somewhat atypical bodies and thereby lack some relatively typical abilities, their bodies and ranges of ability are not *that* atypical. People who are *very* atypically short and who thereby lack *very* typical abilities (e.g., to drive a car) might count as disabled, but plausibly merely "petite" women fail to count as disabled precisely because their range of ability falls within the normal range. Petite women are relatively common, at least compared with, for example, people with achondroplasia.

Of course, it will be very hard to draw a sharp line between bodies and abilities that are within the typical range and those that are not: for many disabilities, there is a continuum between them and merely "ordinary" levels of inability. For example, there is a continuum between a nondisabled person with minor problems with their vision and a disabled person who can barely see at all. There is bound to be no precise point along this continuum where one qualifies as disabled, just as there is no precise number of grains of sand that comprises a heap. But this is not a problem for the inability theory: it shows only that the category of disability has some vague boundaries, as seems independently plausible.

The fourth and final feature of the inability theory that I should clarify is that the theory says that to be disabled you must have atypical levels of ability that are *partly explained* by features of your body that are atypical. Perhaps you (atypically) cannot go outside. And perhaps your body is atypical. But these facts by themselves do not make you disabled: perhaps your abilities are atypical because you are in jail, not at all because of your atypical body. To be disabled, there must be some explanatory connection between the atypical features of your body and the fact that your abilities are atypical.

To this extent, the inability theory claims that disabilities involve inabilities partly explained by your body. But the word "partly" here is crucial. The atypical features of your body need not be the *full* explanation of why you are less able than others at doing some things. For example, perhaps you cannot read

15 Barnes, *The Minority Body*, 17, 20.

*War and Peace*, and this is atypical, and partly explained by your atypical body: you are blind. But your blindness need not be the complete explanation of your inability to read *War and Peace*: perhaps your inability is also partly explained by the unavailability of the book in braille. In this case, we can modify your environment—make the book available to you in braille—and thereby raise your level of ability in this context to the typical level. By doing so, we stop you being disabled in this context, even though we have not changed your body.<sup>16</sup>

That is, the inability theory defines disability in terms of inability, where your range of abilities depends *both* on your body and on your environment.<sup>17</sup> In some environments bodily differences make no difference, and in other environments those same bodily differences do make a difference. In this way, the theory captures both the attractive thought that whether you are physically disabled depends on your body, but also captures the attractive thought that disability can be heavily influenced by the structure of your environment. The inability theory thereby permits that we should sometimes, or often, address disability by changing environments rather than people. We might reduce levels of disability by making environments more accessible, as well as by making changes to people, such as prosthetics. (I return to some related issues in section 6.)

Having described and clarified the inability theory, I now turn to two objections. Both claim that the inability theory is extensionally inadequate in certain respects, but I shall show how to revise the theory to avoid these implications. I cannot rule out the possibility that the inability theory has some revisionary consequences in some cases for what counts as a disability, but I take those revisionary instances to be rarer and less important than objectors suppose.

## 2. OBJECTION: MINOR INABILITY

In recent works, Elizabeth Barnes and Guy Kahane and Julian Savulescu have objected to the inability theory with similar examples. Barnes writes: “[the view that] to be disabled is to lack a physical ability which most people have [is] a non-starter. Most people can ‘roll’ their tongues. To not be able to roll your tongue is to lack an ability that most people have. And yet those of us who can’t

16 Of course, merely making this book available in braille is unlikely to stop you being disabled, since you may continue to lack other abilities. But the underlying point stands that disabilities can be minimized and in principle eliminated through changes to our environment.

17 Cf. Hull, “Defining Disability,” 206–7; Nussbaum, *Women and Human Development*, 84–85; Shakespeare, *Disability Rights and Wrongs*, 31, 56.

roll our tongues are not disabled in virtue of that lack of ability.”<sup>18</sup> Kahane and Savulescu similarly worry about the ability to wiggle one’s ears.<sup>19</sup>

There are two possible ways that we might respond to this objection. The first and simplest would be to qualify the theory so that only sufficiently *life-changing* kinds of inability amount to disabilities.<sup>20</sup> I am in principle open to this broad way of developing the inability theory: it is obviously the simplest way to respond to this objection. But there is a lingering worry that this way of developing the inability theory draws an arbitrary line: Why think that whether something is a disability depends in this way upon its impact? A neater view would say that disabilities can have a larger or smaller effect, and then explain away our tendency to *say* that uninfluential disabilities are not disabilities at all.

Imagine that we are heading for a beer, and I ask you whether you brought any money with you. If you answer in the affirmative, thinking of the £0.10 in your pocket, your answer is true, but nonetheless highly misleading, because it implies that you can afford to buy us beer. For that reason your claim is conversationally inappropriate even though it is strictly true. Or imagine that I ask you whether you have any health problems. If you have a hangnail, that is a health problem, but it would nonetheless be misleading to say that you have a health problem if it is one quite so trivial. Examples like these serve to illustrate Grice’s maxim of relation: your assertions ought to be conversationally relevant.<sup>21</sup> In order to avoid saying things that are misleading, we might sometimes avoid asserting certain truths.

Since this maxim is supposed to be perfectly general, we can also apply it to our talk of disability. Perhaps there are some conditions that are disabilities, but that are not appropriately described as such. Very minor disabilities plausibly fall into this category. After all, when we ask about someone’s disability status we are often interested only in disabilities that are relevant for our specific practical purposes, and very often uninfluential disabilities will not be relevant for those purposes. For this reason, we might think that if you are unable to roll your tongue, or wiggle your ears, that *is* a disability. But saying this is consistent with thinking that we should not *describe* such disabilities as disabilities, since they are so insignificant and of so little practical relevance. In this respect, such minor disabilities are akin to a health problem like a hangnail: real, but insignificantly tiny.

In short, when we remember independently plausible claims about conver-

18 Barnes, *The Minority Body*, 16.

19 Kahane and Savulescu, “The Welfarist Account of Disability,” 45.

20 Cf. Glover, *Choosing Children*, 9; Equality Act 2010, s6(1); Americans with Disabilities Act of 1990, sec. 12102.

21 Grice, “Logic and Conversation.”

sational pragmatics, we can see that the inability theory is perfectly consistent with the thought that it is inappropriate to describe minor inability like those above as disabilities. To that extent the inability theory remains unscathed by such supposed counterexamples.

Before we move on, there is an objection to my proposal I should consider. Imagine that you acquire some disability that is, in one sense, comparatively minor—one day you cease to be able to roll your tongue, say. But imagine further that for very idiosyncratic reasons, this makes a big difference to your life: perhaps you are famed for your unusual whistling, and your unusual whistling relied on your ability to roll your tongue. Does the inability theory classify this as a disability? It does. But it might seem that the above Gricean maneuver fails to apply in this case: given your circumstances, this inability *is* worth mentioning and *does* have practical relevance. One possibility here is to simply accept that under these circumstances, your inability to roll your tongue could appropriately be called a disability. But more plausibly, the same Gricean maneuver does explanatory work even in this case. Although this inability is clearly worth mentioning, it is nonetheless misleading to highlight its status as a *disability*. This kind of case is odd precisely because it is a case where something is highly practically significant despite being only minor *qua* disability: it is a minor restriction *on ability* with surprisingly wide-reaching consequences. So although this inability is worth mentioning, it would be still be misleading to describe it *as a disability* given that its significance is not due to its being a significant disability.

### 3. OBJECTION: IMPOSED INABILITY

I now turn to a second objection to the inability theory. To illustrate the problem, imagine that you are a member of a small minority race, and are a victim of direct discrimination on the basis of your race. This racism might reduce the options you have. To that extent, you might be unable to do certain things, where this inability is partly explained by the atypical features of your body. So it seems as though our theory classifies you as disabled. But plausibly, under these circumstances, you are not necessarily disabled.<sup>22</sup> Similar remarks would apply in many other cases where a bodily minority is constrained by social prejudice.<sup>23</sup>

22 For some discussion, see Barnes, *The Minority Body*, 18–19; Kahane and Savulescu, “The Welfarist Account of Disability,” 34–35.

23 Note that since the inability theory only classifies you as disabled if your body explains your lack of ability, the theory already generates the right results with respect to victims of *indirect* discrimination. Victims of indirect discrimination may lack certain abilities, but if the discrimination is truly *indirect* that inability is explained not by their body but instead by some other property they have that happens to correlate with having a body of that kind.

The best response to this is to qualify the inability theory. We are thinking of disabilities as arising from the way your body combines with your environment. But in the case above, the relevant inability arises not merely from the way your body interacts with your *physical* environment, but in part from the way your body interacts with other people's attitudes. So to fix the problem, we might modify the inability theory as follows:

*The Final Inability Theory:* To be disabled is to be less able to do something than is typical, where this degree of inability (1) is partly explained by features of your body that are atypical, and (2) is not explained by anyone's attitudes toward those bodily features.

Here clause 2 serves to explain why the above kind of inability is not a disability. Though your inabilities are explained by atypical features of your body, your inabilities are explained by the attitudes certain people have toward those atypical features of your body. In other cases of disability, the relevant inabilities are not so explained: it is not only because of how people *think* about paraplegic people that there are some constraints on what such people are able to do.<sup>24</sup>

Some might worry that this qualification rules out too much: Isn't the precise point of the social model of disability that the disadvantages faced by disabled people arise from *social* oppression?<sup>25</sup> I shall discuss the social model in section 6. But here there is one key point to make: the social model may say that the disadvantages faced by disabled people arise from society, and are unjust. But to make these claims, we do not need to also say that the disadvantages faced by disabled people arise from attitudes directed toward disabled people's bodies. It is no doubt true that many of the disadvantages faced by disabled people *do* arise in this way. But when the social model seems persuasive, this is precisely because it does not attribute *all* the disadvantages faced by disabled people to explicit discrimination. Rather, it also focuses on unjust *absences* of attention to access. Such failures need not involve anyone disadvantaging disabled people because of their attitudes toward disabled people. In short, we should distinguish between injustice that arises through explicit discrimination, and injustice that arises through a lack of attention. By defining disability as I have above, we must say that *some* of the constraints on disabled people arise independently of explicit discrimination. That is why the restrictions on disabled people are different than, for example, the restrictions on minority races in racist societies. But saying this is consistent with thinking that many constraints on disabled people

24 Cf. Siebers, *Disability Theory*, 134.

25 Barnes, *The Minority Body*, 19–20.

arise through unjust social choices. So qualifying the inability theory as I do above does not prejudice the issues surrounding the social model.

What should we say in cases where someone's inability is explained by *both* explicit discrimination *and* some absence of attention? The inability theory as stated above says that to the extent their inability is *not* explained by explicit discrimination, it is a disability. To this extent, some disabilities may be constituted in part by disabilities and in part not. For example, imagine that city planners gave very little thought to accessibility for disabled people, but that they did so in part because of discriminatory attitudes (perhaps they thought disabled people just "belong" at home). Under these circumstances, your inability to get to work might be both a disability you have (to the extent that your inability is not wholly explained by their prejudice) but also partially not (to the extent that your inability is partly explained by their prejudice). In some real-world cases, disentangling these factors may be extremely difficult.<sup>26</sup>

This completes my defense of the inability theory. I now turn to compare it with rival theories of disability.

#### 4. THE WELL-BEING THEORY

The well-being theory of disability says, roughly, that a disability is a condition that harms you. The view has received its clearest formulation in the work of Guy Kahane and Julian Savulescu.<sup>27</sup> In its simplest form, the theory says:

*The Well-Being Theory:* To be disabled is to have a property that harms you.

But this theory is inferior to the inability theory: it faces two key objections. First, this theory says that if something is harmless, it is not a disability. This seems controversial at best: think of apparent disabilities such as achondroplasia and deafness, or what we might think of as minor disabilities, such as missing toes. We *might* think that these conditions are on balance harmless and are yet disabilities.<sup>28</sup> In other cases *even if* the well-being theory gets the right answer, it

26 In some cases, what we should say will hinge on wider questions about what counts as an explanation. Imagine that city planners designed the city this way because they explicitly discriminated, but that had they been fairer minded, they would have reached the same conclusion anyway, for independent reasons. Under these circumstances, it is unclear whether the relevant inability is "explained" by discrimination: it was caused by it, but would have existed regardless. I take this to be a general question about the nature of explanation, not a puzzle specifically for the inability theory.

27 Kahane and Savulescu, "The Welfarist Account of Disability"; Savulescu and Kahane, "Disability"; see also Harris, "One Principle and Three Fallacies of Disability Studies."

28 For example, on deafness, see Lane, "Do Deaf People Have a Disability?"

seems to reach it on the wrong grounds: Do we really have to survey paraplegic people on how happy they are in order to gain evidence that paraplegia is a disability?<sup>29</sup> The well-being theory makes it a difficult empirical question whether paradigm disabilities are in fact disabilities, and this seems surprising at best.<sup>30</sup>

The second problem for the well-being theory is that not all harmful conditions are disabilities. Think, for example, of breaking up with your partner or finding a hair in your sandwich. These things might be bad for you, but are quite clearly not disabilities. I take it that cases like these explain why Kahane and Savulescu claim that only *intrinsic* harmful properties constitute disabilities: this restriction serves to exclude external harms like these from constituting disabilities.<sup>31</sup> But this does not solve the underlying problem: there are plenty of intrinsic harmful properties that are not disabilities, such as grumpiness or a mild hangover. Kahane and Savulescu still further qualify the theory so that disabilities are only constituted by *stable* intrinsic harmful properties.<sup>32</sup> But this is wrong in both directions: your grumpy personality might be permanent, and your disabling paralysis might be temporary.

Throughout their papers, Kahane and Savulescu positively admit that their theory has counterintuitive implications like these.<sup>33</sup> Kahane and Savulescu think that we should focus on finding a theory of disability that does interesting ethical work, even if we thereby end up with a theory that has some surprising implications about which things count as disabilities. But there is no need to choose between these two goals: the inability theory shows that the concept of disability is one that does interesting ethical work without the need for radically revising standard disability classifications.

The amount of freedom we have is clearly a topic of central political importance, and our range of abilities is clearly relevant for the amount of freedom we have. In this way, the inability theory understands disability in a manner that

29 Do I here assume some hedonistic theory of well-being? No. Even if happiness is not the only contributor to well-being, it is surely *a* contributor, and that is enough to make it an open empirical question whether paraplegic people might on balance be living lives that are relatively high in well-being because they are happy.

30 Barnes, *The Minority Body*, 11. For interesting data on the happiness of disabled people, see Albrecht and Devlieger, "The Disability Paradox"; Boyd et al., "Whose Utilities for Decision Analysis?"; Daniels, Rose, and Daniels Zide, "Disability, Adaptation, and Inclusion"; Sackett and Torrance, "The Utility of Different Health States as Perceived by the General Public"; Ubel, Loewenstein, and Jepson, "Whose Quality of Life?"; Ubel and Loewenstein, "Pain and Suffering Awards."

31 Kahane and Savulescu, "The Welfarist Account of Disability," 24–25.

32 Kahane and Savulescu, "The Welfarist Account of Disability," 24–25.

33 Kahane and Savulescu, "The Welfarist Account of Disability," 25–33.

connects it to a topic of central political importance. Investigating the exact practical implications of disability, as the inability theory understands it, would take us far afield, but we can at least note the obvious points to be explored. (i) The inability theory may suggest that disabilities are *intrinsically* bad if we think there is something good about freedom itself. To assess this, we would need to ask whether there really is intrinsic value to having more abilities.<sup>34</sup> (ii) The inability theory may suggest that disabilities are *instrumentally* bad insofar as there are often instrumental advantages to having more abilities. To assess this, we would need to ask whether there are always instrumental benefits to having more abilities.<sup>35</sup> Moreover, we would need to ask whether there are sometimes instrumental *benefits* to having fewer abilities.<sup>36</sup>

In short, the inability theory captures the thought that disabilities are highly ethically significant, though it inevitably leaves us with theoretical work to do. To that extent, Kahane and Savulescu are wrong when they imply that a theory of disability can capture the ethical significance of disability only by radically revising our ordinary concept. The inability theory captures the ethical significance of disability while nonetheless remaining tolerably close to our everyday conception of disability. In contrast, the well-being theory might capture something of ethical relevance, but only by radically revising ordinary classifications of disability. *Pace* Kahane and Savulescu, this is a theoretical cost we do not need to pay.

##### 5. CONVENTIONALIST THEORIES

I now turn to a second rival theory of disability, which I label “conventionalist.” Conventionalist theories claim that disabilities are not intrinsic states of people but instead socially constructed properties that are generated by the attitudes of some person or people.<sup>37</sup> I address two such theories.

Our first conventionalist theory says that you are disabled if your society categorizes you as disabled. Call this:

34 See, e.g., Dworkin, *The Theory and Practice of Autonomy*, 62–81, esp. 80.

35 Goering, “You Say You’re Happy, but . . .,” 130; Moller, “Wealth, Disability, and Happiness,” 198–99; Schroeder, “Health, Disability, and Well-Being,” 224–25; Silvers, “On the Possibility and Desirability of Constructing a Neutral Conception of Disability,” 479–80; Wasserman and Asch, “Understanding the Relationship between Disability and Well-Being,” 149–52.

36 Schwartz, “Self-Determination” and *The Paradox of Choice*; Silvers, “On the Possibility and Desirability of Constructing a Neutral Conception of Disability,” 482.

37 Cf. Oliver who claims that “disablement is nothing to do with the body” and that “disability is wholly and exclusively social” (*Understanding Disability*, 35).



*Simple Conventionalism*: To be disabled is to be categorized as disabled by society.

Simple conventionalism is unsustainable.<sup>38</sup> Whether you are disabled does not depend on the attitudes of society, as illustrated by the possibility of a sailor who is shipwrecked and washes up ashore alone on a desert island, paralyzed from the waist down, with no one aware of her disability. Vice versa, imagine that a celebrity becomes a recluse and hides away without any human contact. A rumor goes around to the effect that they have become a recluse in virtue of a disability. Does that rumor suffice to make it true that they are disabled? No. Further investigation might, under these circumstances, show the rumor to be wrong, and this would not be possible if the existence of the rumor sufficed for its own truth.<sup>39</sup>

The point is obviously not that this kind of error is common, but instead that there is no reason to think that society is infallible at identifying disability.<sup>40</sup> We look in the wrong place if we try to determine whether someone is disabled by looking at people's attitudes toward them. Our attitudes toward people aim to be responsive to facts about how those people really are, and our attitudes can be mistaken precisely because the facts are independent of what we think.<sup>41</sup> Equally, if we come across some difficult case where it is unclear whether to classify something as a disability (e.g., deafness), it would be a mistake to try to resolve that uncertainty by surveying what people think of that condition: their views on such matters are formed only in light of other fallible views they have about the nature of the condition and the best way to think of disability.<sup>42</sup> In short, simple conventionalism is implausible for the reason that society can make mistakes about who is disabled.

In her recent book, Elizabeth Barnes objects to a theory along similar lines, but goes on to suggest a nearby theory as a superior replacement.<sup>43</sup> On her theory—our second conventionalist theory—whether you are disabled *is* socially

38 See also Howard and Aas, "On Valuing Impairment," 1123–24.

39 Similar counterexamples would undermine the nearby theory that classifies people as disabled according to whether *they* themselves think they are disabled. Some people might not know they are disabled: think of small children, or people with anosognosia, such as those with Anton-Babinski syndrome. Vice versa, someone might think that they are disabled when they are not (perhaps if they have succumbed to an extreme practical joke).

40 Cf. Stone, "Reactions to Invisible Disability"; Wendell, *The Rejected Body*, 25–26.

41 Bury, "Defining and Researching Disability," 30; Howard and Aas, "On Valuing Impairment," 1123; Shakespeare, *Disability Rights and Wrongs*, 76–77.

42 Cf. Shakespeare, Thompson, and Wright, "No Laughing Matter," 3.

43 Barnes, *The Minority Body*, 28–38.

constructed, but not quite so directly as simple conventionalism suggests. She writes:

A person, *S*, is physically disabled in a context *C*, iff:

- (i) *S* is in some bodily state *x*
- (ii) The rules for making judgements about solidarity employed by the disability rights movement classify *x* in context *C* as among the physical conditions that they are seeking to promote justice for.<sup>44</sup>

The idea is that whether you are disabled *does* depend on the attitudes of one *part* of society, namely, the disability rights movement. But their attitudes alone do not determine whether you are disabled: what matters is whether you *in fact* have the general kind of body that they aim to represent. To this extent, Barnes's theory allows for some mistakes about disability: perhaps you *do* have the kind of body that the disability rights movement aims to represent, but perhaps no-one knows that this is true. Vice versa, perhaps you do not in fact have the kind of body that the disability rights movement represents, but people nonetheless treat you as disabled. The view also makes room for a further kind of mistake. Barnes's theory makes reference to the *rules* that the disability rights movement uses to "make judgements about solidarity"—presumably, the rules they implicitly use to classify things as disabilities. And "this allows that the disability rights movement can be mistaken—even systematically mistaken—about whether a given physical condition is a disability. And that is because they can deploy their own rules incorrectly."<sup>45</sup>

But though it is a vast improvement on simple conventionalism, Barnes's theory nonetheless fails for similar reasons. Just as the disability rights movement might make mistakes about which bodies are disabled, and just as the disability rights movement might make mistakes in deploying their rules as to which kinds of bodies count as disabled, they can also be mistaken in the rules they use for classifying people as disabled or not. In turn, they might make mistakes about which conditions they are seeking to promote justice for.

One way to see this is to note that Barnes endorses the theory above while criticizing various rival theories (such as the inability theory) on the grounds that they have implausible implications.<sup>46</sup> But such criticism is beside the point if Barnes's own theory is correct. If Barnes's own theory is correct, the only way to evaluate any other theory of disability is simply to ask whether it is the one endorsed by the disability rights movement. If *they* endorse the inability theory

44 Barnes, *The Minority Body*, 46.

45 Barnes, *The Minority Body*, 46–47.

46 Barnes, *The Minority Body*, 13–38.

as the way to decide whether someone is disabled or not, then that theory must be correct, regardless of what implausible implications it might have. So as an *ad hominem* point, Barnes is inconsistent when she subscribes to her theory but evaluates other theories on grounds that are independent of the views of the disability rights movement. And more substantively, insofar as we agree with Barnes—when she is in her critical mode—that theories of disability are evaluable independently of whether they happen to coincide with the views of the disability rights movement, then Barnes’s theory has to be mistaken.

Moreover, Barnes’s theory makes it impossible for the disability rights movement to be mistaken about the general nature of disability. But this is implausible: the movement is formed of human beings who can think in confused or ignorant ways.<sup>47</sup> Insofar as they do that collectively, they might deploy rules for classifying people as disabled that are themselves confused or that presuppose falsehoods. For example, perhaps the disability rights movement holds mistaken views about human nature that wrongly inform their views of disability. Or perhaps the movement holds views about disability that are confusedly influenced by their views about evolution. Or perhaps the movement is primarily constituted by men, or by younger people, and as a result the movement employs rules for classifying people as disabled that are wrongly insensitive to issues that primarily affect women or the elderly.<sup>48</sup>

Just as with simple conventionalism, my point is not that the disability rights movement is *actually* mistaken about the nature of disability. The people who compose that movement might well know a lot about disability, and the wisdom of crowds might help the movement as a whole overcome the mistakes of some of its individual members. But it is nonetheless true that the movement is composed of fallible individuals, who are therefore collectively fallible about the nature of disability. But on Barnes’s theory, the disability rights movement is infallible about the nature of disability. This is surely wrong, and so Barnes’s theory must be false.

I conclude that conventionalist theories like those above are mistaken, be-

47 See also Lim, “Disabilities Are Also Legitimately Medically Interesting Constraints on Legitimate Interests,” 982–87.

48 *At least* this objection to Barnes also carries across to Chong-Ming Lim’s modified version of her theory (“Disabilities Are Also Legitimately Medically Interesting Constraints on Legitimate Interests”), according to which disabilities are (roughly) states that the disability rights movement represents *and* which meet further conditions, such as that they limit your functionings and are of medical interest (987). Whatever the other merits of Lim’s view, it shares with Barnes’s the claim that the views of the disability rights community place necessary conditions on disability, and that is mistaken: again, their rules might be defective in ways that lead them to neglect conditions that ought to be recognized as disabilities.

cause they cannot make adequate room for mistakes about the nature of disability. This is one respect in which the social model can seem attractive, since it stresses the social nature of disability while nonetheless allowing for an important role for the body to play, namely in constituting whether an individual has an impairment. To this theory I now turn.

## 6. THE SOCIAL MODEL

The social model of disability was most famously laid out in the Union of the Physically Impaired Against Segregation's (UPIAS) "Fundamental Principles of Disability" document. In its most famous passage, it says:

In our view it is society which disables physically impaired people. Disability is something imposed on top of our impairments by the way we are unnecessarily isolated and excluded from full participation in society. Disabled people are therefore an oppressed group in society. To understand this it is necessary to grasp the distinction between the physical impairment and the social situation, called "disability," of people with such impairment [*sic*]. Thus we define impairment as lacking part of or all of a limb, or having a defective limb, organ or mechanism of the body; and disability as the disadvantage or restriction of activity caused by a contemporary social organisation which takes no or little account of people who have physical impairments and thus excludes them from participation in the mainstream of social activities.<sup>49</sup>

As this passage suggests, the core part of the social model of disability is the distinction between impairment and disability, where impairments are defined in terms of bodily "defects," and disabilities are the socially mediated effects of such impairments.<sup>50</sup> Morris gives some helpful illustrations: "an inability to walk is an impairment, whereas an inability to enter a building because the entrance is up a flight of steps is a disability. An inability to speak is an impairment but an inability to communicate because appropriate technical aids are not made available is a disability. An inability to move one's body is an impairment but an inability to get out of bed because appropriate physical help is not available is a disability."<sup>51</sup> As these examples make clear, it is crucial for the social model that

49 Cited in Oliver, *Understanding Disability*, 22.

50 See, e.g., Barnes, *Disabled People in Britain and Discrimination*; Finkelstein, *Attitudes and Disabled People*; Oliver, *Understanding Disability*; and cf. the distinction between sex and gender.

