

# JOURNAL *of* ETHICS & SOCIAL PHILOSOPHY

VOLUME XXVIII · NUMBER 1

*July 2024*

## ARTICLES

Letting Animals Off the Hook

*Nicolas Delon*

The Challenge for Coronavirus Vaccine Testing

*Bastian Steuwer*

The Person as Environmentally Integrated:  
Dementia, Loss, and Extended Cognition

*Matilda Carter*

Murderers on the Ballot Paper: Bad Apples,  
Moral Compromise, and the Epistemic Value  
of Public Deliberation in Representative  
Democracies

*Richard Beadon Williams*

Three Kinds of Prioritarianism

*Carlos Soto*

## DISCUSSION

Not Living My Best Life: A Reply to Masny

*Guy Fletcher*



JOURNAL *of* ETHICS  
& SOCIAL PHILOSOPHY

VOLUME XVIII · NUMBER 1

*July 2024*

ARTICLES

- 1 Letting Animals Off the Hook  
*Nicolas Delon*
- 30 The Challenge for Coronavirus Vaccine Testing  
*Bastian Steuwer*
- 53 The Person as Environmentally Integrated:  
Dementia, Loss, and Extended Cognition  
*Matilda Carter*
- 83 Murderers on the Ballot Paper: Bad Apples,  
Moral Compromise, and the Epistemic Value  
of Public Deliberation in Representative  
Democracies  
*Richard Beadon Williams*
- 110 Three Kinds of Prioritarianism  
*Carlos Soto*

DISCUSSION

- 141 Not Living My Best Life: A Reply to Masny  
*Guy Fletcher*

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

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

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 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.

*Editors*

Sarah Paul  
Matthew Silverstein

*Associate Editors*

Rima Basu  
Saba Bazargan-Forward  
Brian Berkey  
Ben Bramble  
Dale Dorsey  
James Dreier  
Julia Driver  
Anca Gheaus  
Alex Gregory  
Sean Ingham  
Anthony Laden  
Errol Lord  
Coleen Macnamara  
Elinor Mason  
Simon Căbulea May  
Hille Paakkunainen  
David Plunkett  
Mark van Roojen  
Kevin Toh  
Vanessa Wills

*Discussion Notes Editor*

Tristram McPherson

*Managing Editor*

Chico Park

*Copy Editor*

Lisa Gourd

*Proofreader*

Susan Wampler

*Typesetter*

Matthew Silverstein

*Editorial Board*

Elizabeth Anderson

David Brink

John Broome

Joshua Cohen

Jonathan Dancy

John Finnis

Leslie Green

Karen Jones

Frances Kamm

Will Kymlicka

Matthew Liao

Kasper Lippert-Rasmussen

Stephen Perry

Philip Pettit

Gerald Postema

Henry Richardson

Thomas M. Scanlon

Tamar Schapiro

David Schmidtz

Russ Shafer-Landau

Tommie Shelby

Sarah Stroud

Valerie Tiberius

Peter Vallentyne

Gary Watson

Kit Wellman

Susan Wolf

## LETTING ANIMALS OFF THE HOOK

*Nicolas Delon*

WHAT KINDS of moral agents are there? Computers and cars are not moral agents. Now imagine you are walking past a playground. The children are agents, but you are reluctant to hold them morally responsible. Likewise, maybe the dogs playing at the park are agents of some sort but not *moral* agents.<sup>1</sup> They are playing by some tacit rules, but those are not moral. The realm of agents is larger than that of moral agents. Only the latter are open to attributions of moral responsibility and reactive attitudes. Parents are morally responsible and can be blameworthy for what happens to their children and their dogs. But if children or dogs fight at the park, they may be reprimanded, not held morally responsible. There are important differences between our responses to children and dogs. For instance, children will normally become moral agents; dogs will not. Children need and dogs need not be brought into scaffolding practices where we hold each other accountable and raise budding agents. The standards we apply to children are sensitive not just to what they are but also to what they are starting to become and the contexts in which they grow up.

Consider cases of nonhuman animals (henceforth “animals”) engaging in prosocial helping. These are anecdotes, but they are numerous enough to warrant consideration, and they illustrate growing evidence collected in laboratory and field settings in various species. On a busy highway in Chile, a dog has been hit by a vehicle and lies unconscious in the middle of the road. Another dog weaves in and out of the traffic and manages to drag the dog to safety.<sup>2</sup> Chimpanzees will sometimes help conspecifics without any direct benefit to themselves. In a remarkable video shot in Uganda at a busy road crossing, dominant male

1 Agency does not entail moral agency. For recent work on animal agency, see Arruda and Povinelli, “Two Ways of Relating to (and Acting for) Reasons”; Delon, “Animal Agency, Captivity, and Meaning”; Jamieson, “Animal Agency”; Sebo, “Agency and Moral Status”; Thomas, *Animal Ethics and the Autonomous Animal Self*; Wilcox, “Animals and the Agency Account of Moral Status.”

2 Rowlands, *Can Animals Be Moral?* 6.

chimpanzees aid females and youth to cross safely.<sup>3</sup> Animals such as apes, elephants, cetaceans, and corvids seem to engage in mourning behavior, expressing curiosity, distress, and perhaps grief around the corpses of conspecifics.<sup>4</sup> African and Asian elephants are known to manifest concern over distressed or deceased individuals, assisting the ailing and showing a special interest in dead bodies of their kind. Elephants have demonstrated a capacity for empathic understanding through coalition formation, the offering of protection and comfort to others, retrieving and “babysitting” calves, aiding individuals that would otherwise have difficulty moving, and removing foreign objects attached to others. Moreover, helping and empathetic behavior are not restricted to closely related kin.<sup>5</sup> A female elephant, Grace, was observed trying to help the dying matriarch of another family and distressed when unable to do so effectively.<sup>6</sup>

A growing literature documents animal “proto-morality.” Many primates exhibit “building blocks of morality”: empathy, consolation, conflict resolution, cooperation, and fairness (or inequity aversion).<sup>7</sup> While animals lack full-blown morality, they manifest behavior that is genuinely prosocial and other regarding. Ethologist Marc Bekoff and philosopher Jessica Pierce have argued that many species can follow moral norms (of empathy, fairness, cooperation, and mutual help) but that such norms are species specific: there is human morality, wolf morality, rat morality, and so on. “Animals are moral agents within the limited context of their own communities.”<sup>8</sup> Alongside the empirical literature, philosophical work on animal morality has blossomed.<sup>9</sup>

3 Hockings et al., “Road Crossing in Chimpanzees” (discussed in Andrews and Gruen, “Empathy in Other Apes”). For links to the videos, see <https://www.youtube.com/watch?v=-HJTG6RRN4E> (dogs); <https://www.youtube.com/watch?v=f5QopWSeeZc> (chimpanzees).

4 Gruen, “Death as a Social Harm”; King, *How Animals Grieve*; and Monsó and Osuna-Mascaró, “Death Is Common, So Is Understanding It.”

5 Byrne et al., “Do Elephants Show Empathy?”; Douglas-Hamilton et al., “Behavioural Reactions of Elephants towards a Dying and Deceased Matriarch”; and Plotnik and de Waal, “Asian Elephants (*Elephas maximus*) Reassure Others in Distress.”

6 Douglas-Hamilton et al., “Behavioural Reactions of Elephants towards a Dying and Deceased Matriarch.”

7 De Waal, *Primates and Philosophers* and *The Age of Empathy*; and Flack and de Waal, “Any Animal Whatever.”

8 Bekoff and Pierce, *Wild Justice*, 44.

9 Andrews and Gruen, “Empathy in Other Apes”; Back, “Are Animals Moral?”; Behdadi, “A Practice-Focused Case for Animal Moral Agency”; Clement, “Animals and Moral Agency”; Ferrin, “Nonhuman Animals Are Morally Responsible”; Fitzpatrick, “Animal Morality”; Monsó, “Empathy and Morality in Behaviour Readers” and “Morality without Mindreading”; Monsó and Andrews, “Animal Moral Psychologies”; Monsó, Benz-Schwarzburg, and Bremhorst, “Animal Morality”; Monsó and Wrage, “Tactful Animals”; Musschenga,

This paper seeks to refine our conceptual understanding of the animal morality debate. What would it take for animals to be moral agents, for their conduct to have moral worth? Can animals, as Mark Rowlands argues, act for moral reasons? If so, what do we appraise morally—the act, the motivations, the character? I will force a dilemma on the view that animals can act for moral reasons. If they can, resisting their moral responsibility requires more work if we want to preserve an intermediate category of moral subjects: those who act for moral reasons but are not moral agents. We will need fine-grained conceptual distinctions that may weaken the meaning of “acting for moral reasons,” thus undermining the category of moral subjects. Thus, animals are either *less moral* or *more responsible* than many in the animal morality debate argue. I proceed as follows. Section 1 reconstructs Rowlands’s influential theory of animal morality. The reconstruction leads to a dilemma that puts pressure on the demarcation between moral subjects and moral agents (section 2). I draw on what is known as the *Quality of Will* theory of responsibility for a few reasons. Whereas it originally ruled out animals, some theories of animal morality have explicitly appealed to it, and it bears striking similarities to Rowlands’s view. I argue that even theories of animal morality purporting to eschew claims of responsibility face pressure from *Quality of Will*. Section 3 considers two ways of defusing the dilemma and accommodating moral subjects—by claiming that responsibility has different degrees or faces, respectively. I conclude with some optimism about the liberal horn and recommend some revisions to make it more palatable.

### 1. ANIMAL MORALITY

The inference from prosocial behavior to responsibility is typically blocked by a missing necessary condition: a capacity for deliberation or reflective assessment of motivations, or an understanding of moral concepts.<sup>10</sup> Even arguments that animals could be virtuous stop short of asserting responsibility.<sup>11</sup> Commonly accepted grounds of responsibility include an agent’s actions originating in a reasons-responsive mechanism or being the product of self-government or

---

“Moral Animals and Moral Responsibility”; Rowlands, *Can Animals Be Moral?*; Shapiro, “Moral Agency in Other Animals”; Shupe, “Punishing Moral Animals”; and Vincent, Ring, and Andrews, “Normative Practices of Other Animals.” For earlier arguments, see Clark, *The Nature of the Beast*; DeGrazia, *Taking Animals Seriously*, 199–204; Pluhar, *Beyond Prejudice*, 55; and Sapontzis, *Morals, Reason, and Animals*, 28–33.