51 Morris, *Independent Lives?* x.

disabilities are generated not by impairment alone, but instead by the combination of impairment and environment.<sup>52</sup>

The basic ideas behind the social model are relatively clear. But pinning down the *exact* content of the theory is more difficult.<sup>53</sup> I will set aside three versions of the social model, and then focus in more depth on a fourth.

First, there is:

*The Minimalist Social Model:* Whether you are disabled depends on your environment, not just your body.

Anyone who signs up to the social model in any form presumably signs up to this minimal claim. The minimalist social model is not merely consistent with the inability theory, but in fact positively entailed by the inability theory. As I explained when introducing the inability theory, it permits that disability depends on the environment: it says that disability involves atypical levels of ability that are partly explained by your body but *also* partly explained by your environment. So if we think of the social model as the minimalist social model, then the inability theory and the social model are certainly not competitors and I subscribe to both.

I address the next two options together:

*The Causal Social Model:* (Very) often, the difficulties faced by disabled people arise from unjust choices, such as unjust absences of attention or explicit discrimination.

*The Normative Social Model:* (Very) often, we should solve the difficulties faced by disabled people by changing the environment, not their bodies.

I tend to think that these claims go beyond the minimalist social model in ways that more accurately reflect the actual commitments of adherents of the social model. But despite their interest, my arguments in this paper do not directly bear on these claims. The inability theory is a theory of what disabilities are, and not a descriptive claim about human history or a normative claim about how we should henceforth organize society. To that extent, it does not directly support or undermine the two claims above.

I will now turn to our final and most relevant interpretation of the social model:

52 Cf. World Health Organization, "International Classification of Impairments, Disabilities, and Handicaps," 28; and Disability Discrimination Act 1995.

53 Cf. Oliver, *Understanding Disability*, 41: "[The social model of disability] is not a social theory of disability and it cannot do the work of social theory."

*The Social Theory*: To be disabled is to be harmed or restricted by social response to impairment.<sup>54</sup>

I label this version of the view the social *theory* since it is designed to be a complete theory of the nature of disability that competes with the inability theory. And I take it that this theory does seem to reflect the UPIAS quote above: it maintains a distinction between impairment and disability, and says that the latter should be defined as the socially mediated consequences of the former.

If we understand the social model in this way, how should we adjudicate between it and the inability theory? To make progress with this question, let us first consider a possible objection to the social theory. Again, imagine that a sailor is shipwrecked, and washes up alone on a desert island, paralyzed from the waist down. Is she thereby disabled? Ordinary usage of the label “disabled” would suggest that she is, but according to the social theory above, she is not. Is this a problem for the social theory? I assume that defenders of the theory will say that it is not. They will say that the theory functions partly to stipulate a use for the label “disabled,” rather than to track common usage. Defenders of the social theory ought to be happy to claim that the sailor is not disabled, precisely because—as they use the label—being disabled requires that you interact with society in a certain manner.

Nonetheless, we might still want to say something about the condition that the sailor *does* have, and that they seem to have in common with someone who is in England and paralyzed in the same way. According to the social theory, the thing these individuals have in common is their impairment rather than their disability. This serves to remind us that the social theory relies heavily on the notion of impairment. It relies on that notion to capture commonsense thoughts about things people can have in common across different social contexts.

The social theory also relies on the concept of impairment in a second way: we cannot simply define disabilities as restrictions or disadvantages caused by society, since that would fail to distinguish disability from other socially stigmatized characteristics such as gender or race.<sup>55</sup> Defenders of the social model should claim that disability is distinct from other dimensions of oppression because it involves social response to *impairment* rather than to other bodily features.

In fact, defenders of the social theory are likely to require the concept of im-

54 Cf. Haslanger, “Gender and Race”; Howard and Aas, “On Valuing Impairment,” esp. 1128–29.

55 Bury, “Defining and Researching Disability,” 30; Bickenbach, *Physical Disability and Social Policy*, 176; Howard and Aas, “On Valuing Impairment,” 1118; Shakespeare, *Disability Rights and Wrongs*, 76–77.

pairment for a third purpose: in order to articulate their desires. We might want to bring about a world where disabled people do not suffer from discrimination. But if “disability” is defined so that disabled people by definition suffer from discrimination, this goal would be incoherent. Defenders of the social model can express the relevant hope in a different manner: by saying that we should bring about a world where people with *impairments* do not face discrimination.

So the social theory depends heavily on the concept of impairment. But what is an impairment? The answer to this question is not wholly clear.<sup>56</sup> We might naturally turn to the UPIAS quotation above. It defines impairment as “lacking part of or all of a limb, or having a defective limb, organ or mechanism of the body.”<sup>57</sup> But if *this* is the right way to think about impairment then, surprisingly, the social theory inherits the problems of the medical model of disability with which I began.<sup>58</sup> When we think about impairment, it would be odd to class having chicken pox, a cold, or a graze as impairments. But this is what we must do if we claim that *any* bodily malfunction is an impairment. When we think about what our sailor has in common with a paraplegic person in England, it does not seem to be something also shared by anyone with a cold.

Alternatively, sometimes impairment is defined not as a malfunction, or as a deficiency of any other kind, but instead as a mere difference. For example, Simo Vehmas characterizes the social model so that it treats an impairment as an “organic anomaly,” and this way of thinking fits well with those who think of disabilities as mere differences that are subject to prejudice.<sup>59</sup> But this way of thinking of impairment also seems unsustainable. For example, the sprinter Usain Bolt is not leg impaired, despite having biologically anomalous legs. And again, when we think about what our sailor has in common with a paraplegic person in England, it does not seem to be something shared by Usain Bolt.

So it is not clear exactly how we should define impairment.<sup>60</sup> But since the

56 Cf. Shakespeare, *Disability Rights and Wrongs*, 10–19; and note that neither the Equality Act 2010: Guidance (esp. A3) nor the Americans with Disabilities Act of 1990 offers a definition of the term.

One classic objection to the social model is that it understates the difficulties raised by impairment alone, by attributing *all* of the difficulties faced by disabled people to society (e.g., French, “Disability, Impairment or Something in Between?”). Whatever its merits, this worry is distinct from mine.

57 See also, e.g., Crow, “Including All Our Lives,” 211.

58 Cf. Hughes and Paterson, “The Social Model of Disability and the Disappearing Body,” 329.

59 Vehmas, “Ethical Analysis of the Concept of Disability,” 210–11; see also World Health Organization, “International Classification of Impairments, Disabilities, and Handicaps,” 27.

60 Other defenders of the social model have suggested that impairment itself might be defined in social terms (e.g., Tremain, “On the Government of Disability”); Shakespeare and Watson,

social theory relies heavily on the notion of impairment, the social theory is at best radically incomplete. It cannot adequately capture what our sailor has in common with a paraplegic person in England, and cannot adequately specify what distinguishes disabilities from other kinds of social disadvantage such as gender. Of course, we *could* rescue the social model from this objection by deploying the inability theory as our theory of impairment.<sup>61</sup> If we did this, any remaining disagreement between the inability theory as I have presented it, and this view, would be purely verbal: about whether to define “impairment” as inability, and use “disability” to refer to the socially mediated costs of that inability, or whether to instead define “disability” as inability, and to refer to the socially mediated costs of that inability using some other label. If others want to insist on using terminology in the former way, I am happy to do so: I cannot see that anything deep hangs on this.

In summary, we should distinguish four different versions of the social model. The minimalist social model is positively entailed by the inability theory. The causal and normative social models are independent of the inability theory. And the social theory is plausible only if combined with the inability theory. I therefore conclude that the inability theory is consistent with the social model of disability in any plausible form.

## 7. CONCLUSION

In this paper I have argued that we should define disability as a particular kind of inability. We should endorse:

*The Final Inability Theory:* To be disabled is to be less able to do something than is typical, where this degree of inability (1) is partly explained by features of your body that are atypical, and (2) is not explained by anyone’s attitudes toward those bodily features.

This theory can overcome the objections raised by its opponents, is superior to the well-being theory and to conventionalist theories of disability, and is consistent with the social model of disability. I leave open that there might yet be still

---

“The Social Model of Disability”). But if anything, this would make the case of the shipwrecked sailor all the more problematic since now we cannot say that she is disabled or that she is impaired and yet we still should have some term to capture what she has in common with a paraplegic person in England.

61 For example, Howard and Aas defend the social model, and understand impairment as a kind of malfunction—but they later allow that we might need to understand impairments as only those malfunctions that limit activity (“On Valuing Impairment,” 1120).



other as yet unvoiced objections to the inability theory, but until they are articulated we should tentatively endorse the inability theory of disability.<sup>62</sup>

University of Southampton  
a.m.gregory@soton.ac.uk

## REFERENCES

- Aas, Sean. "Disabled—Therefore, Unhealthy?" *Ethical Theory and Moral Practice* 19, no. 5 (November 2016): 1259–74.
- Albrecht, Gary L., and Patrick J. Devlieger. "The Disability Paradox: High Quality of Life against All Odds." *Social Science and Medicine* 48, no. 8 (April 1999): 977–88.
- Americans with Disabilities Act of 1990, Pub. L. No. 101–336, 104 Stat. 328 (1990).
- Amundsen, Ron. "Against Normal Function." *Studies in History and Philosophy of Biological and Biomedical Sciences* 31, no. 1 (March 2000): 33–53.
- . "Disability, Handicap, and the Environment." *Journal of Social Philosophy* 23, no. 1 (March 1992): 105–18.
- Barnes, Colin. *Disabled People in Britain and Discrimination*. London: Hurst and Co., 1991.
- Barnes, Elizabeth. *The Minority Body*. Oxford: Oxford University Press, 2016.
- Bickenbach, Jerome E. *Physical Disability and Social Policy*. Toronto: University of Toronto Press, 1993.
- Boorse, Christopher. "Health as a Theoretical Concept." *Philosophy of Science* 44, no. 4 (December 1977): 542–73.
- . "On the Distinction between Disease and Illness." *Philosophy and Public Affairs* 5, no. 1 (Autumn 1975): 49–68.
- . "A Rebuttal on Health." In *What Is Disease?* Edited by James M. Humber and Robert F. Almeder, 1–134. Totowa, NJ: Humana Press, 1997.
- Boyd, Norman F., Heather J. Sutherland, Karen Z. Heasman, David L. Tritchler, and Bernard J. Cummings. "Whose Utilities for Decision Analysis?" *Medical Decision Making* 10, no. 1 (February 1990): 58–67.
- Brownlee, Kimberley, and Adam Cureton, eds. *Disability and Disadvantage*. Oxford: Oxford University Press, 2011.
- Buchanan, Allen, Dan W. Brock, Norman Daniels, and Daniel Wikler. *From*

62 I have too many people to thank for helpful comments on this paper, but I should record special thanks to the two excellent referees for this journal, who gave extremely generous feedback.

- Chance to Choice: Genetics and Justice*. Cambridge: Cambridge University Press, 2002.
- Bury, Mik. "Defining and Researching Disability: Challenges and Responses." In *Exploring the Divide*, edited by Colin Barnes and Geoff Mercer, 18–38. Leeds: The Disability Press, 1996.
- Crow, Liz. "Including All Our Lives." In *Encounters with Strangers: Feminism and Disability*, edited by Jenny Morris. London: Women's Press, 1996.
- Daniels, Norman. *Just Health Care*. Cambridge: Cambridge University Press, 1985.
- Daniels, Norman, Susannah Rose, and Ellen Daniels Zide. "Disability, Adaptation, and Inclusion." In Brownlee and Cureton, *Disability and Disadvantage*, 54–85.
- Disability Discrimination Act 1995. Chapter 50. London: HMSO.
- Dworkin, Gerald. *The Theory and Practice of Autonomy*. Cambridge: Cambridge University Press, 1988.
- Equality Act 2010. London: HMSO. <http://www.legislation.gov.uk/ukpga/2010/15/contents>.
- Finkelstein, Victor. *Attitudes and Disabled People*. New York: World Rehabilitation Fund, 1980.
- French, Sally. "Disability, Impairment or Something in Between?" In *Disabling Barriers—Enabling Environments*, edited by John Swain, Vic Finkelstein, Sally French, and Mike Oliver, 17–25. London: Sage Publications, 1993.
- Glover, Jonathan. *Choosing Children*. Oxford: Oxford University Press, 2008.
- Goering, Sara. "'You Say You're Happy, but . . .': Contested Quality of Life Judgments in Bioethics and Disability Studies." *Bioethical Inquiry* 5, nos. 2–3 (June 2008): 125–35.
- Grice, Paul. "Logic and Conversation." In *Syntax and Semantics 3: Speech Acts*, edited by Peter Cole and Jerry L. Morgan, 41–58. New York: Academic Press, 1975.
- Harris, John. "One Principle and Three Fallacies of Disability Studies." *Journal of Medical Ethics* 27, no. 6 (December 2001): 383–87.
- Haslanger, Sally. "Gender and Race: (What) Are They? (What) Do We Want Them to Be?" *Noûs* 34, no. 1 (March 2000): 31–55.
- Howard, Dana, and Sean Aas. "On Valuing Impairment." *Philosophical Studies* 175, no. 5 (May 2018): 1113–33.
- Hughes, Bill, and Kevin Paterson. "The Social Model of Disability and the Disappearing Body: Towards a Sociology of Impairment." *Disability and Society* 12, no. 3 (1997): 325–40.

- Hull, Richard. "Defining Disability: A Philosophical Approach." *Res Publica* 4, no. 2 (September 1998): 199–210.
- Kahane, Guy, and Julian Savulescu. "The Welfarist Account of Disability." In Brownlee and Cureton, *Disability and Disadvantage*, 14–53.
- Lane, Harlan. "Do Deaf People Have a Disability?" *Sign Language Studies* 2, no. 4 (Summer 2002): 356–79.
- Lim, Chong-Ming. "Disabilities Are Also Legitimately Medically Interesting Constraints on Legitimate Interests." *Mind* 127, no. 508 (October 2018): 977–1002.
- Moller, Dan. "Wealth, Disability, and Happiness." *Philosophy and Public Affairs* 39, no. 2 (Spring 2011): 177–206.
- Morris, Jenny. *Independent Lives? Community Care and Disabled People*. Basingstoke: Palgrave Macmillan, 1993.
- Nussbaum, Martha. *Frontiers of Justice: Disability, Nationality, Species Membership*. Cambridge, MA: Harvard University Press, 2006.
- . *Women and Human Development: The Capability Approach*. Cambridge: Cambridge University Press, 2000.
- Oliver, Michael. *Understanding Disability: From Theory to Practice*. Basingstoke: Macmillan, 1996.
- Sackett, David L., and George W. Torrance. "The Utility of Different Health States as Perceived by the General Public." *Journal of Chronic Diseases* 31, no. 11 (1978): 697–704.
- Savulescu, Julian, and Guy Kahane. "Disability: A Welfarist Approach." *Journal of Clinical Ethics* 6, no. 1 (March 2011): 45–51.
- Schroeder, Andrew. "Health, Disability, and Well-Being." In *The Routledge Handbook of Philosophy of Well-Being*, edited by Guy Fletcher, 221–32. London: Routledge, 2016.
- . "Well-Being, Opportunity, and Selecting for Disability." *Journal of Ethics and Social Philosophy* 14, no. 1 (2018): 1–27.
- Schwartz, Barry. *The Paradox of Choice*. New York: Harper Perennial, 2004.
- . "Self-Determination: The Tyranny of Freedom." *American Psychologist* 55, no. 1 (January 2000): 79–88.
- Sen, Amartya. "Capability and Well-Being." In *The Quality of Life*, edited by Martha Nussbaum and Amartya Sen, 30–53. Oxford: Clarendon Press, 1993.
- . "Equality of What?" In *Tanner Lectures on Human Values*, vol. 1, edited by Sterling M. McMurrin, 195–220. Cambridge: Cambridge University Press, 1980.
- Shakespeare, Tom. *Disability Rights and Wrongs*. Abingdon: Routledge, 2006.
- Shakespeare, Tom, Sue Thompson, and Michael Wright. "No Laughing Matter:

- Medical and Social Experiences of Restricted Growth." *Scandinavian Journal of Disability Research* 12, no. 1 (2010): 19–31.
- Shakespeare, Tom, and Nicholas Watson. "The Social Model of Disability: An Outdated Ideology?" In *Exploring Theories and Expanding Methodologies: Where We Are and Where We Need to Go*, edited by Sharon N. Barnartt and Barbara M. Altman, 9–28. Bingley: Emerald Group Publishing Limited, 2001.
- Siebers, Tobin. *Disability Theory*. Ann Arbor: University of Michigan Press, 2008.
- Silvers, Anita. "On the Possibility and Desirability of Constructing a Neutral Conception of Disability." *Theoretical Medicine and Bioethics* 24, no. 6 (November 2003): 471–87.
- Stone, Sharon Dale. "Reactions to Invisible Disability: The Experiences of Young Women Survivors of Hemorrhagic Stroke." *Disability and Rehabilitation* 27, no. 6 (2005): 293–304.
- Terzi, Lorella. "Vagaries of the Natural Lottery? Human Diversity, Disability, and Justice: A Capability Perspective." In Brownless and Cureton, *Disability and Disadvantage*, 86–111.
- Tremain, Shelley. "On the Government of Disability." *Social Theory and Practice* 27, no. 4 (October 2001): 617–36.
- Ubel, Peter A., and George Loewenstein. "Pain and Suffering Awards: They Shouldn't Be (Just) about Pain and Suffering." *The Journal of Legal Studies* 37, no. S2 (2008): S195–S216.
- Ubel, Peter A., George Loewenstein, and Christopher Jepson. "Whose Quality of Life? A Commentary Exploring Discrepancies between Health State Evaluations of Patients and the General Public." *Quality of Life Research* 12, no. 6 (September 2003): 599–607.
- Vehmas, Simo. "Ethical Analysis of the Concept of Disability." *Mental Retardation* 42, no. 3 (June 2004): 209–22.
- . "Philosophy and Science: The Axes of Evil in Disability Studies?" *The Journal of Medical Ethics* 34, no. 1 (February 2008): 21–23.
- Wasserman, David, and Adrienne Asch. "Understanding the Relationship between Disability and Well-Being." In *Disability and the Good Human Life*, edited by Jerome E. Bickenbach, Franziska Felder, and Barbara Schmitz, 139–67. Cambridge: Cambridge University Press, 2014.
- Wendell, Susan. *The Rejected Body: Feminist Philosophical Reflections on Disability*. London: Routledge, 1996.
- . "Unhealthy Disabled: Treating Chronic Illnesses as Disabilities." *Hypatia* 16, no. 4 (November 2001): 17–33.
- World Health Organization. "International Classification of Impairments, Disabilities, and Handicaps." Geneva: World Health Organization, 1980.

## FALSE EXEMPLARS ADMIRATION AND THE ETHICS OF PUBLIC MONUMENTS

*Benjamin Cohen Rossi*

AT LATEST COUNT, over one hundred symbols of the Confederacy have been removed from public spaces since the 2015 attack on African American churchgoers in Charleston, South Carolina, by white supremacist Dylann Roof. However, according to the Southern Poverty Law Center, more than 1,700 symbols remain, many of them protected by state laws in former Confederate states.<sup>1</sup> Moreover, the controversy over public monuments is not limited to Confederate symbols. In 2015, an African American civil rights group at Princeton University orchestrated a walkout to demand that the twenty-eighth U.S. President's name be expunged from the Woodrow Wilson School of Public and International Affairs.<sup>2</sup> Nor is activism centered around monuments limited to the United States. A 2015 Twitter campaign, "Rhodes Must Fall," led to the removal of a statue of Cecil Rhodes from the campus of South Africa's Cape Town University, inspiring a call in 2016 for the removal of a Rhodes statue from Oxford University's campus.<sup>3</sup> It appears that a new generation of activists is reinvigorating debate over the public commemorative landscape.

While this debate is in no way limited to statues, it frequently crystallizes around public representations of historical figures who expressed support for the oppression of certain groups or contributed to their past or present oppression. In this paper, I will consider what should be done about such representations. A number of philosophers have articulated arguments for modifying or removing public monuments. Joanna Burch-Brown grounds her argument for removal in what I call the "honorific" function of such representations—the

1 See Southern Poverty Law Center, "SPLC Report."

2 See Philpott and Weber, "Woodrow Wilson Statue Removal Prompts a Closer Look at His History with Race Relations"; Brait, "Princeton Students Demand Removal of Woodrow Wilson's Name from Buildings."

3 BBC News, "Will 'Rhodes Must Fall' Fail?"

ways in which they express and tend to cultivate admiration for their subjects.<sup>4</sup> In the first two sections of the paper, I develop a novel argument for modifying these representations based on this insight. I argue that leaving such representations unmodified in the public space tends to undermine the dignity of members of oppressed groups as well as their assurance that society and government are committed to their rights and constitutional entitlements. In the paper's third section, I develop a "balancing test" for determining whether the relevant moral and pragmatic considerations favor making a particular representation inaccessible to the public, or recontextualizing it for public consumption. Unlike some of the extant philosophical treatments of honorific representations that focus on particular monuments, this balancing test is designed for general application to any honorific representation that satisfies the presumptive case for modification. To conclude, I offer some reasons why weak forms of recontextualization that do not involve altering institutional context may often be an insufficient remedy for the problems I describe.

#### 1. HONORIFIC REPRESENTATIONS AND THEIR MEANINGS

As I mentioned, Burch-Brown's argument for removing or recontextualizing cultural objects like flags, statues, and place names associated with colonialism and slavery is grounded in their honorific function. Her crucial insight is that statues and place names confer honor and esteem on their subjects.<sup>5</sup> In this section, I develop an account of the mechanisms by which some public representations convey messages of admiration for their subjects, drawing on insights from related work by Alfred Archer and Benjamin Matheson on the ethics of admiration.<sup>6</sup>

For our purposes, the term "honorific representation" designates *any representation of an individual in a public space that depicts that individual as an exemplar of a value or values, such as courage, integrity, or justice*. Thus, what I will say about statues applies equally to paintings, frescoes, and even bare inscriptions. I understand an exemplar as a *person who is the fitting object of admiration on account of their instantiation of a value or values*.<sup>7</sup>

Honorific representations often have multiple meanings, and meaning can be extracted from different features of them. So, what determines their meaning? Answers to this question fall along a spectrum bounded by two extremes. On

4 Burch-Brown, "Is It Wrong to Topple Statues and Rename Schools?" 59–87, esp. 68–69.

5 Burch-Brown, "Is It Wrong to Topple Statues and Rename Schools?" 68.

6 Archer and Matheson, "When Artists Fall."

7 Archer and Matheson make the point that honoring someone picks them out as a fitting target of admiration ("When Artists Fall," 248).

one end is the view that the meaning of a representation is determined solely by what those involved in its creation intended it to mean. On the other end is the view that its meaning is determined solely by whatever members of the public take it to mean.<sup>8</sup> I see no reason why we ought to adopt either of these views. Instead, I propose that for our purposes we treat both the attitudes of the creators and those of the public as legitimate sources of meaning. In a recent paper on the ethics of admiring immoral artists, Archer and Matheson introduce a useful distinction among different kinds of meaning that utterances can possess. In the context of honorific representations, what they call *intended meaning* is the thoughts, attitudes, and concepts that a representation's creators intended it to convey. I will call the thoughts, attitudes, and concepts that an honorific representation's creators *express* through it without necessarily intending to do so its *implicit meaning*.<sup>9</sup> Finally, I will call the thoughts, attitudes, and concepts that the public takes the representation to convey—*whether or not that interpretation is justifiable*—its *public meaning*.<sup>10</sup> It should be clear from these definitions that the same representation may bear multiple meanings. Insofar as a representation is created by multiple people, there may be many, sometimes contradictory intended and implicit meanings. And given that the public is not a monolith, there are likely to be multiple public meanings as well. Indeed, the fact of these multiple public meanings will be important in our discussion of the factors that determine whether we ought to recontextualize or remove these representations.

It may be objected that it is illegitimate to include in our moral calculus those meanings that the public *unjustifiably* takes a representation to convey. Why ought we care about the unreasonable interpretations? The answer is that public meanings are crucial links in the causal chain whereby representations bring about morally relevant effects. Hence, we have reason to care about how the public interprets them even if those interpretations are unreasonable. We can observe a similar line of reasoning in deliberations about public health policy. Even if we think that philosophical or religious opposition to vaccination is unreasonable, a policy of blanket mandatory vaccination may be morally undesirable because of the backlash it tends to provoke and because of the duty to

8 Cf. Schedler, "Are Confederate Monuments Racist?" 287–308.

9 Archer and Matheson call the values and cares implicit in a person's utterance its *attitudinal* meaning ("When Artists Fall," 251). However, because both intended and "attitudinal" meanings express the speaker's attitudes, I think a better label for the latter is *implicit* meaning.

10 Archer and Matheson also define *public* meaning as the meaning that "others can justifiably attribute to our acts given the context in which we perform them" ("When Artists Fall," 251). But for reasons I discuss below, I do not want to limit my analysis to *reasonable* or *justifiable* public interpretations of honorific representations.

respect patient autonomy.<sup>11</sup> By the same token, if we are aware that the public is likely to unreasonably ascribe a certain meaning to a representation and that ascription has morally detrimental effects, then that is a legitimate consideration when deliberating about what to do with the representation.

It might now be objected that because knowledge of intended and implicit meanings typically requires more investigation of a representation's history than members of the public are willing to undertake, intended and implicit meanings can be ignored for our purposes. However, it is not uncommon for the intended and implicit meanings of representations to become more widely known through the efforts of journalists or historians, which in turn shapes the public's interpretations.<sup>12</sup> In this way, intended and implicit meanings can have indirect effects on public meaning. Furthermore, intended and implicit meanings are part of the complete analysis of the harms of these representations. The case for this claim will have to wait until the next section. So, for our purposes it is legitimate to consider the intended, implicit, and public meanings of honorific representations even if these meanings conflict, and even if the interpretations of the public are unreasonable.