10 Korsgaard, “Morality and the Distinctiveness of Human Action”; and Dixon, *Animals, Emotions, and Morality*.

11 Clark, *The Nature of the Beast*; DeGrazia, *Taking Animals Seriously*; and Sapontzis, *Morals, Reason, and Animals*.

conscious deliberation. What matters is that some property demarcates candidates for responsibility from others, even if there exist borderline cases, such as children and psychopaths.

However, the demarcation only holds if moral responsibility does hinge on such features. If there is continuity between animal and human behavior and responsibility does not require conscious deliberation, then what, if anything, blocks the inference? Much of human behavior is automatic, habitual, affective, and opaque and is nonetheless open to moral appraisal.<sup>12</sup> It is then tempting to conclude that animals are open to similar forms of moral appraisal. Thus, work on animal morality suggests that animal behavior may be open to appraisal relative at least to group-specific norms—rules delimiting appropriate behavior within the social group, according to which individuals sometimes evaluate and sanction each other. Even when they do not, we can perform the evaluation.

On the other hand, we could be concerned about the collapse of the demarcation. It is beyond the scope of this paper to argue that seeing animals as morally responsible imposes unfair burdens on them, but let me sketch a rationale. Intuitively, claims about animal morality are not degrading or disrespectful; they do not objectify animals or reinforce prejudices about their inferiority. Quite the opposite. So whence the worry? Elsewhere, I argue that the inductive risk associated with mistakenly attributing morality to other animals is not negligible.<sup>13</sup> Recent work on methodology in animal cognition focuses on the risks associated with failing to ascribe certain cognitive capacities to other animals.<sup>14</sup> The risk of overattribution is usually considered worse than that of underattribution, but this recent work rightly argues that prioritizing false negatives over false positives is misguided for reasons both scientific and ethical. Both are errors, and the former can have high ethical costs. Still, there are risks to attributing capacities that animals lack.<sup>15</sup> Some studies suggest that attributing morally laden capacities to animals can affect our attitudes. Jared Piazza, Justin Landy, and Geoffrey Goodwin have found that perception of harmfulness (having a harmful as opposed to benevolent disposition relative to human welfare) negatively affects attributions of moral standing,

12 Arpaly, *Unprincipled Virtue*; Cova, “Two Kinds of Moral Competence”; Ferrin, “Good Moral Judgment and Decision-Making without Deliberation” and “Nonhuman Animals Are Morally Responsible”; Musschenga, “Moral Animals and Moral Responsibility”; and Railton, “The Affective Dog and Its Rational Tale.”

13 Delon, “Animal Morality and Epistemic Risks.”

14 Andrews and Huss, “Anthropomorphism, Anthropectomy, and the Null Hypothesis”; Birch, “Animal Sentience and the Precautionary Principle”; and Mikhalevich, “Experiment and Animal Minds.”

15 Birch, “Animal Cognition and Human Values.”

independently of animals' sentience or intelligence. Because "perceiving an animal as having a benevolent disposition enhances people's moral consideration for that animal, which is likely to promote better treatment of it," seeing animals as moral or immoral could have unforeseen consequences.<sup>16</sup> If moral subjects are not just moral patients but deserve distinctive protections and respect, or being able to exercise one's moral abilities is constitutive of flourishing, then it matters how we see them.<sup>17</sup> How work on animal morality can affect our treatment of animals is an open question—can it warrant punishment or third-party intervention? After all, morality has many sides, and not all moral animals play nice—predation, aggression, and callousness are pervasive. Our perception of predators could change if we saw them as moral agents. We might see chimpanzees, dolphins, and orcas as sometimes immoral. Our attitudes to coyotes and even wolves, already considered a nuisance by farmers and the US Fish and Wildlife Service, could deteriorate. Such moral costs must be part of inductive risk assessments. I will thus work on the assumption that the costs of overattributing moral characteristics deserve serious scrutiny. With these caveats in the background, let us look closely at the case for animal morality.

The abovementioned anecdotes, for Rowlands, "form parts of a large and growing body of evidence for the claim that some animals can exhibit moral behavior."<sup>18</sup> Most scientists and philosophers deny that possibility by setting stronger conditions on moral behavior: *X can act morally if and only if X can be morally responsible*, and responsibility requires metacognitive abilities that animals lack. While endorsing a standard, reflective picture of moral responsibility, Rowlands argues, *pace* Korsgaard and Dixon, that animals *can* act for reasons despite lacking metacognition. Animals are "motivated to act by moral reasons, not merely causes . . . where these reasons take the form of emotions with identifiable moral content."<sup>19</sup> We can reconstruct Rowlands's reasoning as follows:

1. To be a moral subject is to be motivated to act by moral considerations, which provide reasons for those actions.
2. Moral considerations can take the form of *morally laden emotions*.
3. An emotion is morally laden if it tracks a moral evaluation or judgment as part of its content.
4. Some emotions in some animals have evaluative content under some plausible description.

16 Piazza, Landy, and Goodwin, "Cruel Nature," 121.

17 Rowlands, *Can Animals Be Moral?* 248–54; and Monsó, Benz-Schwarzburg, and Bremhorst, "Animal Morality."

18 Rowlands, "Moral Subjects," 469.

19 Rowlands, *Can Animals Be Moral?* 35.

5. Therefore, some animals are capable of morally laden emotions (from 3 and 4).
6. Therefore, morally laden emotions provide motivating reasons for animals to act (from 1, 2, and 5).
7. Therefore, animals can be moral subjects.<sup>20</sup>

A crucial premise (2) is that animals are capable of *morally laden emotions*, intentional states with identifiable moral content such as “This creature’s distress is bad.” Such emotions have two components (3): cognitive (a representation of a state of affairs) and evaluative (an affective valence). Rowlands uses an intricate “tracking” strategy for ascribing content. It consists in using sentences as “*de dicto*” ascriptions of content to ourselves to explain the behavior of animals.<sup>21</sup> A similar strategy applies to evaluative content. “Emotions, if they are legitimate, track true evaluative propositions, but they do not require that the subject of an emotion entertain, or even be capable of entertaining, such a proposition.<sup>22</sup> Animals can experience moral emotions but cannot form moral judgments:

An emotion, *E*, is morally laden if and only if (1) it is an emotion in the intentional, content-involving, sense, (2) there exists a proposition, *p*, which expresses a moral claim, and (3) if *E* is not misguided, then *p* is true.<sup>23</sup>

The claim that *E* tracks *p* means that there is a truth-preserving relation between *E* and *p* such that *p* is true whenever *E* is correct.<sup>24</sup> Thus, tracking allows us to assess emotions for correctness.<sup>25</sup>

Suppose Rowlands is correct that emotions involve intentional content such that they can (in)correctly represent. Emotions also motivate. A controversial aspect of Rowlands’s view, granted for the sake of the argument, is his externalism about moral motivation. If emotions are responsive to reasons, they can be motivations that track moral reasons, *even if* the subject does not or cannot entertain such reasons. Responsiveness to reasons is responsiveness to morally relevant objective features of the world, such as suffering or distress.

20 Rowlands, *Can Animals Be Moral?* 33–35. I depart slightly from his four-part “unpacked” argument.

21 Rowlands, *Can Animals Be Moral?* 57.

22 Rowlands, *Can Animals Be Moral?* 67.

23 Rowlands, *Can Animals Be Moral?* 69.

24 For an application of the strategy to care and empathy, see Monsó, “Morality without Mindreading.”

25 One subtle difference: in cognitive tracking, the animal does have a belief; in evaluative tracking, the animal’s emotion simply *tracks* the evaluative proposition.

In sum, emotions constitute morally evaluable motivations if they represent features of the world that happen to be reasons for the animal's conduct and if they are efficacious.

Rowlands argues that some animals can be moral *subjects* even though only human beings are moral *agents*. A "minimal moral subject" meets the following sufficient conditions:

*X* is a moral subject if *X* possesses (1) a sensitivity to the good- or bad-making features of situations, where (2) this sensitivity can be normatively assessed, and (3) is grounded in the operations of a reliable mechanism (a "moral module")... Moral subjects are ... sensitive to the good- and bad-making features of situations in the sense that they entertain intentional content emotionally.<sup>26</sup>

In contrast, "the extent to which one is an agent is the extent to which one understands what one is doing, the likely consequences of what one is doing, and how to evaluate those consequences."<sup>27</sup> Moral agency and moral subjecthood are "logically independent." Moral agents possess further capacities to understand *that* certain motives and actions are right or wrong and *why*.<sup>28</sup> Rowlands concedes that moral agency, being a function of understanding, may come in degrees. Yet animals can be moral agents "to such a small extent that, if we were to think of agency as a categorical matter ... then we would almost certainly say [they are] not an agent at all."

Rowlands's key move is to dissociate moral evaluation and responsibility, making moral subjecthood a "desirable" category, and several authors concur.<sup>29</sup> If certain facts or properties can be evaluated morally without presupposing a responsible agent, then animals lacking moral agency may still be open to the evaluation of their behavior or motivations if they are reliably responsive to moral reasons. Remember that reasons need not play a conscious or deliberative role in the animal's mental life.<sup>30</sup> The reasons, however, are implicit in

26 Rowlands, *Can Animals Be Moral?* 230–31.

27 Rowlands, *Can Animals Be Moral?* 240.

28 Rowlands, *Can Animals Be Moral?* 243.

29 Rowlands, "Moral Subjects," 471. Cf. Burgis, "Making Covenants with Brute Beasts"; Monsó, "Empathy and Morality in Behaviour Readers" and "Morality without Mindreading"; and Monsó, Benz-Schwarzburg, and Bremhorst. "Animal Morality."

30 For Monsó, Rowlands endorses realism about moral facts ("Empathy and Morality in Behaviour Readers," 676). To me, notwithstanding his commitment to "a reasonably robust sense of ethical objectivity," he is only committed to *reasons externalism* (Rowlands, "Moral Subjects," 472). The good- or bad-making features are independent of "the subjective states of the *agent*." This is different than saying they are mind independent.

the phenomenology of their emotions. Another's distress is experienced as unpleasant and motivates one to engage in affiliative behavior. The badness of distress, and its motivational pull, color the subject's experience of the target's distress.<sup>31</sup> This makes the presumed "moral module" efficacious.