A morally crucial feature of honorific representations is that they depict their subjects as *exemplars*, or fitting objects of admiration. They can do this by expressing *propositions* or thoughts about their subjects' exemplarity, or by expressing *attitudes* of admiration for their subjects; and they can express both in virtue of their *relations* to other things, by employing visual symbols and metaphors, or by accompanying inscriptions. Honorific representations depict their subjects as exemplars *relationally*: by being located in public space, by being created for certain purposes, or by being informed by certain values, they convey the message that the subjects they represent are to be admired. For example, the Jackson and Lee Monument in Baltimore, which was removed on the orders of the Baltimore City Council in August 2017, was the fruit of a one-hundred-thousand-dollar bequest by local banker J. Henry Ferguson, who intended to hold up the Confederate generals as good examples to Maryland youth.<sup>13</sup> Even where the intention of their creators was not to express admiration for their subjects, that these representations' subjects are exemplars is usually a value judgment that was operative in their creation, and so part of their *implicit meaning*. For

11 See Salisbury, "Should Childhood Vaccination Be Mandatory?"; El Amin et al., "Ethical Issues Concerning Vaccination Requirements"; and Pierik, "Mandatory Vaccination."

12 For an example of journalistic work on the Confederate battle flag, see Appelbaum, "Why Is the Flag Still There?" For a recent example of revisionary historical work on Supreme Court Chief Justice John Marshall, see Finkelman, *Supreme Injustice*.

13 See Kelly, *Outdoor Sculpture in Baltimore*, 198–99.



example, the history of the Robert E. Lee Monument in New Orleans clearly indicates that its erection was motivated by an underlying admiration for him.<sup>14</sup> Moreover, it can be, and often is, reasonably inferred that the mere existence of a representation of a person in a public space implies that its subject is considered an exemplar.<sup>15</sup> Hence, that the subject is an exemplar can be part of the public meaning of its existence in public space.

Honorific representations also *visually* represent their subjects as exemplars through the use of culturally conditioned markers of exemplarity. One way of visually suggesting exemplarity is by placing subjects at a commanding height, e.g., by depicting them in larger-than-life scale. For example, the Lee monument in New Orleans featured a sixteen-and-a-half-foot bronze statue atop a sixty-foot-tall marble column. This technique exploits the metaphorical association between the relation of *being above* and the relation of *being better than*.<sup>16</sup> In addition, honorific representations tend to depict their subjects in physically idealized terms, exploiting the human tendency to infer moral excellence from physical excellence—the “what is beautiful is good” bias.<sup>17</sup> Art-historical references can also be used to reinforce the suggestion of exemplarity. The Lincoln Memorial in Washington, DC, references the Athenian Parthenon in order to suggest Lincoln’s godlike role in preserving American democracy.<sup>18</sup>

Finally, honorific representations can represent their subjects as exemplars through the use of inscriptions. The Jackson and Lee Monument had an inscription on its base that read: “They were great generals and Christian soldiers and waged war like gentlemen.”<sup>19</sup> Similarly, the plaque under an 1895 statue of Edward Colston in the city of Bristol, United Kingdom, reads: “Erected by citizens of Bristol as a memorial of one of the most virtuous and wise sons of their

14 For a concise summary of this history, see Nicholson, “Robert E. Lee Monument.”

15 I invite the reader to consider the many public representations they have encountered in their lifetime. It would be very surprising to me if a large proportion of them did not, on a reasonable interpretation, depict their subjects as objects of admiration.

16 For a comprehensive examination of these metaphorical associations, see Lakoff and Johnson, *The Metaphors We Live By*. See also Moore, “What the Removal of New Orleans’s White Supremacist Monuments Means to My Students.”

17 See Dion, Berscheid, and Walster, “What Is Beautiful Is Good,” 285–90; Langlois et al., “Maxims or Myths of Beauty?”; and Eagly et al., “What Is Beautiful Is Good, but . . .”

18 For an in-depth analysis and history of the Lincoln Memorial, see Greenberg, “With Meaning for All.” Note that according to the article, the Lincoln Memorial was unusual for its time because the statue of Lincoln does not idealize his physical features.

19 For more details about this and other Baltimore Confederate monuments, see Special Commission to Review Baltimore’s Public Confederate Monuments, “Lee Jackson Monument, 1948.”

city.” Notably absent from this inscription is any mention of Colston’s role in the transatlantic slave trade.<sup>20</sup> In some cases, honorific representations are entirely constituted by inscriptions. The Heyward Shepherd Memorial at Harper’s Ferry, Virginia, erected in 1931, honors an African American railroad worker who died in the early hours of John Brown’s famous 1859 raid on the federal arsenal at Harper’s Ferry. The rectangular granite slab has no visual elements, but its inscription reads:

On the night of October 16, 1859, Heyward Shepherd, an industrious and respected colored freeman, was mortally wounded by John Brown’s raiders.... This boulder is erected by the United Daughters of the Confederacy and the Sons of Confederate Veterans as a memorial to Heyward Shepherd, exemplifying the character and faithfulness of thousands of Negroes who under many temptations throughout subsequent years of war. So [*sic*] conducted themselves that no stain was left upon a record which is the peculiar heritage of the American people, and an everlasting tribute to the best in both races.<sup>21</sup>

Although Shepherd was a free man, the inscription depicts him as in some sense representative of faithful slaves who resisted the myriad “temptations” to escape bondage, choosing instead to remain faithful to their masters. Hence, while Shepherd was unlikely to have seen himself as a loyal servant of the South—and certainly not of the Confederacy—the inscription attributes to him motives of loyalty to the South that he likely lacked. Its cavalier attitude toward historical truth reflects its actual function as a piece of propaganda for the Jim Crow social order, about which more will be said shortly.<sup>22</sup>

One may wonder whether the Heyward Shepherd Memorial ought to count as an honorific representation. Its erectors likely did not intend it to represent an exemplar of virtues proper to people *like them*, but to people properly subordinated by post-Reconstruction segregation. Furthermore, what its erectors

20 Between 1680 and 1692, Colston served as a major investor, manager, and eventually deputy governor of the Royal Africa Company (RAC), then Britain’s only official slaving company. During this time, of 84,500 slaves transported on RAC ships, nearly 20,000 died *en route* due to unhygienic conditions onboard, dehydration, scurvy, and dysentery. For an overview of the controversy surrounding Colston in Bristol, see Parks, “Edward Colston.” More information about Colston and the “Countering Colston” campaign can be found at <https://counteringcolston.wordpress.com>.

21 For a thorough account of the monument’s history, see Johnson, “An ‘Ever Present Bone of Contention.’”

22 For an accessible overview of the Reconstruction and Jim Crow periods, see Gates, *Stony the Road*.

may have understood to be virtues proper to subordinates may not be anything like what we *now* recognize as virtues.<sup>23</sup> However, there is good reason to frame a concept of *honorific representation* capacious enough to accommodate these cases. As will become clear in the next section, if we did not do so, we could not fully understand the ways in which such representations are objectionable. Thus, it would be unduly narrow to insist that something is an *honorific representation* only if it exemplarizes subjects for what we currently call “virtues” or “values”; it should be possible to recognize when a culture *honors* one of its members for instantiating a virtue or value we do not share.<sup>24</sup> My account of honorific representation simply requires that it represent a subject as an exemplar of a value, but not necessarily one of *our* values or “correct” values. Nor need it be a value or virtue deemed proper for the representation’s creators. Consider a statue of a Roman woman that depicts her as the embodiment of the virtue of *pudicitia*, a virtue sometimes translated as “chastity.”<sup>25</sup> Surely, we can understand this representation as *honoring* her and underscoring the value of *pudicitia*, even if the statue’s male creators did not regard *pudicitia* as a value *they* ought to instantiate.

## 2. WHAT IS OBJECTIONABLE ABOUT HONORIFIC REPRESENTATIONS?

A survey of the extant literature on the ethics of honorific representations yields three principal objections to them. The first is that they can express a degrading ideology. Both Burch-Brown and Johannes Schulz emphasize this point.<sup>26</sup> The second, which serves as the basis for Travis Timmerman’s argument for destroying Confederate monuments, is that they can cause psychological suffering.<sup>27</sup> Finally, Schulz also claims that they can undermine *self-respect*, a person’s self-perception as “a moral equal, with the same rights and duties that all other persons have.”<sup>28</sup> In this section, I argue that another important moral objection to some

23 Thanks to a reviewer at the *Journal of Ethics and Social Philosophy* for bringing this point to my attention.

24 This is not to say that there is no sense of the term “virtue” that is transculturally shared. It would be difficult for us to understand someone who calls a trait a “virtue,” but does not regard it as something like an appropriate disposition to act, choose, desire, and feel in a certain sphere of human experience. Cf. Nussbaum, “Non-Relative Virtues.”

25 For an examination of this concept, see Langlands, *Sexual Morality in Ancient Rome*. The term *pudicitia* was applied to men as well, but it seems that the “thick” specification of the virtue was different for men and women.

26 See Burch-Brown, “Is It Wrong to Topple Statues and Rename Schools?” 68; Schulz, “Must Rhodes Fall?” 168–72.

27 Timmerman, “A Case for Removing Confederate Monuments.”

28 Schulz, “Must Rhodes Fall?” 172–76.

honorific representations is that they undermine the dignity of members of marginalized groups and their assurance of respect for their dignity. While the other objections all carry serious moral weight, I suggest that a dignity-focused account has some explanatory advantages over the ideological account and is weightier than the psychological suffering account.

There is no question that some honorific representations express a degrading ideology. Sally Haslanger defines “ideology” as a set of social meanings that function to stabilize problematic social hierarchies through masking or illusions that make unjust social arrangements appear as if they are just.<sup>29</sup> With this in mind, consider the Heyward Shepherd Memorial. Even if the admiration memorials like Shepherd’s express for African American “loyalty” is in some sense genuine, it is part and parcel of the ideology of white supremacy. Representations of devoted African American slaves contributed to the effort in the post-Reconstruction South to morally legitimize segregation, violence, and other oppressive practices by depicting subordination as welcomed by African Americans themselves.<sup>30</sup> Indeed, “subordination” is precisely the word used in an 1894 op-ed in the *Confederate Veteran* calling for the erection of monuments to “faithful slaves”:

It seems opportune now to erect monuments to the Negro race of the war period. . . . There is not of record in history subordination and faithful devotion by any race of people comparable to the slaves of the Southern people during our great four years’ war for independence.<sup>31</sup>

Thus, in this context, apparently admiring representations of an African American both assert and promote white supremacy. In this way, they express a degrading ideology that is committed to the denial of basic human rights and constitutional entitlements to African Americans. However, as I will presently argue, the expression of a degrading ideology does not exhaust the problematic aspects of honorific representations, and it is not always clear that every morally problematic honorific representation expresses a degrading ideology.<sup>32</sup>

Consider the General William Tecumseh Sherman Monument in New York City, which commemorates Sherman’s 1864 march through Georgia. If Confederate monuments are plainly expressive of white supremacist ideology, it is

29 See Haslanger, “Racism, Ideology, and Social Movements.”

30 For a detailed account of the relationship between Southern monuments, the Lost Cause myth, and white supremacy, see Leib and Webster, “On Remembering John Winberry and the Study of Confederate Monuments on the Southern Landscape”; and Savage, *Standing Soldiers, Kneeling Slaves*.

31 Confederate Southern Memorial Association, *Confederate Veteran*.

32 As Schulz himself notes (“Must Rhodes Fall?” 172).

less clear in this case. Such an expression was not in the latter case any part of the intended or implicit meaning of the monument.<sup>33</sup> And unlike Schulz's case of Black South African prisoners forced to wear boys' shorts under apartheid, the creators of the monument may not have knowingly erected it in a cultural context in which it would be understood by oppressor and oppressed alike as an expression of a disrespectful ideology.<sup>34</sup> Yet, as overall commander of the US Army after the Civil War, Sherman prosecuted a brutal war against the Plains Indians that included the deliberate mass slaughter of the buffalo, the economic basis of their existence.<sup>35</sup> His justification for the policy of displacing Native Americans in favor of white settlers echoes a narrative of the European settler as the "culturally and morally superior savior who brought civilization to the 'wild' Indians," a narrative that "portrays Indigenous people as lacking the very ability to rationally choose and act on a plan of life."<sup>36</sup> Given this, it would be reasonable for Native Americans to object to expressions of unqualified admiration for such a figure.<sup>37</sup> This suggests that even if the Sherman Monument does not express an ideology of white supremacy, it is not morally unproblematic. The account I articulate presently can explain why.

Honorific representations can undermine both the dignity of members of currently or historically disadvantaged groups and their assurance of respect for their rights. Here I use the word "dignity," following Jeremy Waldron, to denote the "basic social standing" of members of such groups, "the basis of their recognition as social equals and as bearers of human rights and constitutional entitlements."<sup>38</sup> Waldron asks us to consider the effects of "speech" in the broad

33 Compare this to what Johannes Schulz says about Confederate monuments: "The monuments were not erected simply to commemorate military leaders . . . but were *intended* as political statements in defense of the white supremacist ideology of the Confederacy" ("Must Rhodes Fall?" 168, emphasis added).

34 Cf. Schulz, "Must Rhodes Fall?" 169.

35 See Janda, "Shutting the Gates of Mercy"; Smits, "The Frontier Army and the Destruction of the Buffalo."

36 Schulz, "Must Rhodes Fall?" 175–76.

37 Similarly, Schulz points out that a monument to Abraham Lincoln may be offensive to Native Americans because, *inter alia*, he signed the Homestead Act into law ("Must Rhodes Fall?" 175). But to say that the Lincoln Memorial and a statue of Robert E. Lee erected during the civil rights era are equally *expressive* of white supremacy seems to elide some important differences.

38 See Waldron, *The Harm in Hate Speech*, 59. See also Waldron and Dan-Cohen, *Dignity, Rank, and Rights*.

sense—encompassing public imagery, signage, audible speech, and so on—on members of marginalized groups.<sup>39</sup> His primary argument is that

the look of a society is one of its primary ways of conveying assurances to its members about how they are likely to be treated, for example, by the hundreds or thousands of strangers they encounter or are exposed to in everyday life. . . . The point of the visible self-presentation of a well-ordered society, then, is not just aesthetic; it is the conveying of an assurance that they can count on being treated justly.<sup>40</sup>

Drawing on Rawls's conception of a well-ordered society as one in which "citizens accept and know that others likewise accept [basic principles of justice], and this knowledge in turn is publicly recognized," Waldron argues that members of currently or historically oppressed groups require a general and diffuse assurance that other citizens are committed to dealing with them on the basis of the recognition of their rights.<sup>41</sup> Without this assurance, the social world they inhabit is hostile, insecure, and unpredictable; uncountable possible interactions with members of society and government actors become the scene of potential rights violations. This unpredictability could lead to their withdrawal from the public space, making it harder to secure their individual and group interests and eroding democratic legitimacy. Building on these insights, I propose that where honorific representations express a degrading ideology, endorse or express disregard for grave wrongs done to members of a marginalized group, or encourage unqualified admiration for the group's oppressors, they can contribute to the undermining of confidence that society is committed to treating the group's members "with dignity"—with due recognition of their basic rights and constitutional entitlements. The message conveyed by society's omitting to modify these representations to reflect their subjects' roles in the oppression of these groups—or worse, its *defense* of them as elements of a morally unobjectionable "heritage"—is that there is widespread societal approval of their morally objectionable intended, implicit, and public meanings. Since those meanings are inconsistent with respect for the rights and entitlements of members of these groups, leaving these representations unmodified will tend to contribute to the undermining of their assurance that others are committed to treating these groups in ways consistent with their rights-bearing status.

Furthermore, sometimes this lack of confidence will be *accurate* because of the effects of honorific representations: in particular, their endorsement of an

39 Waldron, *The Harm in Hate Speech*, 75.

40 Waldron, *The Harm in Hate Speech*, 82–85.

41 Waldron, *The Harm in Hate Speech*, ch. 4.

inherently degrading ideology or promotion of unqualified admiration of their subjects, which may make it more difficult for the public to acknowledge the injustices in which they were involved. Waldron argues that “one holds a certain status [dignity] not just when one happens to have a given set of entitlements, but when the recognition of those rights or entitlements is basic to how one is in fact dealt with.”<sup>42</sup> Dignity is a social status, and requires not just the possession of human rights and constitutional entitlements but the widespread recognition *that* one bears these entitlements. If this is the case, and if honorific representations can contribute to a general failure of such recognition, then we can say that these representations can undermine dignity itself. These effects are particularly acute when the representations are the responsibility of the state. As the principal guarantor of rights, the state must strongly communicate assurances that it is committed to upholding these rights, and these assurances can be diluted if the state allows representations that elide or endorse oppression to remain unmodified in the public space.

The deleterious effects on oppressed groups’ assurances of just treatment and dignity are also compounded when honorific representations are used as rallying points for further public expressions of hatred and bigotry. For example, at the 2017 “Unite the Right” rally in Charlottesville, Virginia, a statue of Robert E. Lee served as a rallying point for the alt-right protestors. The purpose of such demonstrations is not only to spew hateful speech; it is also, as Waldron points out, “to indicate to others that they are not alone in their racism or bigotry.”<sup>43</sup> In other words, the public show of hatred is aimed at undermining the assurances that are critical to the basic social standing of members of oppressed groups and encouraging the erosion of their dignity.

I mentioned three ways in which honorific representations can undermine dignity and the assurance of dignified treatment: expressing a degrading ideology, endorsing or expressing disregard for grave wrongs done to members of a historically or currently oppressed group, or encouraging unqualified admiration for their oppressors. In what follows, I will explain how honorific representations do this *through* their honorific function. I have already described the way in which they can express degrading ideologies. But in addition, the implicit or public meanings of an honorific representation may be endorsements or elisions of grave historical injustices because the “valuable” traits or deeds for which the representation’s subject are honored were often either *instrumental to* morally objectionable ends or *enabled by* morally objectionable practices. For example, Robert E. Lee may have manifested genuine virtues in his prosecution

42 Waldron, *The Harm in Hate Speech*, 85. See also Rawls, *Political Liberalism*, 66.

43 Waldron, *The Harm in Hate Speech*, 95.

of the war, but the aim of the Confederate war effort was primarily to preserve the institution of slavery. Edward Colston may have manifested benevolence in his philanthropic work for Bristol, but this work was directly enabled by the exploitation of African slaves. Since we tend to think that it is inappropriate to feel or express admiration for traits or deeds that stand in these relations to morally objectionable ends or practices, we can reasonably infer that those involved in these representations' creation either did not see these ends or practices as morally objectionable, or did not see their subjects' traits or deeds as relevantly related to these morally objectionable ends and practices. If the former, then the implicit meaning of the representation is the endorsement of morally objectionable values or the unjustified downplaying of past injustices; if the latter, then the implicit meaning is, in part, historical falsehoods that also function to downplay past injustices. In either case, the public can reasonably infer that the representations either endorse morally objectionable values or elide grave injustices. Moreover, as Archer and Matheson point out, when both attitudes of admiration and of contempt are fitting, the choice to express only attitudes of admiration may, in certain contexts, convey the message that the immoral behavior is condoned.<sup>44</sup> For example, when the Jackson and Lee Monument highlights their "gentlemanly" prosecution of the war while the aims for which they fought pass without comment, this conveys the message that their personal virtue is more important than the fact that had their efforts ultimately borne fruit, slavery might have existed in America well into the twentieth century. Similarly, Burch-Brown quite plausibly claims that publicly honoring those who committed injustices against people of African or indigenous descent "inferiorizes [them] by signaling that the deep injustices suffered by their ancestors are not important enough to the community to cause it to repudiate their actions publicly."<sup>45</sup> For members of historically or currently oppressed groups, these messages can undermine their assurance that society and government are committed to their basic rights and entitlements.

In addition, honorific representations can undermine dignity and the assurance of dignified treatment by cultivating admiration for their subjects, which makes it more difficult to prevent swaths of the public from seeing them as moral exemplars. Relevant here is the distinction between *all-things-considered* exemplars and *qualified* exemplars.<sup>46</sup> Qualified exemplars are fitting objects of admiration on account of *some*, but not the majority, of their traits or deeds, while all-things-considered exemplars are fitting objects of admiration for most of

44 Archer and Matheson, "When Artists Fall," 9–13.

45 Burch-Brown, "Is It Wrong to Topple Statues and Rename Schools?" 70.

46 Cf. Zagzebski, *Exemplarist Moral Theory*.



their traits or deeds. Many of the subjects of honorific representations *are* qualified exemplars, and so are *fitting* objects of *qualified* admiration. But honorific representations themselves rarely do much to indicate the appropriate *limits* of our admiration for their subjects. In many cases, this lack of qualification is clearly deliberate. For example, in presenting the principal players of the Confederacy as morally unimpeachable characters, Southern whites morally laundered the war itself and the restoration of the antebellum racial status quo. As a result of this failure to explicitly delineate the proper bounds of admiration, it will likely overspill these bounds. The failure to explicitly qualify the scope of admiration compounds the natural spreading tendency of admiration highlighted by Archer and Matheson—the way in which admiration for one trait or action tends to influence our evaluations of other traits or actions.<sup>47</sup> In certain contexts, this may help cause the public to see these figures as all-things-considered rather than qualified exemplars, and thus to believe and feel incorrectly about these figures. In a cultural context such as the United States of 2020, in which the meanings of many historical figures and events are still hotly contested, these failures to provide context can help perpetuate degrading ideological narratives that are very much alive.

Yet it might be objected that so long as the prevalent cultural understandings of these representations undergo enlightened change, people will naturally come to acquire the proper attitudes toward their subjects despite public depictions of them as all-things-considered exemplars. It should be emphasized that there is no particular reason to think this will happen with respect to a host of problematic historical figures in the United States, such as Christopher Columbus, Woodrow Wilson, or Thomas Jefferson. Nevertheless, even when these cultural understandings begin to change, the existence of exemplarizing representations may hinder progress owing to the well-documented human aversion to moral ambivalence. Researchers in ambivalence studies distinguish between *attitudinal* ambivalence, which involves conflicting evaluations of an object, and *felt* ambivalence, the experience of being “torn” or having mixed emotions about an object.<sup>48</sup> They have consistently documented the following phenomena: (a) a substantial, if mediated, correlation between attitudinal ambivalence and felt ambivalence; (b) a strong aversion to felt ambivalence; and (c) because of (b), a

47 See Archer and Matheson, “When Artists Fall,” 14–16. For an account of a psychological mechanism possibly at work in this spreading tendency, see Gräf and Unkelbach, “Halo Effects in Trait Assessment Depend on Information Valence.”

48 See Newby-Clark, MacGregor, and Zanna, “Thinking and Caring about Cognitive Inconsistency.”

preference for evaluative consistency.<sup>49</sup> In addition, they have shown that those who experience felt ambivalence often seek to resolve the underlying attitudinal conflict by using biased information processing, such as the selective elaboration of one-sided information *consistent* with their initial attitude about the object or avoidance of information potentially *inconsistent* with their initial attitude about the object.<sup>50</sup> The upshot of this research is that people who have conflicting evaluations of an object tend to experience this conflict aversively and seek to resolve the conflict even at the expense of failing to meet their epistemic duties, such as the duty to gather evidence or think certain things through.<sup>51</sup>

One type of attitudinal ambivalence involves attitudes of opposing moral evaluation, such as admiration and contempt, toward the same person. One might wonder how such judgments are inconsistent, since there is a sense in which they have different objects. For example, Lee's insistence that he not be publicly memorialized after the war might be the object of someone's positive evaluation, while his actions in the service of a slave state might be the object of the same person's negative evaluation. While attitudinal ambivalence need not involve any *logical* inconsistency, it must involve conflicting assessments of the *same object*.<sup>52</sup> The worry is valid as far as it goes, but whether or not evaluative attitudes have the same object is sometimes not easy to determine, since the conception of the object may be more or less fine-grained. For example, if the object is something as coarse-grained as "Woodrow Wilson's presidency," then someone's positive evaluation (say, because of his advocacy for the liberal democratic world order) and negative evaluation (say, for his promotion of racism) have the same object. Moreover, the move from local judgments about particu-

49 See Newby-Clark, MacGregor, and Zanna, "Thinking and Caring about Cognitive Inconsistency"; Nordgren, Van Harreveld, and Van der Pligt, "Ambivalence, Discomfort, and Motivated Information Processing"; Tesser and Conlee, "Some Effects of Time and Thought on Attitude Polarization"; Elliot and Devine, "On the Motivational Nature of Cognitive Dissonance"; Harmon-Jones, "Cognitive Dissonance and Experienced Negative Affect"; and Festinger, *A Theory of Cognitive Dissonance*.

50 See Van Harreveld et al., "The Dynamics of Ambivalence"; Nordgren, Van Harreveld, and Van der Pligt, "Ambivalence, Discomfort, and Motivated Information Processing"; Newby-Clark, MacGregor, and Zanna, "Thinking and Caring about Cognitive Inconsistency"; and Erlich et al., "Postdecision Exposure to Relevant Information."

51 For an account of such duties, see Rosen, "Skepticism about Moral Responsibility."

52 Indeed, psychologists distinguish attitudinal ambivalence and cognitive dissonance in terms of whether the attitudes involved are logically inconsistent. While dissonance is a feeling of aversion arising from logically inconsistent cognitions, attitudinal ambivalence is best understood as the possession of "positive" and "negative" attitudes toward the same object. See Brannon and Gawronski, "Cognitive Consistency in Social Cognition." Thanks to an anonymous referee at the *Journal of Ethics and Social Philosophy* for raising this worry.

lar traits and actions to the global assessment of the whole person seems almost unavoidable. Once we start considering whether a *person* is exemplary or contemptible *in virtue of* those traits or deeds that warrant attitudes of admiration or contempt, we are in the realm of global assessment, and these judgments all have the whole person as their object. So, we are likely to experience some ambivalence if we are exposed to information that tends to show that a person is both exemplary in certain respects and contemptible in others. Given our aversion to such inconsistency, we are likely to seek resolution of the conflict. And if our initial attitude is one of admiration, then, as the ambivalence studies suggest, we are likely to pursue resolution of the conflict through the selective elaboration of information that favors that initial attitude or the avoidance of information inconsistent with that attitude. This means that we will come to see the subjects of representations as all-things-considered exemplars, since we will actively resist both information and emotions inconsistent with this status. This is particularly likely to happen in cultural contexts in which a process of “working through” the injustices of the past is only in its nascent stages.<sup>53</sup> Thus, if some historical representation successfully helps to cultivate admiration for its subject, it will tend to make it harder for people to *subsequently* acquire important information about that figure or to feel the proper negative moral attitudes toward them.