Consider the Chilean dog again:

His companion lies unconscious on a busy road. This is, let us suppose, a bad-making feature of the situation. The first requirement is that the would-be rescuer is sensitive to this bad-making feature. Such sensitivity does not require that the dog is able to think thoughts such as "This is bad!" The appropriate sensitivity can, in fact, be purchased by other means [e.g., empathetic capacities or response to distress]... Nowhere in this general picture is there any suggestion that the dog has control over his sentiments, still less that he is able to critically scrutinize them.<sup>32</sup>

The last bit is crucial. Much of Rowlands' argument consists in burden-shifting, aimed at the scrutiny-control-normativity-motivation (SCNM) schema or nexus.<sup>33</sup> The initial appeal of the idea that morality depends on metacognitive abilities "rests on the fallacy of the miracle-of-the-meta."<sup>34</sup> According to SCNM,

the ability to critically scrutinize one's motivations gives one control over them. This control permits these motivations to make a normative claim on their subject, and so makes them the sort of motivations that might be moral.<sup>35</sup>

Rowlands argues at length that the appeal to control leads to regress and rests on confusion about its role in making motivations normative. His central thesis is that the moral value of an action is logically distinct from the blame- or praiseworthiness of the agent. While the latter requires control and so, perhaps, metacognition, the former does not. Thus, an animal's motivation can be moral without metacognition.

Moral motivations may come cheap, but Rowlands has not argued that responsibility requires metacognition. Nor has he shown that moral subjecthood is not sufficient for responsibility. The dilemma arises from dismantling the SCNM nexus: weakening the conditions for having moral motivations weakens the conditions for responsibility; on the other hand, reinstating stringent conditions on

31 Monsó, "Morality without Mindreading," 351.

32 Rowlands, "Moral Subjects," 473.

33 Rowlands, *Can Animals Be Moral?* chs. 6 and 7, and "Moral Subjects."

34 Rowlands, *Can Animals Be Moral?* 189.

35 Rowlands, "Moral Subjects," 470.

the latter presupposes something like SCNM. It is unclear why a stringent view of responsibility would welcome an entirely separate (“logically independent”) category of moral evaluation. I will return to this possibility in section 3.

Before presenting the dilemma, let us recap. There are moral agents and mere agents. Some mere agents are moral subjects, causally responsible for their actions, not morally responsible, yet capable of acting for moral reasons. The challenge is to prevent sufficient conditions for responsibility from trickling down into our evaluations of moral subjects. As noted, being too liberal with our attributions is risky, so we should be wary of expanding the scope of responsibility without sufficient epistemic and practical reason.

## 2. ANIMALS ON THE HOOK

### 2.1. Rowlands’s Dilemma

Rowlands’s argument, when combined with certain views about moral responsibility, entails that some animals can be morally responsible. My argument does not generalize to all theories of responsibility, but its focus is not arbitrary. First, the view I focus on, Quality of Will, bears revealing parallels to Rowlands’s picture of moral motivation. Furthermore, it is a prominent theory, as a quick glance at recent discussions of responsibility responses and reactive attitudes shows. Maybe the best theory of responsibility does not entail that all moral subjects are also moral agents. But since Rowlands does not offer or endorse a positive conception of moral responsibility, the question is open. In any case, we can take the forthcoming argument to be conditional on the plausibility of Quality of Will.

Nomy Arpaly’s influential account of “moral worth” brings the problem into relief. On her account, blameworthiness and praiseworthiness are constitutive of moral responsibility and depend on responsiveness to moral reasons, which is manifested by a depth of concern for what *happens to be* moral rather than what an agent *takes to be* moral. I will consider each of these features shortly. Admittedly, Arpaly does not share Rowlands’s view of animals’ capacity to act for moral reasons. She writes:

Creatures not acting for reasons at all cannot be either morally praiseworthy or morally blameworthy. . . . One cannot blame or praise a creature who cannot be expected to perceive the morally relevant features of situations any more than an elephant can be expected to perceive legal factors, aesthetic factors, or contexts in which a baseball player should not bunt.<sup>36</sup>

Rowlands would agree. But she also writes:

<sup>36</sup> Arpaly, *Unprincipled Virtue*, 131.

The dog's mind presumably cannot grasp—nor can it track, the way even unsophisticated people can—such things as increasing utility, respecting persons, or even friendship. . . . Thus, even if this animal can act for reasons, to some extent, it cannot respond to *moral reasons*, even though it may occasionally come close.<sup>37</sup>

If animals were responsive to reasons, they could be open to moral praise and blame, but responsiveness to reasons requires a capacity for moral concern, which presupposes conceptual understanding. Importantly, animals are not blame- or praiseworthy according to Arpaly, not because they lack “agent-autonomy,” the capacity to reflect, deliberate, and determine their motives (she denies that responsibility presupposes autonomy), but because acting for moral reasons requires more demanding cognitive capacities than it does according to Rowlands.

For Rowlands, some animals are responsive to moral reasons. They lack “understanding,” but according to his conception of reasons responsiveness and moral content, morally laden emotions are sufficient for moral motivation. If so, some animals are capable of what Arpaly calls “moral concern.” However, if this really is moral concern, then, by the same token, animals are morally responsible. If they are not, then they are not reasons responsive. Both pressures are real. As noted, the empirical evidence for animal proto-morality is growing. The same evidence suggests that, maybe, some animals could be moral agents. Rowlands has only shown that animals can be moral without being responsible *given some disputed theoretical demarcation*. This is not to say the demarcation is unacceptable but simply that the category of moral subject hinges on theoretical commitments. If we lower the standards for moral subjecthood, why not also lower the standards for moral agency?

It is interesting that Rowlands's qualms regarding *human* responsibility surface throughout the book. While claiming that at least most humans but no animals can be moral agents, he seems to think that the standard picture of agency is too demanding even for us. So, if human beings are morally responsible, then maybe we should reconsider our criteria for responsibility. But if we do, we risk collapsing the moral subjecthood/agency distinction. This is *Rowlands's dilemma*:

*Liberal Horn*: Accept moral subjecthood *and* moral agency for some animals.

*Conservative Horn*: Deny moral agency for animals *but also* deflate the meaning of moral subjecthood.

37 Arpaly, *Unprincipled Virtue*, 146.

The rest of the paper will motivate each horn and attempt to defuse the dilemma. The next section motivates the liberal horn, which proponents of animal morality should find the most attractive.

## 2.2. *Quality of Will*

According to Quality of Will,

1. A person is morally responsible for an action when that action expresses her quality of will, that is, her goodwill, ill will, or indifference or lack of concern.
2. Goodwill consists in attitudes such as a desire for the right or the good or a concern for what is morally good or right.<sup>38</sup>

Importantly, the agent acting with goodwill is responsive to moral reasons *de re*, that is, to what happen to be reasons for the action rather than the fact that it is good; the content of the agent's attitude is not *de dicto* concern for morality.<sup>39</sup> Mark Twain's Huckleberry Finn is praised for helping his slave friend Jim escape, even though Huckleberry views himself as flouting what he believes to be the right reasons (property rights, the law). His praiseworthiness derives from his being responsive to moral considerations *de re*. He does the right thing (helping Jim escape) for the right reason (Jim is a friend and a person) but without consciously entertaining this being the right thing as his motivating reason. It is not that he is not deliberating. He is, in fact, torn. But his acting upon the right reasons is not the product of his deliberative process. Had it been, Jim might have concluded he was doing the *wrong* thing!<sup>40</sup>

Rowlands denies that 2 above is a necessary condition for moral evaluation. Animals can have moral motivations even without any understanding of the

38 Arpaly, *Unprincipled Virtue*; McKenna, *Conversation and Responsibility*; and Strawson, "Freedom and Resentment." David Shoemaker distinguishes between three interpretations of "quality of will": *character*, *judgment*, or *regard*, yielding three "noncompeting conceptions of responsibility" and targets for distinct subsets of responsibility responses ("Qualities of Will"). Perhaps some animals exhibit quality of regard since they have affective and cognitive attitudes such as seeing a conspecific as being in distress and to be helped or a companion as worthy of trust and reciprocity.

39 Arpaly, *Unprincipled Virtue*; and Markovits, "Acting for the Right Reasons."

40 In contrast, for Johnson King, Finn's act lacks moral worth because he is *accidentally* doing the right thing; he has no idea that he is performing an act of the right type ("Accidentally Doing the Right Thing"). Rather, he is motivated by the right-making features but does not understand the relationship between those features and the act's rightness. Moral worth requires *deliberately* doing the right thing. As a reviewer notes, this criticism, which would otherwise block the liberal horn of the dilemma, is not compatible with Rowlands's tracking account of moral motivation, since tracking is reliable. According to Johnson King's more demanding view, animals' behavior cannot be moral.

concepts of right or wrong. But while some Quality of Will views do require some such understanding for responsibility, this is not a core commitment. Moral concern is understood *de re* rather than *de dicto*. Insofar as an animal is motivated by what makes an action right, she has the required kind of concern. Moreover, Rowlands's tracking strategy seems specifically designed to allow for such attributions. The relevant moral proposition is implied by an animal's having the relevant moral emotion that does not misfire.

If goodwill does not require autonomy, can animals manifest it? Rowlands's tracking strategy enables the attribution of identifiable moral content to animals. Animals are responsive (*de re*) to features of the environment that our (*de dicto*) attributions identify as morally relevant: he is my buddy, she helped me last time, he is in distress, and so on. Jennifer Lynn Burgis and Asia Ferrin argue that animals can manifest goodwill, thus taking the liberal horn of the dilemma. Burgis specifically argues that some animals can understand morally relevant considerations (*de re*) by Arpaly's lights.<sup>41</sup> Recall the example of Grace the elephant. She was acting for the right reasons in manifesting (*de re*) concern for the welfare of the matriarch, acting upon motivations whose content is responsive to moral considerations. She likely experienced empathy (distress by proxy) and sympathy (other-regarding concern) for a group mate in distress. If such content is sufficient for goodwill, and if autonomous deliberation is not necessary, Grace is responsible according to Quality of Will. To express goodwill is to act for the right reasons, as Grace seems to have done.