This is surely a bad thing in its own right, but I propose that it can contribute to a broader elision or downplaying of ongoing, historically rooted injustices. As I have already noted, the case of Confederate monuments illustrates how honorific representations can be connected to, and promote, ideologies that mask unjust social arrangements. In the cases of Robert E. Lee and other Confederate figures, *part of* this ideology was the partial and misleading account of their characters, their actions, the ends for which they acted, and the unjust practices in which they were involved. If acknowledging such injustices requires acknowledging the roles of various historical figures in their genesis and perpetuation, then acceptance of the admiration expressed by these representations will make it difficult for the public to acknowledge the related injustices. This failure of acknowledgment can, for reasons already discussed, undermine the assurances of historically or currently oppressed groups that their dignity will be respected. Furthermore, if these representations tend to encourage people to see their subjects as all-things-considered exemplars, and given admiration’s natural spreading tendency, we have reason to worry that such representations will encourage the formation of positive, or at least forgiving, attitudes toward their morally objectionable traits and deeds. This could, in turn, *actually* undermine society’s commitment to the basic rights and constitutional entitlements of members of

53 See Schulz, “Must Rhodes Fall?” 177–83.

oppressed groups by making the public more receptive to degrading ideological narratives that play on their admiration for these historical figures.

To conclude, I would like to contrast my case against honorific representations with two other criticisms I mentioned at the opening of this section: that they can cause psychological harm and can undermine self-respect. Travis Timmerman argues that because public Confederate monuments cause “unavoidable suffering”—at least for those who know the racist history that the monuments make salient—there is a strong moral reason to remove them.<sup>54</sup> He compares the harm caused by making salient the racist history of the United States to the harms done to survivors of World War II by punk musicians who wore swastika armbands in the 1970s.<sup>55</sup> While I do not deny the existence of these harms, it seems to me that the kind of harm highlighted by my account—the erosion of the dignity of members of oppressed groups and of the assurances that their moral rights and civic entitlements will be recognized by society and government—may be more important to well-being than the kind of psychological distress that Timmerman makes the centerpiece of his argument. A survivor of World War II could feel psychologically distressed by swastika armbands without having her dignity undermined or losing confidence that her dignity will be recognized by society and government officials, and this seems preferable to the loss of dignity and confidence *without* this sort of psychological distress.<sup>56</sup>

Johannes Schulz objects to honorific representations on the grounds that

54 Timmerman, “A Case for Removing Confederate Monuments,” 2–3.

55 Timmerman, “A Case for Removing Confederate Monuments,” 4.

56 Since I find that the unavoidable harms of certain honorific representations provide a strong moral reason for their modification, but not necessarily their destruction, it may seem that my view denies a key premise of Timmerman’s argument: that if the existence of monument *M* unavoidably harms an undeserving group, then there is a strong moral reason to end the existence of *M* (“A Case for Removing Confederate Monuments,” 2). However, this disagreement is merely verbal. Timmerman claims that what I call strong *recontextualization* of a monument, such as placing it in a museum in its proper historical context, causes it to cease to be a monument. Timmerman’s quick argument is that monuments are “reverential in nature,” but if placed in their proper historical context, they may no longer express reverence (“A Case for Removing Confederate Monuments,” 5). Notice that ordinary linguistic usage provides *prima facie* evidence against this claim: a museum may still refer to a statue as a “monument” even if the museum has properly contextualized it. Furthermore, if what makes a monument reverential are its implicit, intended, and public meanings, then recontextualization cannot alter the reverential *nature* of a monument but can only comment on it (see section 3 below). Thus, I would deny Timmerman’s premise on the semantic grounds that what is properly understood as *recontextualization* may be capable of preventing harm without destroying an honorific representation as such. However, we agree about the substantive point that the harms of honorific representations provide a strong moral reason either to recontextualize or remove them.

they can undermine people's self-respect.<sup>57</sup> This seems quite plausible, but it does not fully describe the nature of the harms they can cause.<sup>58</sup> As I have already mentioned, for Schulz self-respect is the sense that one is "a moral equal, with the same rights and duties that all other persons have"; crucially, it is a certain kind of perception of *self*, rather than of others.<sup>59</sup> This contrasts with the key concept in my account, *dignity*. As we have seen, dignity is a social status requiring not merely *possession* of rights and entitlements, but others' *recognition* of one's rights-bearing status. Accordingly, part of my case against honorific representations is that they can undermine this recognition: they can erode the commitment of members of society and government actors to the rights and entitlements of the disadvantaged. In this sense, my account focuses not on the *self-perception* of members of oppressed groups, but on the attitudes of *others* toward them. I have also argued that honorific representations not only undermine dignity, but the *confidence* of members of oppressed groups that they will be treated "with dignity." This line of argument, like Schulz's, focuses on the attitudes of members of oppressed groups; but whereas Schulz's account calls our attention to the effects on their *self-perceptions*, my account underscores their perceptions of *others*, and in particular their perceptions of society and government's commitment to their rights and entitlements. In short, my point is that honorific representations may affect not only how disadvantaged people view themselves but how others view them, in terms of recognizing their dignity; how they view others, in terms of their confidence of this recognition; and how these effects may impair their ability to operate in the social world. Finally, Schulz's emphasis on *equality* in his definition of self-respect leads him to argue that the failure of society to create conditions of self-respect can wrong not only the members of oppressed groups, but former and current oppressors. For example, he argues that statues of Confederate generals, while perhaps a source of *self-esteem* for neo-Confederates, are not a source of self-respect, since they prevent neo-Confederates from seeing themselves as *equals* in the moral community.<sup>60</sup> Similarly, Burch-Brown focuses on honorific infrastructure's effect on the self-perceptions of both the marginalizers and the marginalized: "Living in cities

57 Schulz, "Must Rhodes Fall?" 172.

58 Schulz plausibly argues that the duty to provide the conditions of self-respect may entail a duty *not only* to remove or recontextualize existing honorific representations, but to *add* new representations aimed at promoting self-respect ("Must Rhodes Fall?" 174–76). While this point goes beyond the scope of this paper, the importance of promoting the public good of assurance described by Waldron provides additional support for this positive duty.

59 Schulz, "Must Rhodes Fall?" 172–76.

60 Schulz, "Must Rhodes Fall?" 173–74.

with innumerable landmarks named after people who kept enslaved laborers has the effect of leading people to internalize racially hierarchical social images.”<sup>61</sup> By contrast, since my account concerns honorific representations’ effects on the dignity and assurances of members of oppressed groups, I do not claim that they have similar negative effects on members of socially dominant groups.<sup>62</sup>

I have argued that honorific representations are morally objectionable because they undermine historically or currently oppressed group members’ dignity and assurances of respect for their rights. Along with the harms identified by Schulz, Timmerman, and Burch-Brown—psychological distress, degradation, and harms to self-respect—these reasons seem to provide very strong moral support for not leaving such representations unmodified. Thus, there is a strong presumption in favor of modifying an honorific representation if (a) the subject of the representation endorsed the subordination of groups that are presently or were recently subject to oppression, discrimination, or systematic disadvantage; (b) the subject committed acts that contributed to or helped constitute the past or present oppression of such groups; or (c) the representation conveys a degrading ideology. However, I have not argued that singly or taken together, these reasons always ground an all-things-considered duty to modify these representations. Moreover, even when the *presumption* in favor of modification grounds an all-things-considered duty to modify, we can modify a representation in at least two ways: by removing an honorific representation from public view, or by recontextualizing it in some way. In the next section, I discuss the considerations that ought to guide deliberation about these options.

### 3. RECONTEXTUALIZATION OR REMOVAL: A BALANCING TEST

I have argued that honorific representations can be morally objectionable for a host of reasons, creating a strong presumption in favor of their modification. But modification can take at least two forms: making them inaccessible to the public, or recontextualizing them. I argue in this section that whether we ought to opt for one or the other option is a function of at least ten moral and pragmatic factors. I develop an account of how these factors ought to be weighed, and I defend this “balancing” approach against a number of objections.

61 Burch-Brown, “Is It Wrong to Topple Statues and Rename Schools?” 70.

62 All of that said, I do not deny that the notions of moral equality and rights are deeply connected. James Griffin argues that equal respect is “a value the content of which is itself best expressed in terms of” rights (*Well-Being*, 234). It follows from this that a plausible substantive theory of moral rights will assign the same rights, and rights of the same strength, to most human beings (with some exceptions at the margins). Cf. Griffin, *Well-Being*, ch. 11.

First, let me say a bit more about how I define the two options. “Removing a representation from public view” means making it inaccessible to the public, and it is “inaccessible” if it is either destroyed or turned over to the care of private individuals who do not hold it open for members of the public to view *as* members of the public. If an honorific representation is in a private collection that is not intended by its owner to be viewed by the public *as* members of the public—as opposed to, say, *as* friends—then it is inaccessible to the public in my sense. But if it is in a private museum, then it is accessible to the public whether or not the museum charges admission.

“Recontextualization” refers to the act of *commenting upon* a semantic object by changing its context. Such commentary interacts with the meanings of the object in two primary ways: foregrounding and generating. To *foreground* a meaning is to *take note* of one of the object’s meanings with the aim of making that meaning another person’s *primary interpretation* of the object. To *generate* a meaning is to produce an interpretation of the object that was not heretofore part of its public meaning. It seems clear that recontextualization can only directly influence public meaning, since intended and implicit meanings are determined by the attitudes of the representations’ creators. However, we have also seen that public meaning can be influenced in various ways by intended and implicit meanings. The latter, then, are some of the contextual factors that contribute to public meaning, and recontextualization can alter the quality and degree of their contribution to public meaning. Another contextual factor that helps determine public meaning is what I will call *semantic* context, the set of semantic entities that is presented or juxtaposed with the original object. These semantic entities can be texts, but they can also be other representations or symbols. A third contextual factor is *institutional* context. This refers to the institutional setting of a semantic object, where an “institution” is a complex of positions, roles, norms, and values lodged in particular types of social structures and organizing relatively stable patterns of human activity.<sup>63</sup> Unlike in the case of intended and implicit meaning, recontextualization can directly alter the semantic and institutional contexts of honorific representations.<sup>64</sup>

Hence, recontextualization can be targeted at different contextual factors, including implicit and intended meaning, semantic context, and institutional context. In general, we can say that some act of recontextualization will vary in “strength” with the number of contextual factors it targets and the degree to which it alters them. A form of recontextualization could slightly alter only the

63 For a defense of this definition, see Turner, *The Institutional Order*, 6.

64 The question of the *degree* of public access to institutions is among the factors relevant to determining the kind of recontextualization that ought to be adopted.

representation's semantic context with the addition of some text. For example, in 2018 the Bristol City Council proposed adding a plaque to the Colston statue describing Colston's role in the Atlantic slave trade.<sup>65</sup> This act of recontextualization would foreground a certain public interpretation of the monument according to which it symbolizes both the role of the slave trade in Bristol's history and the obscuring of that role by means of honorific representations like the statue itself. A stronger form of recontextualization can aim at deeper alterations of the representation's semantic context, including the use of multiple types of semantic objects. For example, in Pretoria, South Africa, there is an equestrian statue of South Africa's first prime minister, Louis Botha, that is now paired with a monumental statue of Nelson Mandela, erected in 2013.<sup>66</sup> To the extent that the two statues are intended to semantically interact, this act of recontextualization seems to be a form of foregrounding aimed at making a certain interpretation of the original object salient.

Often, the classic form of recontextualization through alteration of institutional context—removal of the representation to a museum—also involves alteration of semantic context and making explicit the implicit and intended meanings of the representation. The Jefferson Davis statue that the University of Texas at Austin removed from its campus in 2015 is now housed in the university's Dolph Briscoe Center for American History, where a permanent exhibit tells the story of how the statue came to be and why it was eventually removed.<sup>67</sup> The curator of the exhibit stated that the presence of the statue in an educational exhibit, as opposed to a place of honor, underlines that Davis's ideas and actions are no longer commemorated by the university. This example clearly shows how recontextualization can perform multiple functions: both *commenting* on the meaning of the object by *foregrounding* its intended and implicit meanings, and discursively *generating* moral distance between the object's implicit or intended values and those of the institution. Indeed, the general ethical aims of recontextualization are the generation of this moral distance between society and government, on the one hand, and the representation's objectionable meanings on the other; and the *obstruction* of the representation's tendency to promote all-things-considered admiration for its subject. Hence, strategies of recontextualization must be morally evaluated principally on the basis of how well they achieve these aims.

There are at least ten factors that must be considered when deliberating about

65 See BBC News, "Bristol Slave Trader Edward Colston Proposed Plaque 'Not Impartial.'"

66 This example comes from Demetriou and Wingo, "The Ethics of Racist Monuments," 350.

67 See Crowe, "What Happened When One University Moved a Confederate Statue to a Museum."



whether a particular honorific representation morally ought to be recontextualized or removed, as well as the form such recontextualization should take. These are: (1) whether or not the representation expresses a degrading ideology; (2) the degree to which the representation tends to undermine the dignity of members of oppressed groups; (3) the degree to which the representation tends to undermine their assurances that society and government are committed to the groups' moral rights and constitutional entitlements; (4) the degree to which the representation tends to undermine the *self-respect* of members of oppressed groups; (5) the degree to which the representation supports the dignity of members of oppressed groups; (6) the degree to which the representation *provides* the same kinds of assurances referenced in (3); (7) the degree to which the representation tends to *promote* the self-respect of members of oppressed groups; (8) the extent to which people care, for weighty and morally unobjectionable reasons, about the representation itself or its subject(s); (9) any other morally relevant consequences that might plausibly result from either recontextualization or removal; and (10) the practical feasibility of recontextualization. Of the consequences referenced under criterion 9, I include the psychological distress highlighted by Timmerman, and I would like to examine three in particular: first, the pedagogical benefits that may accrue from proper recontextualization; second, the likelihood that removal will lead to productive or unproductive political or social tension; and third, the contributions that a representation makes to public commitment to important moral values. I will discuss these points in turn. First, I will make a few remarks about how these criteria ought to be used.

With respect to the decision about whether to recontextualize or remove, the criterion of practical feasibility (criterion 10) controls the weighting of the other criteria. There will be any number of cases in which recontextualization is not practically feasible: recontextualization is relatively resource intensive, especially when it involves alteration of institutional context; many public representations have little artistic or art-historical value; and there is little historical reason for a museum to acquire multiple examples of the same type of representation—for example, a hundred Robert E. Lee statues. Nevertheless, if a particular representation can be properly recontextualized within practical constraints, then my proposal is as follows. Since criteria 1–8 are grounded in the fundamental civic values of justice and respect, they are at least much weightier than the sorts of consequences referenced under criterion 9. Indeed, some might claim that the sorts of considerations underlying the former criteria trump or exclude considerations of overall benefit or harm.<sup>68</sup> I will not commit to this stronger position,

68 For the notion of “exclusionary norms,” see Raz, *Practical Reasons and Norms*, 39–40, 73–84, 90–97.

but I will propose that criteria 1–8 are weighted such that only *very substantial* negative or positive consequences (criterion 9) can affect the verdict issuing from consideration of the former criteria. Next, 1–7 should be given roughly equal weight, while less weight should be assigned to 8. The reason for this is that the positive duty to respect others' cares and concerns seems to be less morally weighty than the duty *not* to undermine their self-respect, dignity, or assurances of respect for their rights and constitutional entitlements. In general, and setting aside 8 and 9, if the negative effects referenced in 1–4 significantly exceed the positive referenced in 5–7, then strong forms of recontextualization or removal are morally mandatory. If the reverse is true, then generally speaking moderate recontextualization is morally mandatory. Again, this weighting is for when the particular representation can be feasibly recontextualized. If it cannot, then the case against leaving public representations unmodified will tend to decisively support removal from public view. That is to say, if a representation cannot be feasibly recontextualized, thus “failing” criterion 10, then this means that criteria 1–4 will usually have decisive priority over criteria 5–9: barring extraordinary circumstances, the representation *must be removed* if it tends to undermine the dignity, self-respect, and assurance of respect of members of oppressed groups.

This system is only proposed as a rough set of guidelines for deliberation. Moreover, my balancing test is intended as a guide to determining the moral rightness of removing or recontextualizing a given honorific representation, but it does not take into account that determining what to do with honorific representations is often a *political* exercise, and thus subject to various additional moral requirements on the *procedures* of democratic decision-making. In this respect, I have nothing to add to Schulz's plausible remark that such decision-making must “actively include the voices and perspectives of those who constitute a marginalized group.”<sup>69</sup> In addition, it seems desirable that such public procedures be informed by historically grounded interpretations of the representation's intended and implicit meanings. Finally, it may be *presumptively* desirable that public actions proceed legally and in ways calculated to maximize their perceived legitimacy, but I am not convinced that this is always all-things-considered required.<sup>70</sup>

I have already argued that there is a presumptive moral case against leaving a representation unaltered to the extent that it tends to undermine the dignity of members of oppressed groups and their assurances that society and government are committed to their rights and entitlements. Schulz adds that representations can undermine self-respect. But these tendencies are not all or nothing: they

69 Schulz, “Must Rhodes Fall?” 174.

70 Cf. Demetriou and Wingo, “The Ethics of Racist Monuments,” 349.

come in degrees, and there is clearly no easy way to measure them. It is clear, for example, that many African Americans object strongly to Confederate monuments; this much can be gleaned from popular culture, journalism, political activism, and the behavior of politicians and other leaders who represent significant African American constituencies. Referenda, public forums, polls, and the like might also be used to get a more accurate picture of the attitudes of members of oppressed groups. One important point is that these attitudes can change over time, particularly as information about the subjects of honorific representations becomes more widely known. In turn, this ought to prompt a reassessment of the appropriateness of the representation's existence in public space.

Complicating matters further, the *same* honorific representation may *promote* dignity and self-respect or *provide* assurances to one oppressed group that society and government will respect their rights while *undermining* the dignity, self-respect, or assurances of another group. Consider a statue of the feminist Margaret Sanger to which disabled rights activists object on the grounds that she endorsed wrongful eugenic practices, but which is viewed by many women as a public affirmation of their rights. These countervailing tendencies must be weighed in our deliberation about what to do with such representations. If possible, it may be morally preferable to *replace* an objectionable statue that has these countervailing tendencies with one that is less objectionable but equally affirming; for example, a statue of Margaret Sanger might be replaced with a statue of a non-eugenicist feminist.<sup>71</sup>

In the South, calls to remove Confederate monuments from the public space have provoked fierce opposition from large segments of the public, who perceive these demands as an implicit rejection of a morally unobjectionable Southern heritage and culture.<sup>72</sup> Whether or not this is true in the case of Confederate monuments, there is a more general point to be made.<sup>73</sup> That the fate of these honorific representations can become entangled with people's sense of self, history, and place raises an important ethical concern about destroying them. We have a duty to respect every person, and this plausibly entails that we have a duty not to express disrespect for the objects they care about provided that their reasons for caring about them are not themselves morally objectionable. Therefore, to the extent that people care about honorific representations because of their

71 Thanks to a reviewer at the *Journal of Ethics and Social Philosophy* for alerting me to this point.

72 For an argument that the notion of "Southern honor" can be disentangled from racial oppression and racism in this way, see Schedler, "Are Confederate Monuments Racist?"

73 For another discussion of this point, see Burch-Brown, "Is It Wrong to Topple Statues and Rename Schools?" 84–85.

connection to their history, identity, and (non-objectionable) values, we have reason to treat these representations as objects worthy of care. Minimally, treating an object as worthy of care requires refraining from destroying it or making it inaccessible to those who care about it. Therefore, we have a duty not to wholly remove these objects from public view. This defeasible duty constitutes the eighth factor in our balancing test. As I already mentioned, since this is a merely *pro tanto* duty of respect, it may be outweighed by the wrongs or injustices that recontextualization or removal are aimed at correcting. As Burch-Brown writes,

in cases where there are deep and legitimate (i.e. not inherently unjust) attachments to the cultural objects in question, it is important to ask searchingly whether the injustice in question can be redressed in ways that protect and affirm the legitimate parts of people's positive attachments and identities. . . . The political risks involved, the expenditure of energy, and the costs to others (for instance, the cultural losses or threats to personal identity and so on) should be proportional to the injustice that is correctable through the measures proposed.<sup>74</sup>

Another argument sometimes made in favor of leaving honorific representations unmodified is that removal of these monuments amounts to "erasing history."<sup>75</sup> This argument can be cashed out in a number of ways. One version says that honorific representations educate the public about important individuals and events in the community's past, so removing them contributes to the public's historical ignorance.<sup>76</sup> However, cases such as the Heyward Shepherd Memorial, which grossly distorts the historical record, suggest that many of these representations are more likely to mislead the public than to educate it. This is a natural consequence of the fact that the primary function of these representations is to express and cultivate admiration, rather than to educate. Nevertheless, these representations may have significant pedagogical value *if properly recontextualized*: the distortions, elisions, and implicit evaluative claims that proper recontextualization can reveal help teach us about the political, economic, social, and moral struggles that continue to roil society, while at the same time bringing to the public's awareness the varied experiences of diverse historical actors. This is particularly important if, as some philosophers claim, society has a collective responsibility to preserve the memory of past atrocities to reduce the likelihood

74 Burch-Brown, "Is It Wrong to Topple Statues and Rename Schools?" 85.

75 This argument is also addressed by Timmerman, "A Case for Removing Confederate Monuments"; Schulz, "Must Rhodes Fall?"; and Burch-Brown, "Is It Wrong to Topple Statues and Rename Schools?"

76 Cf. Demetriou and Wingo, "The Ethics of Racist Monuments," 348.

of their repetition and to help promote relations of respect in the present.<sup>77</sup> The pedagogical argument, then, actually favors recontextualizing honorific representations for the purpose of preserving the historical record—in particular, the record of society’s failures to uphold the dignity of some of its members. Crucially, this pedagogical value would be lost if we opted to make the representations inaccessible to the public. Hence, the pedagogical value of honorific representations primarily weighs in favor of some kind of recontextualization and against *either* leaving them unmodified or removing them from public view.

Partly because honorific representations can become bound up with people’s sense of identity, history, and culture, and partly because they embody values and ideals with which at least a portion of the population strongly identifies at any given time, campaigns for their removal or strong forms of recontextualization can be politically polarizing. For example, in response to moves by Memphis officials to remove statues of Nathan Bedford Forrest and Jefferson Davis from public parks, Tennessee passed legislation in 2018 that would allow any group or individual with an interest in a memorial to seek an injunction against the public entity involved in the memorial’s removal.<sup>78</sup> More dramatically, and as noted previously, the “Unite the Right” rally in Charlottesville that ultimately led to deadly rioting was organized specifically to protest the proposed removal of a Robert E. Lee statue.<sup>79</sup> That political or social tension is a likely outcome of efforts to reshape the public space is itself a fact of ethical significance, since it may lead to an intensification of social tensions and ultimately serve to perpetuate the very kinds of ongoing historical injustice that campaigners wish to combat.

Burch-Brown argues that the fact that social tensions are likely to result from removal campaigns does not necessarily weigh against them:

This is because, as Frederick Douglass and many others have observed, power never gives up power willingly. Shifts in power will only take place when the discomfort of maintaining the status quo becomes greater than its benefits. Indeed, the purpose of the 1960s direct action campaigns was to generate enough social tension as to open the door to negotiation.<sup>80</sup>

77 For a discussion of the nature, justification, and orientation of such responsibility, see Radzik, “Historical Memory as Forward- and Backward-Looking Collective Responsibility”; see also Adorno, “The Meaning of Working through the Past.”

78 See Ebert, “Legislation in Response to Memphis’s Confederate Statue Removal Signed by Gov. Haslam.”

79 See Morlin, “Bickering Galore Precedes ‘Unite the Right’ Rally.”

80 Burch-Brown, “Is It Wrong to Topple Statues and Rename Schools?” 83.

Burch-Brown's point is that by challenging the social practices and meanings connected to monuments and other symbols that undergird unjust social arrangements, campaigns for their removal can generate social tension, which can in turn be productive of progressive social change under the right circumstances. Similarly, Johannes Schulz argues that campaigns for removal may be particularly effective when a wider social process of working through the unjust past to which the honorific representation is connected does not yet exist.<sup>81</sup> Schulz writes, for example, that in the context of South Africa's post-apartheid "amnesia" concerning that system's lingering effects, the campaign to remove the statue of Cecil Rhodes from the University of Cape Town's campus "triggered one of the most vivid and visible debates about its own past that South Africa has engaged in since the apartheid regime fell."<sup>82</sup> I agree with Burch-Brown and Schulz that the raising of social tension that may attend removal or strong recontextualization campaigns is not invariably a mark against them. Nevertheless, it is important to distinguish here between the ethics of campaigning for removal and the ethics of removal. Schulz may be right that publicly campaigning for removal can sometimes trigger politically productive discussion, since it means pushing our unjust past out into the "open arena of the public sphere, where [it] become[s] a matter of public debate and political struggle."<sup>83</sup> It does not follow from this that the *actual* removal of a representation is the most politically productive outcome.