Ferrin draws on the empirical literature to argue that empathetic capacities are sufficient for the capacity to act for moral reasons.<sup>42</sup> Frans de Waal's Russian doll metaphor describes layers of empathy, from (1) state-matching (emotional contagion) at the core to (2) sympathetic concern (consolation) to (3) perspective taking (targeted helping) on the outside.<sup>43</sup> Many animals exhibit at least 1, including rodents; many primates at least 2.<sup>44</sup> Ferrin defends two claims. First, empathy (affective and cognitive) is sufficient for moral action, especially responsiveness to others' states. Second, both affective and cognitive empathy are found to various degrees across species, including apes, cetaceans, and elephants. These animals meet the criteria for manifesting quality of will. Accordingly, their actions can have moral worth.

41 Burgis, "Making Covenants with Brute Beasts," 132.

42 Ferrin, "Nonhuman Animals Are Morally Responsible," 138–42. Also see Ferrin, "Good Moral Judgment and Decision-Making without Deliberation."

43 De Waal, *The Age of Empathy*.

44 On rodents, see Bartal, Decety, and Mason, "Empathy and Pro-social Behavior in Rats."

By the same token, Grace could have *failed* to show proper concern for the matriarch. And if she is morally responsible, she may be blameworthy. Oddly, such an implication is rarely considered. Work on animal morality typically focuses on morally admirable behavior.<sup>45</sup> While for Rowlands, “praise would be an inappropriate attitude to bear toward [moral subjects],” for Burgis, moral animals are open to praise but not blame.<sup>46</sup>

More generally, there is a widely held praise/blame asymmetry.<sup>47</sup> Some argue that praise and blame have different control conditions—the ability to do otherwise is a condition of blame but not praise.<sup>48</sup> One can be blameworthy only if one had alternative possibilities, while one can be praiseworthy even if one did not. Praise merely requires acting for the right reasons; the ability to do otherwise is a more stringent condition. On this view, some animals could meet the conditions for praiseworthiness but not blameworthiness because they lack control-relevant abilities but can still act for the right reasons. Rowlands would agree with the verdict but cannot avail himself of this justification of the asymmetry since he rejects control as a condition of moral evaluation. Moreover, theories of responsibility do not distinguish between moral subjects and moral agents, so it is unclear how these justifications mesh with his view.

A different but related assumption is that a higher bar must be met for blame than praise.<sup>49</sup> Indeed, the risks of harm are lower in praising than in blaming mistakenly: praise tends to benefit the target; blame tends to harm.<sup>50</sup> The asymmetry is reinforced by the fact that our access to animals’ motivations is opaque, so we should be charitable about their motivations. This echoes the caution favoring false positives over false negatives in animal research. But these are epistemic and pragmatic considerations that do not bear on whether animals are *worthy* of blame or praise.

A natural thought is that elephants cannot express ill will when failing to help others in distress. Yet one could argue that orcas tormenting baby seals

45 Though see Monsó, “Is Predation Necessarily Amoral?”; Monsó and Wrage, “Tactful Animals”; and Shupe, “Punishing Moral Animals.”

46 Rowlands, *Can Animals Be Moral?* 142; and Burgis, “Making Covenants with Brute Beasts,” 132.

47 For an empirical review, see Anderson, Crockett, and Pizzaro, “A Theory of Moral Praise.”

48 Nelkin, “Responsibility and Rational Abilities”; and Wolf, “Asymmetrical Freedom.”

49 Vilhauer, “Free Will and the Asymmetrical Justifiability of Holding Morally Responsible”; and Wolf, “Asymmetrical Freedom.”

50 Argetsinger, “Blame for Me and Not for Thee”; Mackenzie, “Culpability, Blame, and the Moral Dynamics of Social Power”; McGeer, “Co-reactive Attitudes and the Making of Moral Community”; and Pickard, “Responsibility without Blame” and “Responsibility without Blame for Addiction.”

and chimpanzees brutally killing infant chimps are manifesting what seems like ill will, cruelty, or indifference toward suffering.<sup>51</sup> Still, we usually assume that animals hurting or failing to help others do not manifest such attitudes. Animals are at least excused when their actions fail to express proper concern, either because they lack a crucial capacity or because of their circumstances (e.g., diet, scarcity). And so we admire or praise the nice chimpanzees and let the nasty ones off the hook.

### 2.3. *Accountability*

Ferrin writes, “though animals are sometimes morally responsible, we may not be able to engage in practices of holding them responsible given the communication barrier and lack of overlapping social context.”<sup>52</sup> On the other hand, “some animals seem to experience reactive attitudes toward each other such as resentment, indignation, hurt feelings, anger, gratitude, reciprocal love, and forgiveness.” Ferrin suggests that animals are likely responsible to each other (“intraspecies accountability”) but not across species boundaries (“interspecies accountability”). We probably should not *hold* animals responsible; reactive attitudes are only locally applicable by and to group members, even if *we* can recognize that their actions have moral worth. Thus, the recognition of moral subjecthood in other animals may entail intra- but not interspecies responsibility.