Whether or not strong recontextualization or removal are likely to lead to productive or unproductive tension is clearly an empirical question, and writers on this topic seem not to agree on what the available evidence tells us. For example, Dan Demetriou and Ajume Wingo object to the removal of racist monuments primarily on the grounds that it will lead to politically unproductive tension, undermining what they call "civic sustainability":

Efforts to cleanse the landscape of racist monuments are unacceptably damaging to civic cohesion and ultimately frustrate antiracist goals. Widespread removalism will tend to "resurrect" forgotten monuments, confirm the suspicions of white separatists and nationalists [e.g., that they are being culturally "replaced" by non-whites], and lend credence to [some radical removalists'] belief that the national identity of these lands is inextricably based on white hegemony.<sup>84</sup>

81 Schulz, "Must Rhodes Fall?" 177–83.

82 Schulz, "Must Rhodes Fall?" 183.

83 Schulz, "Must Rhodes Fall?" 179.

84 Demetriou and Wingo, "The Ethics of Racist Monuments," 352.

In addition to these worries, recall the Margaret Sanger example discussed earlier. If removal violates a defeasible duty of respect, and if many people care about the Sanger statue for non-objectionable reasons, then the removal of the statue may be unlikely to win the kind of broad-based support of those whose votes must be won in order to bring about change in the existing legal and political order. Hence, it is difficult to see how this instance of removal could contribute to the broader amelioration of social injustices. Along with Demetriou and Wingo, I believe removal will often be unhelpful in furthering larger political and social goals. But two points need to be emphasized. First, even Demetriou and Wingo do not support leaving racist monuments unmodified; instead, they seem to favor moderate recontextualization that leaves these monuments *in situ*.<sup>85</sup> Second, I break with them by insisting that considerations of justice and respect are *not* on a moral par with “civic sustainability.”<sup>86</sup> Cleansing our public spaces of representations that degrade or undermine dignity, self-respect, and assurances of respect for rights might be worth the cost of some short-to-medium-term damage to civic cohesion. That does not mean that *any* degree of the former kind of disvalue can outweigh even the most extreme threats to the latter value; but the claim that we should tolerate monuments that to some degree strike at the heart of innocent people’s self-respect, dignity, and assurance of rights because removal might validate white supremacist narratives seems like a moral miscalculation. Moreover, when the public landscape is littered with such representations, it is likely that civic cohesion is *already* relatively weak. This is why Demetriou and Wingo’s analogy of an interracial couple decorating their home misses the mark.<sup>87</sup> A person might tolerate her partner’s hanging a picture of a racist ancestor for the sake of domestic tranquility, but that is *because* the couple already enjoys a relationship of trust and respect, a sort of well-ordered society in microcosm. The citizens of a flawed democracy such as the United States are in a very different position, still unsure of other citizens’ basic commitment to their rights and constitutional entitlements. In this far-from-ideal political order, members of oppressed groups are not necessarily obliged to tolerate others’ veneration for racist figures in the name of their “heritage.”

A final consideration for our balancing test is an honorific representation’s contribution to the public’s commitment to important moral values. Consider the Jefferson Memorial in Washington, DC, which features a bronze statue of Thomas Jefferson along with a number of passages from Jefferson’s letters and the Declaration of Independence. These passages emphasize his commitment to

85 Demetriou and Wingo, “The Ethics of Racist Monuments,” 350–52.

86 Demetriou and Wingo, “The Ethics of Racist Monuments,” 349.

87 Demetriou and Wingo, “The Ethics of Racist Monuments,” 351–52.

equality, freedom from tyranny, and freedom of religion, as well as his opposition to slavery. This last quotation comes from a 1786 letter to George Washington that reads: “Commerce between master and slave is despotism. Nothing is more certainly written in the book of fate than these people are to be free.” However—and this part is left off of the monument—Jefferson continued, “nor is it less certain that the two races, equally free, cannot live in the same government.” Jefferson’s treatment of Sally Hemings is now well known, as is the fact that he owned slaves and did not free many of them during his lifetime. Nonetheless, one will not find any reference to Jefferson’s ownership and treatment of slaves or his views about African Americans on the National Park Service’s web page for the monument.<sup>88</sup> In encouraging *unqualified admiration* for a person who, indeed, did support a white supremacist ideology, and by eliding his involvement in the subordination of African Americans, the monument makes it more difficult for citizens to recognize the ways in which white supremacy was conjoined to the American project. And that could undermine African Americans’ assurance that government and society are committed to their basic rights and undermine their dignity. Hence, this representation meets my presumptive standard for modification. However, many people see Jefferson as a symbol of the highest American ideals of freedom, equality, and democracy, and the Jefferson Memorial clearly presents him as the philosophical wellspring from which America’s articulation of and commitment to these ideals flowed. Martha Nussbaum has argued for the indispensable importance of powerful and positive emotions, and particularly love, in cementing citizens’ attachments to moral and political ideals.<sup>89</sup> It is conceivable that by eliciting powerful emotional responses, monumental representations of figures like Jefferson can help foster such attachments. And this may, in turn, constitute a reason *not* to remove such representations from public view.

There are, then, many factors to weigh when deciding what is to be done with a given honorific representation. This conclusion is more cautious than some other philosophers’ claims. For example, Joanna Burch-Brown’s inference from the claim that it is inappropriate to honor and esteem people who have carried out grave injustices to the conclusion that statues of those people should be removed—in the sense of “removal” meaning “not left *in situ*”—suggests that she intends “inappropriate” as an all-things-considered prohibition.<sup>90</sup> But given the many factors that may weigh in favor of recontextualization *in situ*, it seems plausible that sometimes the moral reasons not to honor and esteem the subject of an honorific representation *do not* all things considered require removal, even

88 See National Park Service, “Thomas Jefferson.”

89 Nussbaum, *Political Emotions*.

90 Burch-Brown, “Is It Wrong to Topple Statues and Rename Schools?” 68.



if not removing the representation means retaining some of its honorific function. For example, even if the subject of an honorific representation committed a grave injustice, it is possible that this representation does not *currently* cause much harm, and it is also possible that the moral good the subject did counterbalances the evil. It is even conceivable, though exceedingly unlikely, that we ought all things considered to leave the statue unmodified.<sup>91</sup>

To illustrate this point, and how my balancing test might be implemented in a particular case, let us consider the Jefferson Memorial in light of all the factors highlighted above. This exercise will also help explain *why* the Jefferson Memorial seems like a “hard case.” As mentioned, the memorial easily meets my presumptive case for modification. However, unlike Confederate monuments, the Jefferson monument may not *express* a degrading ideology, and it may be the case that it undermines African Americans’ self-respect, dignity, and assurance of respect for their rights to a lesser degree. In addition, there are clear pedagogical benefits of proper recontextualization, there is the potential to inculcate positive emotional support for important moral values, and there is a high likelihood that removal will lead to politically or socially unproductive tension. That is at least partly because Jefferson himself is a morally mixed character. His ringing denunciations of slavery, for example in his *Notes on the State of Virginia*, were morally compromised by his personal behavior, failure to take more practical steps toward the elimination of slavery, and racist views. In the *Notes*, he echoed the common view that Blacks possessed qualities that make human beings worthy of “esteem-respect” to a much lesser degree than whites.<sup>92</sup> Yet he denied that this undermined their basic moral equality.<sup>93</sup> Ambivalence, rather than unalloyed condemnation, might be the proper response to him.<sup>94</sup> Thus, it seems to me that at this time, the morally right course of action is to recontextualize the monument to a moderate degree. An effective solution might be to add a substantial piece of public art adjacent to or within the monument that

91 I am thinking here of a standard evil demon case; e.g., an evil demon threatens to kill every African American unless a Lee statue remains unmodified.

92 By “esteem-respect” I mean what Darwall calls “appraisal respect”: an attitude of positive appraisal for a person’s achievements or traits. See Darwall, “Two Kinds of Respect,” 36–49.

93 The historiography on Jefferson’s views about slavery and race is enormous. For three strikingly divergent views, see Magnis, “Thomas Jefferson and Slavery”; Dawidoff, “The Jeffersonian Option”; and Holowchak, *Framing a Legend*.

94 This point once again raises the issue of our aversion to ambivalence. If the ambivalence research is correct, then the public will likely not *enjoy* an ambivalent Jefferson Memorial as much as the current unqualifiedly admiring version. Nevertheless, this loss of satisfaction is morally preferable to allowing the monument *itself* to contribute to unqualified admiration for a morally ambivalent figure.

acknowledges Jefferson's contradictions, along with descriptions of these problematic aspects on the National Park Service's website.

#### 4. CONCLUSION

In this paper, I have considered whether we ought to leave honorific representations unmodified, remove them from public view, or recontextualize them in various ways. I argued that in many cases honorific representations should not be left unmodified in the public space because they can undermine oppressed group members' dignity and assurance of respect for their rights and civic entitlements. I argued that a balancing test is the appropriate way to think about whether a particular representation ought to be removed or recontextualized. As we have seen, the term "recontextualization" captures a wide variety of strategies of differing "strength" aimed at commenting on a representation by changing its context. I doubt that a single strategy is appropriate for all cases. Nevertheless, in what follows I want to suggest reasons why strategies of minimal recontextualization may be an insufficient remedy for the problems I have described.

One problem with recontextualization strategies involving only small changes in semantic context, such as the addition of signage, is that they will not necessarily prevent people intent on rallying the forces of hatred and bigotry from using these representations as focal points for their demonstrations. I have argued that such demonstrations serve to undermine the dignity of members of oppressed groups and deprive them of the assurance that others will recognize their rights. In addition, there is at least a potential that such strategies will merely convey to the public a sense of moral ambivalence rather than solid condemnation. The juxtaposition of a clearly exemplarizing honorific representation with signage that emphasizes the morally problematic aspects of the subject's traits, deeds, and beliefs may prompt members of the public to infer that society and government are evenly divided between admiration and condemnation of the figure, or that society and government consider both attitudes of admiration and contempt equally legitimate.<sup>95</sup> Sometimes, as in the case of the Jefferson Memorial, ambivalence might be the all-things-considered appropriate response to the subject, and might be the response we should encourage the public to experience when considering him or her. In other cases, such as some Confederate monuments, we probably ought to encourage a more negative overall response.

These considerations suggest that in many cases, alteration of honorific representations' institutional contexts will be required. Of course, such moves do

95 For a discussion of "dual heritage" strategies in the American South, see Upton, *What Can and Can't Be Said*; and Labode, "Reconsideration of Memorials and Monuments."

not guarantee the amelioration of their morally problematic effects. As Janeen Bryant et al. note, “statues remain powerful—and physically imposing—visual forms that will keep speaking even when they are in new settings. They can and certainly will shape social experiences in ways that [museum] curators may not be able to anticipate.”<sup>96</sup> The broader point is we ought to be cautious about educational institutions’ ability or willingness to address the disturbing meanings of these representations; “just put it in a museum” is not a strategy. Nevertheless, these legitimate worries should not deter museums and other educational institutions from taking up the challenge. As I have argued, sometimes much rides on their success.

Tulane University  
brossi@tulane.edu

#### REFERENCES

- Adorno, Theodor W. “The Meaning of Working through the Past.” In *Critical Models: Interventions and Catchwords*, 89–104. New York: Columbia University Press, 1998.
- Appelbaum, Yoni. “Why Is the Flag Still There?” *The Atlantic*, June 21, 2015. <https://www.theatlantic.com/politics/archive/2015/06/why-is-the-flag-still-there/396431/>.
- Archer, Alfred, and Benjamin Matheson. “When Artists Fall: Honouring and Admiring the Immoral.” *Journal of the American Philosophical Association* 5, no. 2 (April 2019): 246–56.
- BBC News. “Bristol Slave Trader Edward Colston Proposed Plaque ‘Not Impartial,’” July 24, 2018. <https://www.bbc.com/news/uk-england-bristol-44937719>.
- . “Will ‘Rhodes Must Fall’ Fail?” January 18, 2016. <https://www.bbc.com/news/blogs-trending-35328444>.
- Brait, Ellen. “Princeton Students Demand Removal of Woodrow Wilson’s Name from Buildings.” *The Guardian*, November 23, 2015. <https://www.theguardian.com>.
- 96 Janeen Bryant et al., “Are Museums the Right Home for Confederate Monuments?” Illustrative of this problem is the newly opened American Civil War Museum’s handling of a twelve-by-eight-foot painting depicting Lee and Jackson, one of the largest items in the museum’s collection. The museum displays the painting alongside the text of the Thirteenth, Fourteenth, and Fifteenth Amendments; and on the wall immediately facing the painting is a large, colorized photograph of African American legislators dating to the Reconstruction period. See Connelly, Ayers, and Freeman, “The Civil War in the 21st Century.”

- com/education/2015/nov/23/princeton-woodrow-wilson-racism-students-remove-name.
- Brannon, Skylar M., and Bertram Gawronski. "Cognitive Consistency in Social Cognition." *Oxford Research Encyclopedia of Psychology*, March 28, 2018. <https://oxfordre.com/psychology/view/10.1093/acrefore/9780190236557.001.0001/acrefore-9780190236557-e-314>.
- Bryant, Janeen, Benjamin Filene, Louis Nelson, Jennifer Scott, and Suzanne Seriff. "Are Museums the Right Home for Confederate Monuments?" *Smithsonian Magazine*, May 7, 2018. <https://www.smithsonianmag.com/history/are-museums-right-home-confederate-monuments-180968969/>.
- Burch-Brown, Joanna. "Is It Wrong to Topple Statues and Rename Schools?" *Journal of Political Theory and Philosophy* 1, no. 1 (2017): 59–88.
- Confederate Southern Memorial Association, *Confederate Veteran*, vol. 2. Nashville, TN: S.A. Cunningham, 1894. <https://archive.org/stream/confederatevetero2conf#page/336/mode/2up>.
- Connelly, Nathan, Ed Ayers, and Joanne Freeman. "The Civil War in the 21st Century: A New Museum Marks an Old Conflict." Podcast audio. *BackStory*, May 3, 2019. <https://www.backstoryradio.org/shows/the-civil-war-in-the-21st-century>.
- Crowe, Cailin. "What Happened When One University Moved a Confederate Statue to a Museum." *The Chronicle of Higher Education*, September 10, 2018. <https://www.chronicle.com/article/what-happened-when-one/244481>.
- Darwall, Stephen. "Two Kinds of Respect." *Ethics* 88, no. 1 (October 1977): 36–49.
- Dawidoff, Robert. "The Jeffersonian Option." *Political Theory* 21, no. 3 (August 1993): 434–52.
- Demetriou, Dan, and Ajume Wingo. "The Ethics of Racist Monuments." In *Palgrave Handbook of Philosophy and Public Policy*, edited by David Boonin, 341–55. New York: Palgrave Macmillan, 2018.
- Dion, Karen, Ellen Berscheid, and Elaine Walster. "What Is Beautiful Is Good." *Journal of Personality and Social Psychology* 24, no. 3 (1972): 285–90.
- Eagly, Alice, Richard D. Ashmore, Laura C. Longo, and Mona G. Makhijani. "What Is Beautiful Is Good, but . . . : A Meta-Analytic Review of Research on the Physical Attractiveness Stereotype." *Psychological Bulletin* 110, no. 1 (January 1991): 109–28.
- Ebert, Joel. "Legislation in Response to Memphis's Confederate Statue Removal Signed by Gov. Haslam." *Tennessean*, May 22, 2018. <https://www.tennessean.com/story/news/politics/2018/05/22/governor-signs-measure-bolstering-heritage-protection-act-into-law/565755002/>.

- El Amin, Alvin Nelson, Michelle T. Parra, Robert Kim-Farley, and Jonathan E. Fielding. "Ethical Issues Concerning Vaccination Requirements." *Public Health Review* 34, no. 1 (June 2012): 1–20.
- Elliot, Andrew, and Patricia Devine. "On the Motivational Nature of Cognitive Dissonance: Dissonance as Psychological Discomfort." *Journal of Personality and Social Psychology* 67, no. 3 (September 1994): 382–94.
- Erlich, Danuta, Isaiah Guttman, Peter Schönbach, and Judson Mills. "Postdecision Exposure to Relevant Information." *Journal of Abnormal and Social Psychology* 54, no. 1 (January 1957): 98–102.
- Festinger, Leon. *A Theory of Cognitive Dissonance*. Evanston, IL: Row, Peterson, and Company, 1957.
- Finkelman, Paul. *Supreme Injustice: Slavery in the Nation's Highest Court*. Cambridge, MA: Harvard University Press, 2018.
- Gates, Henry Louis, Jr. *Stony the Road: Reconstruction, White Supremacy, and the Rise of Jim Crow*. New York: Penguin Press, 2019.
- Gräf, Michael, and Christian Unkelbach. "Halo Effects in Trait Assessment Depend on Information Valence: Why Being Honest Makes You Industrious, but Lying Does Not Make You Lazy." *Personality and Social Psychology Bulletin* 42, no. 3 (March 2016): 290–310.
- Greenberg, Allan. "With Meaning for All: The Lincoln Memorial in American Consciousness." *City Journal*, Spring 2013. <https://www.city-journal.org/html/meaning-all-13552.html>.
- Griffin, James. *Well-Being: Its Meaning, Measurement, and Moral Importance*. New York: Oxford University Press, 1986.
- Harmon-Jones, Eddie. "Cognitive Dissonance and Experienced Negative Affect: Evidence That Dissonance Increases Experienced Negative Affect Even in the Absence of Aversive Consequences." *Personality and Social Psychology Bulletin* 26, no. 12 (December 2000): 1490–1501.
- Haslanger, Sally. "Racism, Ideology, and Social Movements." *Res Philosophica* 94, no. 1 (January 2017): 1–22.
- Holowchak, Mark. *Framing a Legend: Uncovering the Distorted History of Thomas Jefferson and Sally Hemings*. Amherst, NY: Prometheus Books, 2013.
- Janda, Lance. "Shutting the Gates of Mercy: The American Origins of Total War, 1860–1880." *The Journal of Military History* 59, no. 1 (January 1995): 7–26.
- Johnson, Mary. "An 'Ever Present Bone of Contention': The Heyward Shepherd Memorial." *West Virginia History* 56 (1995): 1–26.
- Kelly, Cindy. *Outdoor Sculpture in Baltimore: A Historical Guide to Public Art in the Monumental City*. Baltimore, MD: Johns Hopkins University Press, 2011.

- Labode, Modupe. "Reconsideration of Memorials and Monuments." *History News* 71, no. 4 (November 2016): 7–12.
- Lakoff, George, and Mark Johnson. *The Metaphors We Live By*. Chicago: University of Chicago Press, 1980.
- Langlands, Rebecca. *Sexual Morality in Ancient Rome*. New York: Cambridge University Press, 2006.
- Langlois, Judith H., Lisa Kalakanis, Adam J. Rubenstein, Andrea Larson, Monica Hallam, and Monica Snoot. "Maxims or Myths of Beauty? A Meta-Analytic and Theoretical Review." *Psychological Bulletin* 126, no. 3 (May 2000): 390–423.
- Leib, Jonathan, and Gerald Webster. "On Remembering John Winberry and the Study of Confederate Monuments on the Southern Landscape." *Southeastern Geographer* 55, no. 1 (Spring 2015): 8–18.
- Magnis, Nicholas. "Thomas Jefferson and Slavery: An Analysis of His Racist Thinking as Revealed by His Writings and Political Behavior." *Journal of Black Studies* 29, no. 4 (March 1999): 491–509.
- Moore, Michael. "What the Removal of New Orleans's White Supremacist Monuments Means to My Students." *Artsy*, May 15, 2017. <https://www.artsy.net/article/artsy-editorial-removal-new-orleanss-white-supremacist-monuments-students>.
- Morlin, Bill. "Bickering Galore Precedes 'Unite the Right' Rally." Southern Poverty Law Center, August 3, 2017. <https://www.splcenter.org/hatewatch/2017/08/03/bickering-galore-precedes-%E2%80%9CUnite-right%E2%80%9D-rally>.
- National Park Service. "Thomas Jefferson." March 30, 2018. <https://www.nps.gov/thje/learn/historyculture/thjebio.htm>.
- Newby-Clark, Ian, Ian MacGregor, and Mark P. Zanna. "Thinking and Caring about Cognitive Inconsistency: When and for Whom Does Attitudinal Ambivalence Feel Uncomfortable?" *Journal of Personality and Social Psychology* 82, no. 2 (2002): 157–66.
- Nicholson, Amber. "Robert E. Lee Monument." New Orleans Historical. <http://www.neworleanshistorical.org/items/show/1279>.
- Nordgren, Loran, Frenk van Harreveld, and Joop van der Pligt. "Ambivalence, Discomfort, and Motivated Information Processing." *Journal of Experimental Social Psychology* 42, no. 2 (2006): 252–58.
- Nussbaum, Martha. "Non-Relative Virtues: An Aristotelian Approach." *Midwest Studies in Philosophy* 13, no. 1 (September 1988): 32–53.
- . *Political Emotions: Why Love Matters for Justice*. Cambridge, MA: Harvard University Press, 2013.

- Parks, Pamela. "Edward Colston: The Slave Trader Dividing Bristol." *BBC News*, February 26, 2018. <https://www.bbc.com/news/uk-england-bristol-42404825>.
- Philpott, Ben, and Andrew Weber. "Woodrow Wilson Statue Removal Prompts a Closer Look at His History with Race Relations." *KUT*, August 31, 2015. <http://kut.org/post/woodrow-wilson-statue-removal-prompts-closer-look-his-history-race-relations>.
- Pierik, Roland. "Mandatory Vaccination: An Unqualified Defense." *Journal of Applied Philosophy* 35, no. 2 (May 2018): 381–92.
- Radzik, Linda. "Historical Memory as Forward- and Backward-Looking Collective Responsibility." *Midwest Studies in Philosophy* 38, no. 1 (September 2014): 26–39.
- Rawls, John. *Political Liberalism*. New York: Columbia University Press, 1993.
- Raz, Joseph. *Practical Reasons and Norms*. Oxford: Oxford University Press, 1990.
- Rosen, Gideon. "Skepticism about Moral Responsibility." *Philosophical Perspectives* 18, no. 1 (November 2004): 295–313.
- Salisbury, David. "Should Childhood Vaccination Be Mandatory? No." *British Medical Journal* 344 (May 2012): 18–19.
- Savage, Kirk. *Standing Soldiers, Kneeling Slaves: Race, War, and Monument in Nineteenth-Century America*. Princeton, NJ: Princeton University Press, 1997.
- Schedler, George. "Are Confederate Monuments Racist?" *International Journal of Applied Philosophy* 15, no. 2 (Fall 2001): 287–308.
- Schulz, Johannes. "Must Rhodes Fall? The Significance of Commemoration in the Struggle for Relations of Respect." *The Journal of Political Philosophy* 27, no. 2 (June 2019): 166–86.
- Smits, David D. "The Frontier Army and the Destruction of the Buffalo: 1865–1883." *The Western History Quarterly* 25, no. 3 (Fall 1994): 312–38.
- Southern Poverty Law Center. "SPLC Report: More Than 1,700 Monuments, Place Names, and Other Symbols Honoring the Confederacy Remain in Public Spaces." June 4, 2018. <https://www.splcenter.org/news/2018/06/04/splc-report-more-1700-monuments-place-names-and-other-symbols-honoring-confederacy-remain>.
- Special Commission to Review Baltimore's Public Confederate Monuments. "Lee Jackson Monument, 1948." <https://baltimoreplanning.wixsite.com/monumentcommission/leeandjacksonmonument>.
- Tesser, Abraham, and M. C. Conlee. "Some Effects of Time and Thought on Attitude Polarization." *Journal of Personality and Social Psychology* 31, no. 2 (1975): 262–70.
- Timmerman, Travis. "A Case for Removing Confederate Monuments." In *Ethics*,

- Left and Right: The Moral Issues That Divide Us*, edited by Bob Fischer, 513–22. New York: Oxford University Press, 2020.
- Turner, Jonathan. *The Institutional Order*. New York: Longman, 1997.
- Upton, Dell. *What Can and Can't Be Said: Race, Uplift, and Monument Building in the Contemporary South*. New Haven: Yale University Press, 2015.
- Van Harreveld, Frenk, Iris Schneider, Hannah Nohlen, and Joop van der Pligt. "The Dynamics of Ambivalence: Evaluative Conflict in Attitudes and Decision-Making." In *Cognitive Consistency: A Fundamental Principle in Social Cognition*, edited by Bertram Gawronski and Fritz Strack, 267–84. New York: Guilford Press, 2012.
- Waldron, Jeremy. *The Harm in Hate Speech*. Cambridge, MA: Harvard University Press, 2012.
- Waldron, Jeremy, and Meir Dan-Cohen. *Dignity, Rank, and Rights*. New York: Oxford University Press, 2012.
- Zagzebski, Linda Trinkaus. *Exemplarist Moral Theory*. Oxford: Oxford University Press, 2017.



## KIDNEY EXCHANGE AND THE ETHICS OF GIVING

*Philippe van Basshuysen*

OVER 1.2 MILLION PEOPLE worldwide die of kidney disease each year, making it a little-noticed epidemic, comparable in scale to all deaths by road injuries.<sup>1</sup> Increasing the number of kidney transplants from live or deceased donors would save many lives, and would improve the life expectancy and quality of many more. At present, there is a striking shortage of kidneys for transplantation.

Many patients have willing live donors who cannot donate to their loved ones because they are biologically incompatible. Kidney exchange (KE) promises relief. For example, suppose your partner needs a kidney but you cannot donate because you are incompatible. If the same is true of a different donor-recipient pair, it may be possible for you to donate to the other recipient and your partner to get the other donor's kidney. Moreover, some people decide to donate *altruistically*, that is, they give a kidney to a stranger without receiving anything in return. Their gifts can trigger chains of KE, thus multiplying the benefits from a single donation. Below, a detailed description of different types of KE will be provided.

However, KE frequently meets ethical objections, in particular concerning the protection of live donors. These objections are embodied in the transplant laws in many countries, which prohibit live organ donations to strangers, making the implementation of a broad range of KE procedures illegal. In light of the shortage of kidneys for transplantation, it is an urgent matter to clarify the ethics underlying KE.

This paper aims to do that. It examines the implications for live kidney donations of some weak tenets from the ethics of giving. What I call the *effectiveness principle* is such a tenet. It says, roughly, that if a donation can either be allocat-

1 On kidney disease, see Wang et al., "Global, Regional, and National Life Expectancy, All-Cause Mortality, and Cause-Specific Mortality for 249 Causes of Death, 1980–2015," 1483, 1490. The study estimates the number of deaths by road injuries worldwide at 1.3 million to 1.4 million annually (1491).

ed to benefit some persons, or be allocated to benefit more persons than the first allocation, and the benefits to everyone are approximately identical, then we ought to choose the second allocation. This principle implies that, when an autonomous donor is given a choice between donating into a waiting list or into KE, in many cases morality requires the latter. KE is thus instrumental in meeting a moral obligation, which provides a novel argument for KE.

I also examine possible ethical objections to KE that might block the argument. Throughout, the German transplant law is used as a case study that expresses many of those concerns. I seek to distinguish the objections that are well-founded from those that are not and for the former, discuss their implications for different types of KE. It will be argued that even conservative views on donor protection and distributive justice, while possibly justifying restrictions of some types of KE procedures, do not in principle oppose the implementation of KE programs.

Finally, I consider possible implications of KE for donors' motivations to donate. KE programs use matching algorithms that maximize the number of transplantations; in particular, in their presence, a donation can trigger > 1 life savings. Therefore, KE may increase the motivation for donating, which constitutes a further advantage of KE. Together, these arguments make a robust case in favor of providing a legal framework that allows implementing KE programs.

This paper is organized as follows. Section 1 argues for some weak principles from the ethics of giving. Section 2 explains the rationale for and basic procedures of KE. Section 3 examines the implications of the introduced principles for the allocation of kidneys, which provides an argument for KE. Section 4 considers some of the critics' concerns and argues that they might restrict some procedures of KE, but they do not in principle reject the permissibility of KE programs. Section 5 considers motivational aspects of live kidney donations, which constitute a further argument for KE. Section 6 concludes.

## 1. THE CONDITIONAL OBLIGATION TO DONATE EFFECTIVELY

Before turning to the allocation of kidneys, some basic principles from the ethics of giving will be introduced. The philosophy of effective altruism has recently drawn increased attention to this ethical branch. Effective altruists believe that charitable giving should be done in a way that is most efficient in promoting the most good.<sup>2</sup> Effective altruism raises two questions. First, to what extent (if at all) is there a moral imperative for individuals to give resources to good causes;

2 E.g., Singer, *The Most Good You Can Do*; MacAskill, *Doing Good Better*. The former includes a chapter on altruistic kidney donations.

and second, if someone decides to give some of their resources to a good cause, what does morality imply for their allocation? We will only be concerned with the second, conditional question. To answer it, weak principles suffice that many people who are not effective altruists will find acceptable as well. The following trolley problem, which stems from Theron Pummer, may help to find these principles.<sup>3</sup>

A trolley on a track *A* is headed for one innocent person, and another trolley on a track *B* is headed for one hundred different, innocent persons. Each trolley will kill everyone on its track with certainty—unless you stop it. You can stop one but not both trolleys by laying your arm on the respective track. If you do so, you will lose your arm but everyone on that track will be rescued and nothing else will happen to you.

In line with the above, we make no claim as to whether there is a moral obligation to stop a trolley by sacrificing your arm. Instead, we are interested in the conditional question: If you choose to help, which trolley should you stop? Most will agree that you ought to stop trolley *B*. It would not be permissible to stop trolley *A* because you would only save one life instead of one hundred by bearing the identical cost, namely your arm. Your rescuing would be terribly ineffective. This appears to be a robust moral intuition. For instance, suppose that, instead of sacrificing a limb, you can donate an amount of money to stop one of the trolleys (but not both). Or suppose that your donation (be it money or your arm) will not save the lives of the persons on a track but would merely prevent them from having short, miserable, or diseased lives, where this benefit would roughly be the same for each beneficiary, no matter on which track. The felt conditional obligation to benefit many, rather than few, when the benefit is the same for each person, holds under a broad range of variations of the problem. Accordingly, we shall assume the following principle:

*Effectiveness principle:* If one and the same donation can either be allocated to benefit some persons, or be allocated to benefit more persons than the first allocation, and the benefit is roughly identical for each person under the two possible allocations, then, other things being equal, there is a conditional moral obligation to choose the second allocation.

It is worth emphasizing that the effectiveness principle, while implied by effective altruism, is weaker and therefore less controversial than effective altruism.<sup>4</sup> For instance, effective altruism is demanding in terms of the cause that

3 Pummer, “Whether and Where to Give.”

4 Two referees of this journal kindly point out that some philosophers might nevertheless dis-

you should support. If fighting malaria brings about the most good in the most cost-effective way, then effective altruism may imply that you ought not to support charities fighting homelessness in developed countries. The effectiveness principle would only make this recommendation if the benefits were roughly identical for each beneficiary and the beneficiaries under the former intervention more numerous. But typically, these benefits are unequal for the homeless in developed countries and persons suffering from malaria. Furthermore, effective altruism is typically thought to rely on a consequentialist and welfarist moral theory. The effectiveness principle presupposes neither, as the conditional obligation implicated by it could instead be argued to arise, for instance, from fulfilling more claims to our aid.

The effectiveness principle is equipped with an other-things-being-equal clause in order to accommodate factors that can make a difference in various moral theories. For example, suppose the person on track *A* is a friend or family member. This and other “agent-relative” reasons break the symmetry between the two tracks, and they may justify bringing about outcomes that are less than

---

agree that the effectiveness principle applies in general. While the intricate theoretical discussions about the ethics of giving are beyond the scope of this paper, there is a reason why the effectiveness principle is rather innocuous in the context of altruistic kidney donation. To see why, consider a variation of the trolley problem, in which stopping trolley *B* would rescue all 101 persons. For example, this could be incorporated by assuming that stopping trolley *B* also induces trolley *A* to stop, but not vice versa. Various philosophers, e.g., Parfit (“Future Generations”) and Horton (“The All or Nothing Problem”), have discussed similar cases. These cases are weaker than the trolley problem I use here because choosing to do good effectively comes with *no opportunity cost*: by rescuing many on track *B* you also rescue the one on track *A*, whereas you sacrifice the one by rescuing many in the initial problem.

McMahan (“Doing Good and Doing the Best”) argues that a conditional obligation to donate effectively only arises in cases, such as Horton’s and Parfit’s, in which doing good effectively has no opportunity cost. As we shall see, donating a kidney effectively can sometimes have opportunity costs, that is, someone who would have benefitted from a given less-than-effective donation might not benefit from an effective donation. Thus, *prima facie*, it seems that if one maintains, in line with McMahan, that a conditional obligation to donate effectively only arises in no-opportunity-cost cases, one should reject that such an obligation arises in cases of kidney donations.

However, there is a reason to believe that such an obligation does arise in the context of kidney donations, even if one maintains that a conditional obligation to donate effectively only arises in no-opportunity-cost cases. The reason is that an altruistic kidney donor does not usually know *ex ante* whether or not she is in a no-opportunity-cost case because she does not know the identity of the beneficiaries of her possible (effective or less-than-effective) donations. (See section 2 for the details of altruistic kidney donations.) Since she could well be in a no-opportunity-cost case, she might be morally required to donate her kidney effectively. Thus, if she does not donate effectively, she risks violating this moral requirement. Therefore, she ought to donate effectively.

optimal from an impartial point of view.<sup>5</sup> There are also other possible circumstances that may block the consequent of the effectiveness principle, for example, if the person on track *A* would die with certainty whereas the persons on track *B* would have a nonzero chance of survival, or if helping has bad side effects. The other-things-being-equal clause is supposed to capture all such relevant considerations.

I shall next extend the effectiveness principle minimally along three independent lines. Each will be motivated by slightly altering the trolley problem.

1. We assumed above that your sacrificing an arm will stop a trolley with certainty. But suppose instead that your arm only slows a trolley down, stopping it later, so there is a small, positive probability that it will still reach any given person on its track. You have no reason to believe that this probability differs systematically for different persons, or different tracks: from your perspective, indistinguishable strangers stand at roughly the same distance on both tracks from indistinguishable trolleys that approach with roughly the same speed. Arguably, under this variation of the problem the effectiveness principle continues to hold. Since you cannot rationally differentiate between the probabilities of different persons being overrun or saved, the only morally relevant feature that distinguishes tracks *A* and *B* is the number of persons on them, just as before. So we shall assume that *the effectiveness principle holds in cases in which the benefits of the donation accrue to beneficiaries with probabilities smaller than 1, where the donor has no reason to believe that the probabilities of the benefits accruing to the persons under the first allocation differ systematically from the probabilities of the benefits accruing to the persons under the second allocation.*

It is worth noting that this extension of the effectiveness principle is weaker than a principle prescribing the maximization of expected value. The latter would require sacrificing your arm on track *B* even when the probability of surviving for the persons on track *B* were much lower than for the person on track *A*. Our extension is silent about cases in which these probabilities differ.

2. Suppose that your motivation for stopping a trolley is not to rescue lives, but a different goal, such as meeting social expectations, or the desire to be seen as a hero. It seems that if such a non-altruistic motive is the driver behind the decision to help it would nevertheless be morally wrong to rescue only one person instead of many. (Note that, in

5 Cf. Horton, "The All or Nothing Problem," 98.

this case, you do not have agent-relative reasons that would break the symmetry of the options.) So the thought experiment advocates the effectiveness principle *even when the donor's motivation for donating is not altruistic*.

3. Finally, suppose a stranger is in the choice situation and you, instead of stopping the trolley yourself, observe her free choice. If she decides to stop a trolley, no matter on which track, you can prevent her from doing so. Suppose she decides to stop trolley *B*, thus rescuing one hundred persons. Most of us will likely have the intuition that it would be morally wrong of you to stop her from doing so. (Perhaps the case would be more difficult should she decide to stop trolley *A* instead, but for our purposes this can be ignored.) There are limits to this intuition. For example, suppose the donor would not only lose her arm but would also bleed to death. Many think we should prevent people from sacrificing their lives. We shall assume that, *within reasonable limits, one ought not to prevent a donor from exercising her conditional obligation under the effectiveness principle*. We need not take a stance here on how to set the reasonable limits in general, although we will touch on it below with respect to kidney donations.

For the rest of this paper, the effectiveness principle and its three extensions will be assumed. We will employ them in section 3. Before doing so, we introduce KE in some depth.

## 2. KIDNEY EXCHANGE AND ALTRUISTIC DONATIONS

Worldwide, there is a growing number of patients on waiting lists for kidneys. These are patients who suffer from end-stage renal disease, that is, their existing kidneys have failed. In most countries, there is a sizeable shortage of kidneys for transplantation. For example, in the US, 83,978 people were on the deceased donor waiting list for kidneys in 2015, 5,400 in the UK, and around 8,000 in Germany.<sup>6</sup> The average time a person spends on these waiting lists is two-and-a-half

6 For the US data, see [https://www.usrds.org/2017/view/v2\\_o6.aspx](https://www.usrds.org/2017/view/v2_o6.aspx), accessed April 11, 2018. The number refers to dialysis patients only. For the UK data, see <https://www.organdonation.nhs.uk/news-and-campaigns/news/nhs-blood-and-transplant-reveals-nearly-49-000-people-in-the-uk-have-had-to-wait-for-a-transplant-in-the-last-decade>, accessed April 11, 2018. For the Germany data, see [http://statistics.eurotransplant.org/index.php?search\\_type=&search\\_organ=kidney&search\\_region=all+et&search\\_period=by+year&search\\_characteristic=&search\\_text](http://statistics.eurotransplant.org/index.php?search_type=&search_organ=kidney&search_region=all+et&search_period=by+year&search_characteristic=&search_text), accessed April 11, 2018.

to three years in the UK, almost four years in the US, and around six years in Germany.<sup>7</sup>

In the meantime, many receive dialysis. But dialysis diminishes patients' quality of life and their life expectancy, and many die while on the waiting list. Moreover, dialysis is extremely expensive, thus putting a strain on healthcare services and requiring a medical infrastructure that is unavailable in many countries, especially developing countries.<sup>8</sup> Kidney transplants would extend many patients' life expectancy and life quality, and they are in most cases the cheaper alternative to dialysis. Thus, there is an urgent need to increase the supply of kidneys.

Healthy people have two kidneys and can donate one. Live donor kidney transplants offer the best prospects with respect to recipients' life expectancy and quality.<sup>9</sup> However, because of incompatibility, mostly due to blood types, or specific antibodies of the recipient, many willing donors are not eligible to donate to their loved ones.<sup>10</sup> KE promises relief for these patients. KE programs use matching algorithms to determine matches between donors and recipients that increase the number and quality of transplants. To ensure an informed discussion of the ethics of KE, I shall in the following introduce the main KE procedures in depth.

The simplest KE procedure is "two-way kidney paired donation," as shown in part A of figure 1. This procedure matches two incompatible donor-recipient pairs that are mutually compatible. So the donor of the first pair must be compatible with the recipient of the second pair, and the donor of the second pair compatible with the recipient of the first pair, in order for the exchange to happen. This requirement of *reciprocal compatibility* can be relaxed if, additionally, more than two-way paired donations are feasible.<sup>11</sup> For example, part B in figure 1 shows a three-way paired donation.

Other forms of KE alleviate the requirement of reciprocal compatibility by combining paired donations with *altruistic donations*. An altruistic, or "non-di-

7 For the UK data, see <https://www.nhs.uk/conditions/kidney-transplant/waiting-list>, accessed April 11, 2018. For the US data, see [https://www.usrds.org/2017/view/v2\\_06.aspx](https://www.usrds.org/2017/view/v2_06.aspx), accessed April 11, 2018. For the Germany data, see <https://www.dso.de/organspende-und-transplantation/warteliste-und-vermittlung/niere.html>, accessed April 11, 2018.

8 Wang et al., "Global, Regional, and National Life Expectancy, All-Cause Mortality, and Cause-Specific Mortality for 249 Causes of Death, 1980–2015," 1525.

9 E.g., Wallis et al., "Kidney Paired Donation."

10 Referring to data from the Global Observatory on Donation and Transplantation, Biró et al. report that "depending on the country, 40% or more of recipients are incompatible with their intended donors" ("Building Kidney Exchange Programmes in Europe," 6).

11 Roth, Sönmez, and Ünver, "Efficient Kidney Exchange."

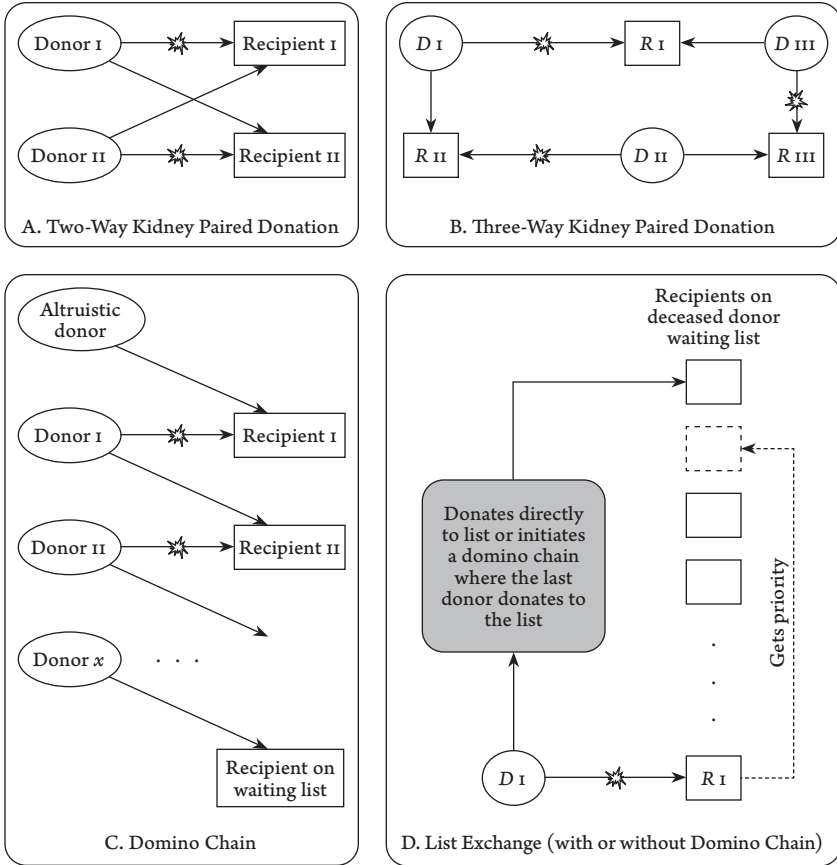


FIGURE 1 Simultaneous KE procedures. A solid arrow from  $a$  to  $b$  denotes an intended kidney donation from  $a$  to  $b$ . An exploding arrow denotes incompatibility of the intended donor.

rected” donor is defined in this context as someone who gives a kidney to a stranger without receiving compensation.<sup>12</sup> In the absence of KE, altruistic donor kidneys are allocated to compatible patients ranked high on the waiting list. When combined with KE, an altruistic donor does not donate directly to the list. Instead, they donate to the recipient of an incompatible pair, whose incompat-

12 Altruistic donors do not typically specify a person they wish to receive their kidney, which is why “non-directed” and “altruistic” refer to the same class of donors. However, in some countries (e.g., the UK), directed altruistic donations are legal, that is, donors may donate to specific but unrelated persons, for example, a patient whose predicament was reported on TV. Directed altruistic donations may give rise to “repugnance” (Roth, “Repugnance as a Constraint on Markets”), for example if they produce markets for attention among patients with end-stage renal disease. However, they appear to be a marginal phenomenon and will not be considered here.



ible donor simultaneously donates to the recipient of yet another incompatible pair, and so on, up to the last donor, who donates to a patient on the waiting list. Since the altruistic donation kicks off various transplants, the resulting chains are called “domino chains.”<sup>13</sup> A domino chain is shown in part C of figure 1.

Similarly, paired donations could be combined with *list exchanges*, as in part D of figure 1. In conventional list exchanges, the donor of an incompatible pair donates to a patient on the waiting list and in return her recipient gets priority on the list. Instead of donating directly to the list, the donor could also donate to the recipient of another incompatible pair, thus kicking off a sequence of simultaneous exchanges, in which the last donor donates to the list and the recipient of the first donor gets priority on the list.<sup>14</sup>

Each of the procedures A–D triggers at least two transplantations, which are carried out *simultaneously*. This is because the promise to give a kidney is not legally enforceable, which poses the problem that in nonsimultaneous chains, possible donors might renege on their promise to donate once their recipient receives a kidney. The size of simultaneous chains is circumscribed, in particular by hospitals’ logistics (each transplantation requires two operating rooms, and hospitals cannot accommodate many transplantations simultaneously), or by geography (kidneys must be transplanted quickly and should therefore not travel far).

However, there have been successful, “nonsimultaneous, extended, altruistic donor” (NEAD) chains.<sup>15</sup> These consist of segments of domino chains, as shown in figure 2. The last donor of a segment becomes a “bridge donor”: instead of simultaneously donating to the waiting list, they initiate a new segment at a later date. There are two types of NEAD chains: Closed NEAD chains specify a last donor, who donates to the waiting list simultaneously with the other donations of the last segment. Open-ended NEAD chains, in contrast, consist of indefinitely many segments. They end only when a bridge donor is ineligible to donate (e.g., because of a difficult-to-match blood type), or if they renege on their promise to donate. Being nonsimultaneous, NEAD chains alleviate the logistical obstacles that confine simultaneous chains. They promise a further increase in chain lengths, some of which have reached over thirty recipient-donor pairs in recent

13 Montgomery et al., “Domino-Paired Kidney Donation”; Roth et al., “Utilizing List Exchange and Nondirected Donation through ‘Chain’ Paired Kidney Donations.”

14 Roth et al., “Utilizing List Exchange and Nondirected Donation through ‘Chain’ Paired Kidney Donations.”

15 Rees et al., “A Nonsimultaneous, Extended, Altruistic-Donor Chain.”

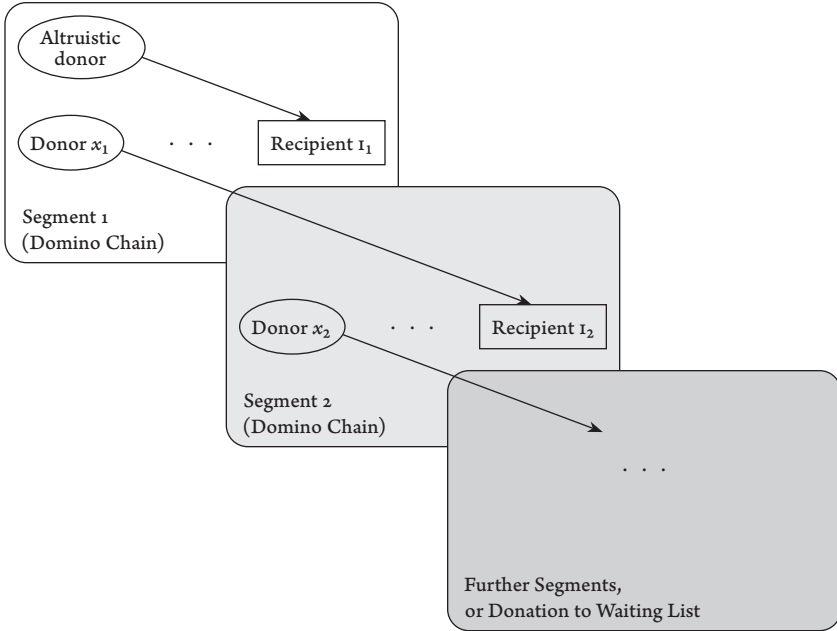


FIGURE 2 Nonsimultaneous, extended, altruistic donor (NEAD) chain. An altruistic donor initiates a domino chain (segment 1). The last donor from segment 1 (denoted  $x_1$ ) becomes a bridge donor and initiates another domino chain (segment 2) at a later date. The last donor of segment 2 either donates to the waiting list, in which case the NEAD chain ends, or becomes a bridge donor and initiates segment 3, and so on.

years.<sup>16</sup> However, they raise ethical and motivational issues for the bridge donors, which we will encounter below.

KE programs are being increasingly implemented in many countries.<sup>17</sup> Further developments that promise additional increases in the numbers and the quality of transplantations include global KES, some of which have already taken place.<sup>18</sup> The integration of compatible pairs into KE is also possible if those pairs would profit from the exchange, or if they wish to engage in altruistic behavior.<sup>19</sup>

However, while KE is expanding, it is at the same time meeting ethical concerns. These are embodied in the transplant laws in various countries that vir-

16 Ashlagi et al., “Nonsimultaneous Chains and Dominos in Kidney-Paired Donation—Revisited.”

17 E.g., Roth, “Transplantation” and “Marketplaces, Markets, and Market Design.” For Europe, cf. Biró et al., “Building Kidney Exchange Programmes in Europe.”

18 Rees et al., “Kidney Exchange to Overcome Financial Barriers to Kidney Transplantation.”

19 Roth et al., “Kidney Paired Donation with Compatible Pairs”; Wallis et al., “Kidney Paired Donation.”

tually ban KE programs. For instance, German legislation permits donations only from persons of first and second degree of relationship, or who otherwise “manifestly stand in a special, personally close relationship” to the recipient.<sup>20</sup> This restriction makes most KES and all altruistic donations, which are by definition provided by strangers, illegal. There have been piecemeal two-way kidney paired donations where the two incompatible pairs established personally close relationships with each other, thus allowing for legal paired donation. However, since the requirements on exhibiting such a relationship are high, and in the absence of an appropriate clearinghouse, success via this path is unlikely. This is even more so for three-way exchanges, the integration of which would achieve a more efficient use of the donor pool. Domino and list exchanges are rendered altogether impossible.

Germany is not an isolated case. In Europe, similarly restrictive laws are in effect in Bulgaria, Estonia, Finland, and Hungary, among others.<sup>21</sup> Less but still relatively restrictive laws prevail in Belgium, France, Greece, Poland, and Switzerland, which are more permissive concerning kidney paired donation, but prohibit altruistic donations, thus ruling out domino chains.<sup>22</sup> Thus, there seem to be concerns in particular about altruistic donations, leading many countries to impose restrictive transplant laws that rule out various forms of KE.

We shall uncover the possible objections to different forms of live kidney donations and their implications for KE in section 4. Before proceeding to that, we will examine KE in view of the effectiveness principle. For the time being, we shall assume that all types of live kidney donations, including altruistic donations to strangers, are feasible.

### 3. KIDNEY EXCHANGE AND THE EFFECTIVENESS PRINCIPLE

It will not come as a surprise what the effectiveness principle and its extensions imply for the allocation of live kidney donations. Yet we need to take care not to jumble different types of kidney donors. This section will consider the most important types of donors successively, and it will end by examining the implications for KE programs.

First, consider *altruistic donors*. When an altruistic donor donates to the wait-

20 The author’s translation from paragraph 8 of the German Transplant Law, version from September 4, 2007 (BGBl. I S. 2206), changed by article 2 from November 21, 2016 (BGBl. I S. 2623).

21 Wissenschaftlicher Dienst, “Die Cross-over-Lebendspende,” 17; and Lopp, *Regulations Regarding Living Organ Donation in Europe*.

22 Biró et al., “Building Kidney Exchange Programmes in Europe,” esp. 12 and table 1.

ing list, they may help one patient on that list. If they donate into KE instead, they trigger a KE chain, thereby helping at least two, but possibly many more patients if the transplantations are successful. From this and the effectiveness principle it follows that, when an altruistic donor is offered the choice between donating into a waiting list or into KE, morality requires the latter. Moreover, KE programs use optimization algorithms that maximize the number of possible matches within the pool of possible donors and recipients, subject to quality constraints (see below). The use of these algorithms guarantees that no possible allocation of kidneys in this pool could be more effective, given those quality constraints. Therefore, an altruistic donor donating into KE is thereby donating as effectively as possible.

So far, we have talked as if comparing successful transplants in the presence versus the absence of KE. But there is no guarantee for success: in a small number of cases, graft loss or other complications occur for the recipient. When an altruistic donor is offered the choice between donating into a waiting list or into KE, they do not know who would receive their kidney in each case and the respective chances of success. Before KE programs were in effect, altruistic donations were typically allocated to patients ranked high on the waiting list in such a way that takes the match quality into account—including factors such as blood type compatibility, sensitization, and age. This increases the chances of success. Similarly, the optimization algorithms used in KE may include all the factors that figure into the list allocations, thus achieving similarly high levels of success.<sup>23</sup> So altruistic donors have no reason to believe that the chances of success differ systematically for donating into KE versus donating to the waiting list. In this situation, the first extension of the effectiveness principle applies. According to this extension, the effectiveness principle is in effect when the benefits of the donation accrue with probability  $< 1$  and there is no reason to believe that this probability differs systematically for the beneficiaries under the two possible allocations. Thus, this principle requires donating into KE in the probabilistic case as well.

In contrast to altruistic donors, *directed donors* have agent-relative reasons that may block the obligation to donate into KE, as the other-things-equal clause of the effectiveness principle applies when a donor wishes to help a relative or friend. If they are compatible, it is usually uncontroversial that that person will receive the organ. If they are not compatible, it might nevertheless be possible for the recipient to receive a kidney through paired donation. So they may engage in KE and thereby achieve an effective allocation, but effectiveness results as a by-product: it is not the effectiveness that requires the donor to engage in the

23 Rees et al., "A Nonsimultaneous, Extended, Altruistic-Donor Chain," 1100.

paired donation, but the agent-relative reason that a donation to a stranger will provide a transplant for their loved one.

Next, consider *donors eligible for bridge donation in NEAD chains*. Remember that a bridge donor donates to a stranger after their recipient receives the transplant and the donation triggers a domino chain that constitutes a segment of a NEAD chain. A donor who is eligible for bridge donation may be offered the choice between donating to a waiting list simultaneously to their recipient receiving the transplant, or triggering a domino chain at a later date. Their motivation for donating is to help the intended recipient and thus differs from that of an altruistic donor. If they decide to become a bridge donor, they donate to a stranger because they honor their promise to do so after the recipient received the transplant. They typically would not have donated to a stranger otherwise, and would have donated to the recipient if compatible. Bridge donors' motivation gives rise to concerns that they might renege on their promise to donate once their recipients receive the transplant, as discussed in section 4. Here, we note that the second extension of the effectiveness principle applies to persons who are eligible for bridge donation. It states that the effectiveness principle is indifferent to the motivation for donating. Thus, for a donor who is offered the choice to donate into a list (simultaneously with their recipient receiving a transplant) or to trigger a chain (at a later date), morality requires the latter.<sup>24</sup>

It is less clear whether the same holds for *donors of compatible pairs* who are offered the choice to take part in KE. They may accept this offer for a range of reasons, for example that their recipient would profit from a kidney paired donation because of a better match. Or they refuse the offer, for example when the compatible recipient prefers the organ of a related party to a stranger's. In such a case, there might be legitimate agent-relative reasons for not donating into KE and the effectiveness principle does not apply. In other cases, in which there are no relevant agent-relative reasons to prefer donating directly to their loved ones, the effectiveness principle applies, requiring them to take part in KE.

Finally, consider the third extension of the effectiveness principle: it is not permitted to prevent a donor from exercising their conditional obligation to do-

24 If a donor has the choice to donate at the point in time in which her recipient also receives a kidney, or to trigger a chain at a later date, the time lag might constitute a morally relevant factor for her decision. This is because, for reasons of time discounting, a kidney transplant for a single person in need now may have more value than kidney transplants for many persons in the distant future. However, discounting is unlikely to be significant in this context because the time frames in question are relatively short. For instance, Fumo et al. argue that bridge donors' waiting times should be subject to a maximum limit of seven to twelve months ("How Long Is Too Long?"). Arguably, only under unreasonably high discount rates would this justify sacrificing the value of a future chain for a single transplant now.

nate effectively. Concerning the relevant donor groups above—altruistic donors, donors eligible for bridge donation, and some compatible pairs—KE is instrumental in meeting the conditional obligation to donate effectively. Thus, it is morally wrong to prevent these donors from donating into KE. It is then only a small step to conclude that it is wrong not to allow KE. This is because it is plausible to assume that if it is morally wrong to prevent someone from acting in an institution and there are no good independent reasons for prohibiting that institution, then it is morally wrong not to allow the institution. We shall see in section 4 that the alleged reasons against KE programs are in fact not good reasons.<sup>25</sup> It follows that it is morally wrong not to allow KE programs.

To sum up, weak principles from the ethics of giving have two important implications for live kidney donation: (i) there is a conditional obligation for altruistic donors and donors eligible for bridge donation, and for some compatible pairs, to donate into KE instead of into a waiting list, if they can choose to do so; and (ii) since KE is instrumental in meeting the conditional obligation to donate effectively, it is wrong to prohibit KE if there are no independent reasons for doing so (a condition that will be discussed below).

The effectiveness principle may have other implications that will not be considered in depth here. For example, it might be applied to make a case for global KE, which has the potential to substantially increase pool sizes and thus numbers and quality of transplants. However, global KE may give rise to separate ethical issues, for example, possible organ trafficking and unreliable medical care in developing countries, and it has generated opposition on these grounds.<sup>26</sup> These are difficult issues, worthy of a separate paper, and will therefore not be considered here.

#### 4. THE SCOPE OF KE AND THE DESIGN OF TRANSPLANT LAWS

Where do the ethical concerns stem from, which are embodied in many transplant laws that virtually ban KE programs? There are various potential issues that we have ignored in the argument from the effectiveness principle. We shall first discuss influential arguments against donations from strangers, which rule out

25 Note that allowing KE programs does not necessarily mean allowing every possible instance of KE; as we shall see, there might be good reasons for restricting the scope of KE programs. But even if there is reason to think we should not allow some instances of KE, say, those resulting from NEAD chains, this could easily be institutionalized in a KE program that does not offer NEAD chains (in fact, most existing KE programs are of this kind), so it is not a reason to prohibit KE programs. Thanks to an anonymous reviewer of this journal for pressing me on this point.

26 Delmonico and Ascher, "Opposition to Irresponsible Global Kidney Exchange."

most KES. The Research Section of the German Federal Parliament provided a rich source of these arguments in a technical report.<sup>27</sup> We then discuss narrower arguments for restricting the scope of KE programs, which stem from concerns about specific types of KE, especially NEAD chains. Finally, the implications of this discussion for the design of transplant laws will be considered.

#### 4.1. Arguments against Donations from Strangers

*Protecting donors from possible harms.* There is no evidence that live kidney donations significantly decrease donors' life expectancy or quality of life.<sup>28</sup> However, like any invasive surgery, they entail small risks of medical complications, including a very small, nonzero probability of death. These are possible harms to healthy persons who receive no medical benefits from the surgery. Therefore, removing a kidney from such a person might be argued to violate the physician's duty to "do no harm." Accordingly, one of the reasons why the German transplant law prohibits donations from strangers is to protect live donors from such harms that their decision to donate might entail.

But risk does not imply harm. Prohibiting live donations on the basis of donor protection would require a duty to incur no risks of harm, or, more reasonably, no risks above certain thresholds, which must arguably be set relative to the benefits to the recipient. This is not the place to argue for a specific threshold that is acceptable for live kidney donations, but it may nevertheless be helpful to compare their risks to some other risks that many people face in their daily lives. It is estimated that 3.1 per 10,000 kidney donors die during or within the first ninety days of their donation.<sup>29</sup> This mortality rate is comparable to working in refuse and recyclable material collection for a year, according to statistics on occupational hazards.<sup>30</sup> It is five times smaller than a year working in logging, which is listed as the most dangerous profession in these statistics. It has been argued, not least by medical practitioners, that these risks are reasonably low.<sup>31</sup> Moreover, they are arguably far outweighed by the benefits to the recipient. Fur-

27 Wissenschaftlicher Dienst, "Die Cross-over-Lebendspende."

28 For a detailed discussion, cf. Tenenbaum, "Bartering for a Compatible Kidney Using Your Incompatible, Live Kidney Donor," 136–39.

29 Segev et al., "Perioperative Mortality and Long-Term Survival Following Live Kidney Donation."

30 See Bureau of Labor Statistics, "National Census of Fatal Occupational Injuries in 2016."

31 For example, Richard B. Freeman writes, "We expose patients to all kinds of risks every day for presumed benefits. Moreover, people willingly assume risks in their everyday lives, often much greater than those imposed by donor surgery, that have little or no direct benefit to their health. The risk that the harms from kidney donation will occur is very small compared with many risks we all face in everyday life" ("The Limits of Altruism," 273).

thermore, presupposing that donors are mentally healthy and not subject to coercion or exploitation (concerns that will be discussed below), it seems they have a right that their autonomous choice be respected.<sup>32</sup>

More central for our purposes is the fact that Germany and other countries allow directed donations while prohibiting donations from strangers. Their implicit assumption seems to be that the risk of harm to directed donors is justifiable but the risk of harm to non-directed donors is not. But these risks do not differ systematically. So the claim that non-directed donors are more in need of protection from their decision to donate must be based on other considerations than risk: it must be based on considerations justifying that the same level of risk is acceptable for directed but not for non-directed donors.

A possible consideration that would justify this is that non-directed donors are mentally ill, while directed donors are not. This is a suspicion that altruistic donors' apparently selfless decision to give an organ to strangers sometimes arouses. But it seems unreasonable to take donors' altruism as evidence for mental illness. Of course, it should be ruled out that a person is donating an organ as a result of impaired judgment due to mental illness. In practice, there is extensive screening of all kinds of living donors—not altruistic donors alone—in order to rule out mental health issues.<sup>33</sup> Skeptics might challenge the reliability of such tests. But it would be unreasonable to suppose that a directed donor who passed the test is mentally healthy while casting doubt on the mental health of an altruistic donor who passed the test, which is what this line of argument for justifying the restriction of the donor pool would require. We shall next examine another possible justification of the restriction.

*Coercion.* Germany also justifies the restriction of the legal donor pool with the need to rule out the possibility of coerciveness of donations and to secure their voluntariness. If it could be argued that anonymous donations entail an element of coercion that directed donations to family members and especially personally close persons do not, this would indeed constitute an argument for the restriction. To examine whether this is the case, we shall consider what a coerced donation could amount to.

In most countries, human kidneys are not for “valuable consideration.” This is a legal term, meaning that it is prohibited both to donate and to receive kidneys in exchange for money or other valuable goods or services. A promise is only le-

32 Cronin, “Allowing Autonomous Agents Freedom.”

33 For the practices in various European countries, cf. Working Group on Living Donation, “Toolbox Living Kidney Donation.” For screenings of non-directed donors in the us, see Tenenbaum, “Bartering for a Compatible Kidney Using Your Incompatible, Live Kidney Donor,” esp. 148–49 and the references in n. 177.



gally enforceable if it is for valuable consideration. Thus, a promise to donate a kidney is not legally enforceable. This rules out the strongest form of coercion, which would subject the provision of a kidney to a legally binding contract. It also rules out possible exploitation of the poor, as it is impossible to sell kidneys.

However, as legal scholars point out,

consideration is a slippery doctrine. . . . [D]onors are allowed to direct that their kidneys be given to certain people: family members, friends, and others. This might seem like a transfer without valuable consideration, but that is not necessarily the case. The donor might transfer to such people rather than to a stranger because she expects to receive something in return—for example, household services or help in some other matter. Only a donation to an anonymous stranger could clearly be without consideration. Nonetheless, the common law of contract generally treats intrafamily transfers as occurring without consideration, and regulated entities and regulators have apparently taken this position for kidney donations to friends and family, as well.<sup>34</sup>

Thus, compensation and coercion are harder to rule out when someone donates to a relative or friend. The organization of KE programs in countries where they are legal reflects this concern that personal relations can be instrumental for exercising coercion. Not only do transplant centers seek to rule out coercion through extensive background checks of potential donors, interviews, and education; it is typically also made impossible for mutually unacquainted persons in KES to contact each other prior to the donation. There are various practical measures to enforce this, such as using different hospital sites. Some countries, e.g., Australia, discourage donors and recipients to meet even after the donation in order to rule out the possibility of posterior compensations, or of raising accusations, for example after graft loss.

Alas, there is no guarantee that donations are always entirely free of some soft forms of coercion, in particular in emotionally close relationships. The argument that restricting the donor pool to especially close persons would help secure the voluntariness of a donation gets it the wrong way around. The altogether different conclusion here is that if you want to allow directed, e.g., intrafamily donations—as most countries, including Germany, do—then there is no reason based on coerciveness for prohibiting anonymous donations, including altruistic donations.

There is a more subtle issue concerning coercion in NEAD chains, which will be considered below.

34 Choi et al., “Altruism Exchanges and the Kidney Shortage,” 290–91.

*Slippery slope: KE and the commercialization of human organs.* Germany also adduces the prevention of organ trade as a reason for the limitation of the donor pool. However, this argument is not convincing. It is empirically unfounded, as most countries condemn the practice of buying and selling organs and there is not a single country that has commercialized organ donations after implementing KE programs.<sup>35</sup> Concerning black markets, there is no reason to believe that they are more likely to develop in the presence of KE. (It might be argued that the opposite is the case because KE helps to decrease the demand for kidneys.) Finally, the argument also commits the fallacy encountered before: Why should it help for preventing valuable consideration in kidney donations to restrict the donor pool to especially close persons, where the risk of valuable consideration is higher?

#### 4.2. Concerns about Specific Types of KE

*Trade-offs between efficiency and fairness.* As we have seen, without KE, altruistic donations are allocated to compatible patients ranked high on a deceased donor waiting list. Now, suppose that an altruistic donor decides to donate into KE instead of to the list. The resulting concern is most visible in open-ended NEAD chains. Remember that these chains end only when a bridge donor becomes ineligible or reneges on their promise to donate. As a result, there is no kidney that will be allocated to the waiting list. Closed NEAD chains, in which a last donor is specified who will donate to the waiting list, are not necessarily subject to this diversion of kidneys from the list. But even in closed NEAD chains, the last donor might not end up donating to the list, for example, if the bridge donor of an earlier segment reneged on their promise to donate, which terminated the chain prematurely.

The allocations of kidneys resulting from NEAD chains raise concerns because deceased donor waiting lists incorporate medical but also fairness principles, such as time already spent waiting, or priority of children over adults. Thus, NEAD chains, by diverting kidneys from the list, might go against a fair allocation of kidneys. Furthermore, when kidneys are not allocated to the list, those particularly vulnerable patients on the list who do not have living donors might be disadvantaged because they are not eligible to participate in KE.

On the other hand, NEAD chains achieve large numbers of transplants. Thus, these chains can be seen as promoting efficiency at the expense of fairness. Proponents of NEAD chains typically counter this concern by arguing that the efficiency that NEADs achieve helps patients on the waiting list too, namely by removing multiple patients from the list.<sup>36</sup> Yet, diverting altruistic donations from

35 Iran is currently the only country where the sale of kidneys is legal.

36 Rees et al., "A Nonsimultaneous, Extended, Altruistic-Donor Chain," 1100.

the list may disadvantage at least some patients on the list. In particular, a patient does not profit if lower-ranked patients are removed from the list, so patients that are already highly ranked can be expected to be disproportionately disadvantaged.

Unlike NEAD chains, other types of KES do not in principle divert live donor kidneys from waiting lists. However, combining KE with waiting lists may disadvantage blood-type O patients on the list. The reason is, very roughly, that blood-type O patients can receive kidneys only from O donors, whereas O donors can donate to all blood types. Now, consider as an example a domino chain that an altruistic donor triggers and that ends with an incompatible donor donating to the list. The distribution of blood types among altruistic donors resembles that of the general population. Therefore, there is a high probability that this donor is O and will donate to a hard-to-match O recipient. But it is unlikely that the incompatible donor who donates to the list is O, otherwise they would likely be compatible with the recipient. Thus, KE may systematically divert highly demanded O kidneys from the list.<sup>37</sup>

Woodle et al. suggest that the crucial ethical question concerning the trade-off between efficiency and fairness is this: How many additional transplants must the inclusion of altruistic donors into KE chains generate in order to justify the diversion of altruistic donors (in NEAD chains), or of blood-type O altruistic donors (in general) from the waiting list?<sup>38</sup> Transplant laws could do justice to a specific answer to this question by stipulating that the inclusion of altruistic donors into KE requires a minimum number of transplants. Moreover, concerning the loss of O donors, they suggest “a requirement that, for every [altruistic donor] kidney donated to initiate a KE chain, a kidney of the same blood type must be donated to the [waiting list] at the end of the KE chain.”<sup>39</sup>

*Risks for bridge donors.* NEAD chains entail the risk that bridge donors renege on their promise to donate. However, the rates of renegeing bridge donors appear to be small, and it has been argued that the utility benefits from NEAD chains outweigh these risks.<sup>40</sup>

There is yet another worry concerning bridge donors. NEAD chains are

37 A similar concern arises in list exchanges (conventional or combined with domino chains). For a discussion, see den Hartogh, “Trading with the Waiting-List.”

38 Woodle et al., “A Prospective, Randomized, Multicenter Study Evaluating Early Corticosteroid Withdrawal with Thymoglobulin in Living-Donor Kidney Transplantation.”

39 Woodle et al., “A Prospective, Randomized, Multicenter Study Evaluating Early Corticosteroid Withdrawal with Thymoglobulin in Living-Donor Kidney Transplantation,” 1464.

40 Wallis et al., “Kidney Paired Donation”; Tenenbaum, “Bartering for a Compatible Kidney Using Your Incompatible, Live Kidney Donor.”

formed on the understanding that the bridge donors will donate to initiate a segment of transplants after their partners receive transplants (assuming that they continue to be medically and psychologically eligible). Because they gave this promise beforehand, they may feel obliged to donate after their recipients receive transplants. Bridge donors know that if they bail out they thereby break the promise they gave, on the basis of which their partners received their transplant and on which various persons in need of kidneys rely. This might impose pressure on them, which may be felt as a form of coercion. It has been argued that it is morally problematic to put people in this position, and NEAD chains have been criticized on these grounds.<sup>41</sup>

There are various measures to diminish the risk that bridge donors feel coerced through their promise to donate. Transplant centers select and educate possible bridge donors carefully. It may also be possible to relieve them from some of the felt pressure to be triggering a great number of transplants simply by not telling them how long the chain will be prior to their donation. Moreover, since evidence suggests that the level of felt coercion increases with time, time limits can be set within which their donation should happen, otherwise their promise is void. Alternatively, a more conservative solution would be to restrict KE programs to the simultaneous cases.

#### *4.3. Implications for the Design of Transplant Laws*

The principled arguments against donations from strangers, which preclude most KES, have been found wanting. However, it might be ethically required to restrict the scope of KE procedures. In general, the efficiency gains from allowing broader KES must be weighed against increasing concerns with respect to the diversion of altruistic donors, especially type-O donors, from waiting lists and, in the case of NEAD chains, the potential felt coercion of bridge donors. My aim was not to argue for a specific weighting. Instead, the argument is the following. Suppose we take a conservative view and put heavy weight on avoiding the diversion of altruistic donors, especially O donors and on minimizing the possibility that bridge donors feel coerced into donating. A transplant law embodying this view might restrict or even prohibit NEAD chains. It may also require that KE chains divert altruistic donations from the list only when they achieve a large number of transplants, and it may prescribe the prevention of O donor loss. The result would be a transplant law that places heavy weight on donor protection and allocative fairness with respect to patients without living donors. The point is that this legislation would not resemble the transplant laws we encountered

41 For discussion, see Tenenbaum, "Bartering for a Compatible Kidney Using Your Incompatible, Live Kidney Donor."

earlier, such as the German transplant law, which require a personally close relationship between donor and recipient. Even a conservative view on live kidney donation, if sound, does not in principle reject KE programs.

##### 5. THE ATTRACTION OF EFFECTIVENESS

KE may generate motivational benefits for donors. Consider the following report from Dylan Matthews, who altruistically donated his kidney in 2016:

The very same day that I donated, [the recipient's] relative had their kidney taken out as well and flown to the West Coast. This second recipient also had a friend or relative agreeing to an exchange; so did the third recipient, who got the second recipient's friend's kidney. Our chain will let people enjoy 36 to 40 years of life they would've otherwise been denied.

Our four kidneys were pretty good, but some chains can go even longer. A chain started by a 44-year-old man in California named Rick Ruzzamenti wound up getting 30 people kidneys. Ruzzamenti's chain let people live 270 to 300 years longer. You can literally measure the years of life his kidney donation chain gave in centuries.<sup>42</sup>

Matthews does not go so far as to suggest that he, or Ruzzamenti, decided to donate *because* of the potentially large numbers of life years that their donations would enable. But the passage provides clear evidence for the awe that donors experience when considering the large impact of their donations in terms of life years gained. This naturally suggests the hypothesis that, other things being equal, a donor's motivation is higher if the possible number of transplants triggered, or of life years saved, is higher.

This hypothesis, if true, has implications for KE. As we have seen, KE programs use optimization algorithms that increase the number and quality of transplants. Thereby, they increase the number of lives saved, or of life years gained.<sup>43</sup> It follows that, if the motivation for donating is partly determined by and increases with the impact of the donation, KE increases the motivation for donating. The argument would apply to all donors who are partly motivated by altruism. This includes altruistic donors, but also other types—such as directed donors, bridge donors, and compatible pairs—as they may often be partly motivated by altruism as well.

42 Matthews, "Why I Gave My Kidney to a Stranger."

43 Different algorithms may have unequal implications concerning numbers of lives saved and of life years gained. We can neglect this point here because an altruistic donation will typically increase both variables if KE is in effect as compared to the default of no KE.

In light of the striking shortage of kidneys for transplantation, if KE programs promote altruistic behavior, this constitutes a significant advantage. Thus, the hypothesis that the motivation for kidney donations is partly determined by the amount of good they can be expected to achieve, if true, constitutes a second, motivational argument for the implementation of KE programs.

This is not the place to investigate whether the hypothesis is true. It is an empirical hypothesis that could be confirmed by comparing trajectories of altruistic donations in countries where KE programs exist to countries where they do not. We note here merely that the available evidence is consistent with the hypothesis. In many countries in which centralized KE programs exist, e.g., in the US and the UK, the numbers of altruistic donations have been increasing in recent years.<sup>44</sup> More generally, there is evidence that donating effectively can boost donors' motivation.<sup>45</sup> We conclude that it is a reasonable hope that KE promotes the emergence of altruism.

In contrast, transplant laws that restrict the donor pool to relatives and persons manifestly close to recipients entail problematic incentives. Concerning altruistic donations, they convey the image that there is something unethical about the gift to a stranger. Moreover, concerning kidney paired donations, members of mutually compatible pairs will have incentives to pretend that there are personally close relationships even when there are not. But such incentives to "game the system" cannot be in the interest of legislative authorities, and they are detrimental for building trust in the system. These motivational considerations speak against prohibiting live organ donations to strangers, and in favor of making effective use of them through KE.

## 6. CONCLUSION

Weak principles from the ethics of giving make a strong case for KE programs. These programs are instrumental in allowing kidney donors to meet the conditional moral obligations implied by those principles. Therefore we ought not to preclude people from fulfilling these obligations by banning KE. There might be ethical reasons for restricting specific procedures of KE, but these reasons do not

44 For data on altruistic donations in the UK, see Robb et al., "Annual Report on Living Donor Kidney Transplantation." For the US, see, e.g., Tenenbaum, "Bartering for a Compatible Kidney Using Your Incompatible, Live Kidney Donor."

45 Parbhoo et al., "Best of Intentions," 21. In a survey, 85 percent of donors revealed that they paid very close attention to effectiveness when giving to charities, and even though they care about effectiveness, few donors spend time investigating the effectiveness of the charities they give to. So admittedly, the evidence is somewhat mixed.

in principle reject KE programs. Finally, KE may achieve motivational benefits that constitute a further argument in its favor.

The arguments given here are not wedded to a specific moral theory. They will appeal to effective altruists, but because of their weak, conditional premises, many people who are not committed effective altruists will welcome them as well. They are also consistent with conservative views on donor protection and allocative justice concerning patients on waiting lists. I hope that these arguments will lead to a clarification of the debates about the ethics underlying KE programs, particularly in countries that have hitherto banned these programs.

This paper calls for various follow-up projects. First, we explicitly excluded global KE, which has the potential to substantially increase the numbers and quality of transplants. Ethicists are called for to weigh these benefits against the concerns that have been raised about global KE, for example, whether the risk of organ trafficking can be ruled out sufficiently in some developing countries. Second, the hypothesis about donors' motivation on which our argument from attractiveness draws should be investigated empirically. Third, in many countries that currently prohibit live donations to strangers, changes to legislation, for which we argued here, may not be feasible in the short or medium term. In the meantime, some "slim" forms of KE programs might be implemented in those countries. For instance, restrictions of the donor pool to persons that are emotionally close to the recipient provide the possibility to match donor-recipient pairs that could in the next step meet in person and establish the required relationship. This could enable some types of KE, in particular kidney paired donations, which are currently conducted only sparsely in those countries. Making the most of existing transplant laws would improve the predicament of many people suffering from kidney disease, but it does not excuse decision makers' inaction.<sup>46</sup>

*Leibniz University Hannover*

*London School of Economics and Political Science  
philippe.van.basshuysen@philos.uni-hannover.de*

46 I thank Luc Bovens, Paul Daniell, Jurgis Karpus, David Kinney, Theron Pummer, Bryan Roberts, Thomas Rowe, Bastian Steuwer, Alex Voorhoeve, and Jannik Zeiser, as well as anonymous referees of this journal, for helpful comments and/or discussions at various stages of this research. I also thank the audience at the Cognition, Value and Behaviour Lab Meeting at LMU Munich for valuable discussions. Part of this research was supported by Volkswagenstiftung within the project "Bias and Discrimination in Big Data and Algorithmic Processing: Philosophical Assessments, Legal Dimensions, and Technical Solutions."

## REFERENCES

- Ashlagi, Itai, Duncan S. Gilchrist, Alvin E. Roth, and Michael A. Rees. "Nonsimultaneous Chains and Dominos in Kidney-Paired Donation—Revisited." *American Journal of Transplantation* 11, no. 5 (May 2011): 984–94.
- Biró, Péter, Bernadette Haase-Kromwijk, Tommy Andersson, et al. "Building Kidney Exchange Programmes in Europe: An Overview of Exchange Practice and Activities." *Transplantation* 103, no. 7 (July 2019): 1514–22.
- Bureau of Labor Statistics. "National Census of Fatal Occupational Injuries in 2016." News release, December 19, 2017. [https://www.bls.gov/news.release/archives/cfoi\\_12192017.pdf](https://www.bls.gov/news.release/archives/cfoi_12192017.pdf).
- Choi, Stephen J., Mitu Gulati, and Eric A. Posner. "Altruism Exchanges and the Kidney Shortage." *Law and Contemporary Problems* 77, no. 3 (2014): 289–96.
- Cronin, Antonia J. "Allowing Autonomous Agents Freedom." *Journal of Medical Ethics* 34 (2008): 129–32.
- Delmonico, Francis L., and Nancy L. Ascher. "Opposition to Irresponsible Global Kidney Exchange." *American Journal of Transplantation* 17, no. 10 (October 2017): 2745–46.
- Den Hartogh, Govert. "Trading with the Waiting-List: The Justice of Living Donor List Exchange." *Bioethics* 24, no. 4 (May 2010): 190–98.
- Freeman, Richard B. "The Limits of Altruism: Selecting Living Donors." *American Medical Association Journal of Ethics* 14 (2012): 272–77.
- Fumo, David, Timothy Suttle, Ryan Flynn, Laurie Reece, Stanislaw Stepkowski, Michael Rees, and Jonathan Kopke. "How Long Is Too Long? The Impact of Bridge Donor Waiting Time on Paired Donation Transplants." *Journal of Urology* 193, no. 4S (April 2015): e1015.
- Horton, Joe. "The All or Nothing Problem." *Journal of Philosophy* 114, no. 2 (February 2017): 94–104.
- Lopp, Leonie. *Regulations Regarding Living Organ Donation in Europe: Possibilities of Harmonisation*. Berlin: Springer, 2013.
- MacAskill, William. *Doing Good Better: How Effective Altruism Can Help You Make a Difference*. London: Random House, 2016.
- Matthews, Dylan. "Why I Gave My Kidney to a Stranger—and Why You Should Consider Doing It Too." *Vox*, April 11, 2017. <https://www.vox.com/science-and-health/2017/4/11/12716978/kidney-donation-dylan-matthews>.
- McMahan, Jeff. "Doing Good and Doing the Best." In *The Ethics of Giving: Philosophers' Perspectives on Philanthropy*, edited by Paul Woodruff, 78–102. New York: Oxford University Press, 2018.
- Montgomery, Robert A., Sommer E. Gentry, William H. Marks, Daniel S. War-



- ren, Janet Hiller, Julie Houp, Andrea A. Zachary, J. Keith Melancon, Warren R. Maley, Hamid Rabb, Christopher Simpkins, and Dorry L. Segev. "Domino-Paired Kidney Donation: A Strategy to Make Best Use of Live Non-Directed Donation." *Lancet* 368, no. 9533 (July 2006): 419–21.
- Parbhoo, Omar, Katy Davis, Robert Reynolds, Piyush Tantia, Pranav Trewn, and Sarah Welch. "Best of Intentions: Using Behavioral Design to Unlock Charitable Giving." Ideas42 Report, 2018. [http://www.ideas42.org/wp-content/uploads/2018/08/ideas42-Best-of-Intentions\\_Charitable-Giving-1.pdf](http://www.ideas42.org/wp-content/uploads/2018/08/ideas42-Best-of-Intentions_Charitable-Giving-1.pdf).
- Parfit, Derek. "Future Generations: Further Problems." *Philosophy and Public Affairs* 11, no. 2 (Spring 1982): 113–72.
- Pummer, Theron. "Whether and Where to Give." *Philosophy and Public Affairs* 44, no. 1 (Winter 2016): 77–95.
- Rees, Michael A., Ty B. Dunn, Christian S. Kuhr, et al. "Kidney Exchange to Overcome Financial Barriers to Kidney Transplantation." *American Journal of Transplantation* 17, no. 3 (March 2017): 782–90.
- Rees, Michael A., Jonathan E. Kopke, Ronald P. Pelletier, et al. "A Nonsimultaneous, Extended, Altruistic-Donor Chain." *New England Journal of Medicine* 360, no. 11 (March 12, 2009): 1096–101.
- Robb, Mathew, Chloe Brown, and Lisa Mumford. "Annual Report on Living Donor Kidney Transplantation." NHS Blood and Transplant, October 2017. [https://nhsbt.dbe.blob.core.windows.net/umbraco-assets-corp/5706/annual-report-on-living-donor-kidney-transplantation-2016\\_17.pdf](https://nhsbt.dbe.blob.core.windows.net/umbraco-assets-corp/5706/annual-report-on-living-donor-kidney-transplantation-2016_17.pdf).
- Roth, Alvin E. "Marketplaces, Markets, and Market Design." *American Economic Review* 108, no. 7 (July 2018): 1609–58.
- . "Repugnance as a Constraint on Markets." *Journal of Economic Perspectives* 21, no. 3 (Summer 2007): 37–58.
- . "Transplantation: One Economist's Perspective." *Transplantation* 99, no. 2 (February 2015): 261–64.
- Roth, Alvin E., Tayfun Sönmez, and M. Utku Ünver. "Efficient Kidney Exchange: Coincidence of Wants in Markets with Compatibility-Based Preferences." *American Economic Review* 97, no. 3 (June 2007): 828–51.
- Roth, Alvin E., Tayfun Sönmez, M. Utku Ünver, Frank L. Delmonico, and Susan L. Saidman. "Utilizing List Exchange and Nondirected Donation through 'Chain' Paired Kidney Donations." *American Journal of Transplantation* 6, no. 11 (November 2006): 2694–705.
- . "Kidney Paired Donation with Compatible Pairs." *American Journal of Transplantation* 8, no. 2 (February 2008): 463.
- Segev, Dorry L., Abimereki D. Muzaale, Brian S. Caffo, et al. "Perioperative Mor-

- tality and Long-Term Survival Following Live Kidney Donation.” *Journal of the American Medical Association* 303, no. 10 (March 10, 2010): 959–66.
- Singer, Peter. *The Most Good You Can Do: How Effective Altruism Is Changing Ideas about Living Ethically*. New Haven: Yale University Press, 2015.
- Tenenbaum, Evelyn M. “Bartering for a Compatible Kidney Using Your Incompatible, Live Kidney Donor: Legal and Ethical Issues Related to Kidney Chains.” *American Journal of Law and Medicine* 42, no. 1 (March 2016): 129–69.
- Wallis, C. Bradley, Kannan P. Samy, Alvin E. Roth, and Michael A. Rees. “Kidney Paired Donation.” *Nephrology Dialysis Transplantation* 26, no. 7 (July 2011): 2091–99.
- Wang, Haidong, Mohsen Naghavi, Christine Allen, et al. “Global, Regional, and National Life Expectancy, All-Cause Mortality, and Cause-Specific Mortality for 249 Causes of Death, 1980–2015: A Systematic Analysis for the Global Burden of Disease Study 2015.” *The Lancet* 388, no. 10053 (October 2016): 1459–1544.
- Wissenschaftlicher Dienst. “Die Cross-over-Lebendspende. Zum Stand in Deutschland und in ausgewählten europäischen Ländern.” *Deutscher Bundestag*, WD9-3000-022/17 (2017).
- Working Group on Living Donation. “Toolbox Living Kidney Donation.” March 2016. [https://ec.europa.eu/health/sites/health/files/blood\\_tissues\\_organ/docs/eutoolbox\\_living\\_kidney\\_donation\\_en.pdf](https://ec.europa.eu/health/sites/health/files/blood_tissues_organ/docs/eutoolbox_living_kidney_donation_en.pdf).
- Woodle, E. S., J. A. Daller, M. Aeder, R. Shapiro, T. Sandholm, V. Casingal, D. Goldfarb, R. M. Lewis, J. Goebel, and M. Siegler. “Ethical Considerations for Participation of Nondirected Living Donors in Kidney Exchange Programs.” *American Journal of Transplantation* 10, no 6. (June 2010): 1460–67.

## ADDRESSED BLAME AND HOSTILITY

*Benjamin De Mesel*

**B**ENJAMIN BAGLEY sets out a dilemma for addressed blame, that is, blame addressed to its targets as an implicit demand for recognition.<sup>1</sup> The dilemma arises when we ask whether offenders would actually appreciate this demand, via a sound deliberative route from their existing motivations. “If they would, their offense reflects a deliberative mistake, and blame’s hostility seems unnecessary. If they would not, addressing them is futile, and blame’s emotional engagement seems unwarranted.”<sup>2</sup>

Bagley wants to resolve the dilemma in such a way that addressed blame’s distinctive elements of hostility and emotional engagement can be accounted for.<sup>3</sup> He draws on two ideas from Bernard Williams. First, what someone would conclude via sound deliberation can be indeterminate. Second, addressed blame works proleptically:

In general, blame is proleptic when the ability of its targets to recognize and respond to our demands may depend on our blame itself, and when the nature of our blame reflects this fact.... When we proleptically address blame to offenders, we presuppose that they have a sound deliberative route to the recognition we demand, but we do not presuppose that this is the only such route open to them. The former gives us reason to actively care about their recognition; the latter gives us reason to be hostile.<sup>4</sup>

I accept that hostility and emotional engagement are distinctive elements of addressed blame, and I believe that Bagley’s focus on addressed blame’s proleptic mechanism helps to resolve his puzzle. I will argue, however, that Bagley has difficulties accounting for the element of hostility in addressed blame, which means that his solution to the dilemma cannot work (section 2). I will suggest that an alternative account of addressed blame avoids these difficulties (section

1 Bagley, “Properly Proleptic Blame.”

2 Bagley, “Properly Proleptic Blame,” 852.

3 Bagley, “Properly Proleptic Blame,” 853.

4 Bagley, “Properly Proleptic Blame,” 854.

3), and reply to an objection (section 4). I will start by introducing Bagley's paradigm example of addressed blame (section 1).

#### 1. ADDRESSED AND NON-ADDRESSED BLAME

Bagley illustrates the difference between addressed and non-addressed blame by way of a passage from E. M. Forster's novel *Howards End* in which Margaret Schlegel blames her husband Henry Wilcox.

"Not any more of this!" she cried. "You shall see the connection if it kills you, Henry! You have had a mistress—I forgave you. My sister has a lover—you drive her from the house. Do you see the connection? Stupid, hypocritical, cruel—oh, contemptible!—a man who insults his wife when she's alive and cants with her memory when she's dead. . . . Only say to yourself: 'What Helen has done, I've done.'"

"The two cases are different," Henry stammered. . . .

"In what way different? You have betrayed Mrs. Wilcox, Henry, Helen only herself. . . . You have the insolence to talk to me of differences, Henry?"

Oh, the uselessness of it! Henry's retort came.

"I perceive you are attempting blackmail. . . . My rule through life has been never to pay the least attention to threats, and I can only repeat what I said before: I do not give you and your sister leave to sleep at Howards End."

Margaret loosed his hands. . . . For a little she stood looking at the Six Hills, tombs of warriors, breasts of the spring. Then she passed out into what was now the evening.<sup>5</sup>

Margaret's blame at the beginning is very different from her blame at the end. The latter is cold rejection and a paradigm example of non-addressed blame. The former is a paradigm example of addressed blame. Addressed blame is proleptic. When Margaret addresses her blame to Henry, it is not determinate whether his conduct embodies relationship-impairing values. Her blame enables him to specify his values, and the way in which he does so will reveal whether their relationship has a future. Non-addressed blame is not proleptic: Henry has revealed relationship-impairing values, no opportunity to respond is provided, and the case is closed.

5 Forster, *Howards End*, 323–24.

## 2. ADDRESSED BLAME AND HOSTILITY

Bagley intends to show how addressed blame's elements of emotional engagement and hostility can be justified. According to Bagley, what gives us reason to be hostile in addressing blame is that we do not presuppose that the route to recognition of our demands is the only sound deliberative route open to the offender.<sup>6</sup> Consider also the following passage:

Because [the offender's] conduct also doesn't determinately not embody their values, it can't be written off as a deliberative mistake. This is what explains addressed blame's element of hostility. While you do not have reason to revise your attitudes toward the offender in the ways a defect in their values would warrant, you also do not have reason to respond favorably toward them in the ways you would if their underlying values were in good shape.<sup>7</sup>

The idea seems to be that we have reason to be hostile whenever we do not have reason to respond favorably. We would have reason to respond favorably if the offender's values were in good shape, if the route to recognition were the only sound route open to them. But we do not presuppose this in addressing blame: we leave open the possibility that they have another route, that their values are not in good shape, and as long as that possibility is left open we have no reason to respond favorably. But do we really have reason to be hostile whenever we do not have reason to respond favorably? That does not seem enough: we need a reason *not* to respond favorably. And even *that* is not enough: if we accept that some responses are neither hostile nor favorable, we need a positive reason to respond with hostility.

Bagley claims that addressed blame is hostile because "we are ready to op-

6 Bagley, "Properly Proleptic Blame," 854. One might object that only *facts* about an offender, and not merely what we *presuppose* or believe about them, can give us reason to be hostile or can justify hostility. It is important to point out here that Bagley's use of "justification" and "reason" is inspired by Williams, "Internal and External Reasons." For a summary of Williams's perspectivist view, see Alvarez, "Reasons for Action." Objectivists about reasons and justification presuppose that whether an agent has a reason to act and whether her action is justified depend solely on the facts. According to perspectivists, by contrast, the "reasons an agent has [and the justification of his action] depend in an important sense on his epistemic perspective, and so an agent can have a normative reason that is a false belief" (Alvarez, "Reasons for Action," sec. 2). As Bagley notes, those who object to the use of "reason" and "justification" in a perspectivist sense "are free to substitute another phrase" (Bagley, "Properly Proleptic Blame," 859). I will use "reason" and "justification" in the way Bagley uses them.

7 Bagley, "Properly Proleptic Blame," 874.

pose these people should they ultimately refuse to [accept our considerations].”<sup>8</sup> But to be “ready to oppose if . . .” is not to be hostile. I am ready to oppose you should you hurt my daughter, and I leave open the possibility that you will, but that does not make hostility toward you justified. To be hostile, my dictionary explains, is to be “opposed.” To be “ready to oppose if . . .” is not yet to be opposed. Bagley writes that addressed blame’s element of hostility “expresses a willingness, should the invitation [to accept our demand] be rejected, to reject the offender in turn.”<sup>9</sup> Again, to be hostile is not just to be “willing to reject if . . .”; it is to reject.

In general, Bagley seems to think that we have reason to be hostile whenever we take it to be *possible* that offenders will refuse to accept our considerations, that they do not have a sound deliberative route to recognition, or that their values are relationship impairing. He writes that Margaret “treats the relation between Henry’s conduct and his values as an open question.”<sup>10</sup> Because she takes it to be possible that his values are relationship impairing, she has no reason to respond favorably. But if there is an open question, she also takes it to be possible that his values are *not* relationship impairing, so why would she have reason to respond with hostility?

### 3. ADDRESSED BLAME: AN ALTERNATIVE SUGGESTION

I will suggest that, in addressing blame, we respond to what we take to be a *fact*, rather than a mere *possibility*, about the offender’s values, and that this way of understanding addressed blame helps us to understand why it is hostile.

It does not seem right to describe Margaret as treating the relation between Henry’s conduct and his values as an open question. If she would, would she then cry “Not any of more of this!”, “Stupid, hypocritical, cruel—oh, contemptible!—a man who insults his wife when she’s alive and cants with her memory when she’s dead”? Margaret does not express the readiness to oppose Henry—she opposes him. She is not asking an open question about the relation between Henry’s conduct and his values. Rather, her addressed blame involves an implicit claim about that relation, something like “Your conduct embodies relationship-impairing values.” The difference between her addressed and non-addressed blame lies indeed, as Bagley points out, in the fact that the latter is final in a way that the former is not. But rather than seeing the relation between them as the relation between an explicitly final claim (“No further response allowed!”) and an *open question*, I propose seeing it as the relation between an explicitly final

8 Bagley, “Properly Proleptic Blame,” 854.

9 Bagley, “Properly Proleptic Blame,” 876.

10 Bagley, “Properly Proleptic Blame,” 874.

claim and a *claim that allows for a response* (which is not the same as a question or a claim that *asks for* or *invites* a response). Margaret claims that Henry's conduct embodied relationship-impairing values; the burden is on Henry to show that it did not. Margaret specifies his values in a certain way ("Only say to yourself: 'What Helen has done, I've done.'"). He gets the opportunity to protest against her way of doing it. He is accused but has the opportunity to appeal.<sup>11</sup>

This way of understanding Margaret's addressed blame makes it possible to see where her hostility comes from. She believes (that is, takes it to be a fact) that Henry's values are relationship impairing, that he does not have a sound deliberative route to recognition, and this belief justifies her hostility or, at least, comes closer to justifying it than the belief that it is possible that his values are relationship impairing. One could object that the values of psychopaths are relationship impairing, and that they do not have a sound deliberative route to recognition, while many think that hostile blame toward them would be unjustified. Those who think that hostile blame toward psychopaths is unjustified, however, tend to think that blaming them is in general unjustified: to the extent that psychopaths are not responsible agents, they are not blameworthy, which is why hostile blame cannot be justified. By contrast, Henry is treated by Margaret as a responsible agent, as someone who is at least partly responsible for the fact that his values are relationship impairing. She does not see him as someone who *cannot* connect, but as someone who has "*refused* to connect."<sup>12</sup> His refusal gives her reason to be hostile; were he *unable* to connect, the case would be different.

It could be asked why Margaret, if she really believes that Henry's conduct embodies relationship-impairing values, *addresses* blame to him instead of finally rejecting him? Why does he get the opportunity to respond? I believe that it would be an overintellectualization to think that one stops caring as soon as one believes that an offender's conduct embodies relationship-impairing values. Because Margaret cares, she wants to be certain of what she believes before ending the relationship. As long as she is not certain, that is, has not ruled out the possibility that Henry's conduct does not embody relationship-impairing values, her

11 Both my view and Bagley's can be understood as specifications of McKenna's account. McKenna writes that, when we address blame, the blamed agent "is afforded the opportunity to offer some explanation" ("Directed Blame and Conversation," 133). This can be specified in at least two ways. According to Bagley, we *invite* them to offer some explanation; we ask for it. According to my view, we make a *claim that allows for* a response. In both cases, we make a move in a conversation (a question or invitation versus a claim), and in both cases the blamed agent is afforded the opportunity to make a further move by offering some explanation.

12 Forster, *Howards End*, 350, emphasis added.

emotional engagement is fitting (I agree with Bagley here).<sup>13</sup> This implies that, in my view (as opposed to Bagley's), there is an asymmetry between the justification of emotional engagement and the justification of hostility: while hostility cannot be justified by the belief that it is possible that another will oppose me, emotional engagement can be justified by the belief that it is possible that another will recognize my demands. I cannot say more about this consequence of my view here, but it seems intuitively right.

If hostility is justified by what we take to be a fact about the offender's values, am I then not justifying hostile *non*-addressed blame as well? Bagley suggests that non-addressed blame is *not* hostile, but how could that be the case if it involves the certainty that an offender's values are relationship impairing?<sup>14</sup> I cannot do justice to the complexity of *non*-addressed blame here, but my point about emotional engagement helps to sketch a tentative answer. As long as we merely *believe* that an offender's values are relationship impairing, we do not rule out the possibility that they are not, and emotional engagement is fitting. When we are *certain* that an offender's values are relationship impairing, the possibility that they are not has been ruled out, and emotional engagement is no longer justified. But if emotional engagement is not justified, then neither is hostility, because hostility *requires*, or is even a kind of, emotional engagement.

My alternative account of addressed blame preserves a crucial difference between addressed and non-addressed blame, thus incorporating Bagley's main insight: addressed blame is proleptic (Henry has the opportunity to respond) in a sense in which non-addressed blame is not, and the proleptic character of addressed blame helps to avoid his dilemma. According to Bagley, that addressed blame is proleptic implies that there is "an essential connection between addressed blame and indeterminacy."<sup>15</sup> I agree, if "indeterminate" means that we do not rule out the possibility that, in response to our blame, the offender might convincingly specify their values in a way that conflicts with our way of specifying them. I disagree if "indeterminate" means that, in addressing blame, we raise a question or manifest uncertainty about the offender's values.<sup>16</sup>

13 On the relation between belief and certainty, as I see it here, see Hacker, *The Intellectual Powers*, 170–74, 208–9. "Someone who is certain . . . has ruled out the possibility of things not being so" (172). "To say that one believes that things are so is, among other things, to imply that not all doubt can rationally be excluded, *even though one has no doubts*" (208–9).

14 Bagley, "Properly Proleptic Blame," 856.

15 Bagley, "Properly Proleptic Blame," 854.

16 See Bagley's formulations "raising certain questions" and "the sense of uncertainty and tension that [addressed] blame distinctively registers" ("Properly Proleptic Blame," 875, 882).



## 4. AN OBJECTION

Bagley writes in a footnote that “conduct warranting addressed blame certainly reveals relationship-impairing attitudes, just not relationship-impairing values.”<sup>17</sup> In light of this remark, it is possible to understand Margaret, in addressing her blame, as saying “Your conduct has revealed relationship-impairing attitudes, but it is not clear whether it has revealed relationship-impairing values.” This reading seems to leave Bagley’s account of addressed blame intact, and one may think it helpful in order to correct Bagley’s claim that the justification of addressed blame’s hostility is to be found in a possibility about Henry’s values: maybe what justifies Margaret’s hostility is not her belief in a *possibility* about Henry’s *values*, but her belief that his conduct has *actually* revealed relationship-impairing *attitudes*.

There is a difficulty about the difference between relationship-impairing attitudes and relationship-impairing values. According to Bagley, Margaret’s conclusion that Henry’s conduct embodies relationship-impairing values is the conclusion that she can no longer have a loving relationship with him. It justifies final rejection. Similarly, relationship-impairing attitudes might be understood as attitudes that make a relationship impossible. If they do, it becomes unclear why, in Bagley’s view, relationship-impairing attitudes do not justify final rejection, while relationship-impairing values do. So he must mean something different by “relationship-impairing attitudes.” Here is another possibility. Suppose that a friend promises to water my plant while I am away but then forgets about it. According to Bagley, this is conduct warranting addressed blame, and so it can be said to reveal relationship-impairing attitudes.<sup>18</sup> But if this is what relationship-impairing attitudes are, they do not justify hostility. Suppose that my friend has a perfectly good explanation for why she forgot to water the plant. Then our relationship may not be impaired at all, to use the term “relationship impairing” for the attitude revealed in her conduct is misleading, and hostility was never justified. The bare fact that someone has done something cannot justify hostility as long as we are unclear about its *meaning*, about whether it expresses ill will or not.

The most plausible interpretation of “relationship-impairing attitudes” lies somewhere in between the two interpretations mentioned above: to reveal relationship-impairing attitudes is to reveal poor quality of will. It is not yet to reveal relationship-impairing values, because the latter require the poor quality of will to be rooted in underlying evaluative commitments. Suppose that we un-

17 Bagley, “Properly Proleptic Blame,” 874.

18 Bagley, “Properly Proleptic Blame,” 875.

derstand the relation between relationship-impairing attitudes and values in this way. Would it then be helpful to understand Margaret, in addressing her blame, as saying “Your conduct has revealed relationship-impairing attitudes, but it is not clear whether it has revealed relationship-impairing values,” where the first part accounts for her hostility? I see some problems with this reading. First, Margaret is not saying that she is unclear about something. It could be argued, in response to this, that the “but it is not clear” part should be left out of Margaret’s addressed blame. The suggestion would then be that her addressed blame expresses something like “Your conduct has revealed relationship-impairing attitudes,” while her non-addressed blame is rather like “Your conduct has revealed relationship-impairing values.” But this suggestion cannot be Bagley’s: there is no open question, no invitation, no uncertainty. Moreover, it does not capture what Margaret communicates to Henry. She does not just sum up disconnected cases where Henry has shown disregard; she suggests that these cases exhibit a *pattern* (“a man who insults his wife . . . cants with her memory . . . ruins a woman for pleasure . . . casts her off to ruin other men . . . gives bad financial advice”), and that is why her addressed blame is closer to “Your conduct has revealed relationship-impairing values.”

## 5. CONCLUSION

I conclude that Bagley’s account of addressed blame makes it difficult to understand why addressed blame is hostile. My alternative suggestion (1) makes better sense of Bagley’s paradigm example of addressed blame, (2) avoids Bagley’s dilemma in the way Bagley’s original solution does, because it preserves addressed blame’s proleptic character, and (3) can account for addressed blame’s elements of emotional engagement *and* hostility.<sup>19</sup>

KU Leuven

*benjamin.demesel@kuleuven.be*

## REFERENCES

Alvarez, Maria. “Reasons for Action: Justification, Motivation, Explanation.”

<sup>19</sup> This work was funded by the Research Foundation Flanders (FWO). I would like to thank Samuel Reis-Dennis and two anonymous reviewers for very helpful comments and suggestions.

- Stanford Encyclopedia of Philosophy* (Winter 2017). <https://plato.stanford.edu/entries/reasons-just-vs-expl/>.
- Bagley, Benjamin. "Properly Proleptic Blame." *Ethics* 127, no. 4 (July 2017): 852–82.
- Forster, E. M. *Howards End*. 1910. London: Penguin, 2012.
- Hacker, P. M. S. *The Intellectual Powers: A Study of Human Nature*. Oxford: Wiley-Blackwell, 2013.
- McKenna, Michael. "Directed Blame and Conversation." In *Blame: Its Nature and Norms*, edited by D. Justin Coates and Neal A. Tognazzini, 119–40. Oxford: Oxford University Press, 2013.
- Williams, Bernard. "Internal and External Reasons." In *Moral Luck*, 101–13. Cambridge: Cambridge University Press, 1981.