Dorna Behdadi takes a different route to the conclusion that some animals, who participate in “moral responsibility practices” (MRPs), are accountable to each other.<sup>53</sup> Behdadi’s alternative to “capacity-focused approaches” sees moral agency as “the participation in certain social, inter-relational practices” and argues specifically, from evidence on canine cognition and social play, that canids participate in MRPs and hence are moral agents. According to practice-focused approaches (which overlap with Quality of Will), participants in MRPs “share a strong disposition to internalize norms and to participate in the attitudes, expressions, and practices that surround them.”<sup>54</sup> Canids are competent participants in canid normative “communicatory practices,” which are “a relevant analog to at least *some* forms of moral exchange in terms of *asking for reasons, explanations, or acknowledgment and responding by providing*

51 Monsó, “Is Predation Necessarily Amoral?”; and Monsó and Wrage, “Tactful Animals.”

52 Ferrin, “Nonhuman Animals Are Morally Responsible,” 146.

53 Behdadi, “A Practice-Focused Case for Animal Moral Agency.”

54 Behdadi, “A Practice-Focused Case for Animal Moral Agency,” 230. For the overlap with Quality of Will, see Strawson, “Freedom and Resentment”; McGeer, “Co-reactive Attitudes and the Making of Moral Community”; McKenna, *Conversation and Responsibility*; and Vargas, *Building Better Beings*.

*explanations, excuses, or acknowledging transgressions.*<sup>55</sup> Canids can thus be appropriate targets of blame when shared norm communication is possible—when canids are, as a Strawsonian could put it, potential moral interlocutors.<sup>56</sup> (Indeed, the dispositions and inclinations relevant to MRPs coincide with abilities enabling quality of will. Accordingly, they can adopt something akin to the Strawsonian “participant attitude” to each other.)

Ferrin’s and Behdadi’s views have two implications. First, morality is species specific, and the evaluation of moral subjects is relative to context.<sup>57</sup> Moral subjects are off the hook relative to us.<sup>58</sup> By the same token, moral subjects internalize different norms and act for different reasons than we do. We may identify whether and when they act for moral reasons, but our respective spaces of moral reasons may not overlap much. So, we lack standing to adopt the participant attitude toward them. The second implication, however, is that if interspecies communication and sufficient social overlap could be secured, interspecies accountability would make sense. Perhaps our “relations of mutual trust and affection” with companion animals provide such a context.<sup>59</sup>

#### 2.4. *Protecting Moral Subjects*

In sum, we have philosophical and empirical reasons to extend Quality of Will to some animals, but the meaning and scope of these animals’ responsibility remain unclear. My argument turns on the plausibility of Quality of Will, and since Rowlands does not discuss it, I can only surmise what his response would be. Two cases he has offered to maintain the separation between moral motivation and moral responsibility will help.

First, consider the real-life case of Robert Thompson and Jon Venables, two ten-year-old English boys who on February 12, 1993, abducted, tortured, and murdered three-year-old Jamie Bulger. Thompson and Venables became “the youngest convicted murderers in English history.”<sup>60</sup> As Rowlands notes, “under questioning, they revealed that they had planned to abduct and murder a child that day,” so we presume they acted intentionally and were motivated to inflict suffering and kill. Even though, because of their age, they fell below the threshold of responsibility, Rowlands expects the reader to agree that their motivations were morally evil.

55 Behdadi, “A Practice-Focused Case for Animal Moral Agency,” 236.

56 Watson, “Responsibility and the Limits of Evil.”

57 Bekoff and Pierce, *Wild Justice*.

58 Though see Shupe, “Punishing Moral Animals.”

59 Scanlon, *Moral Dimensions*, 166.

60 Rowlands, “Moral Subjects,” 471.

Unfortunately, Rowlands's only supporting claim is that denying that the boys had morally bad intentions, "if one is not in the grip of a peculiarly warped moral psychology, is as counterintuitive as a claim can get."<sup>61</sup> Let us concede, then, that their motivations were evil. Does it not follow, according to Quality of Will, that they were somewhat morally responsible? They did the wrong thing for the wrong reasons. The question is whether their motivations were morally laden. They clearly manifested a lack of moral concern, but should we expect the boys to manifest such concern? No less but also no more than what we expect of moral subjects. Indeed, for Rowlands, their motivations would be evil even if the boys were mentally ill or under the influence of factors beyond their control. They are moral subjects, open to moral evaluation, but not moral agents. Why not hold them accountable? The condition of their exemption is that they are children, though it is worth noting that according to a Strawsonian view, extreme evil serves as its own exempting condition by placing wrongdoers outside the bounds of the moral community. Gary Watson underscored the ambivalence of extreme evil between antipathy and sympathy, blame and exemption. This could be clouding our intuition regarding the boys.<sup>62</sup>

A few things cast doubt on the moral status of the boys' motivations, though. First, they were held legally responsible and convicted, presumably partly on account of their motivations. According to Peter Strawson and Watson, children gradually become moral agents even if they lack full moral understanding. This suggests that the subject/agent distinction is porous. If, however, the boys were not responsible, this is because the moral psychology we deploy to explain their behavior discounts the moral status of their motivations. They may be malicious or vicious, and we may justifiably harbor antipathy toward them, but not evidently in a moral sense. Whether such psychology is "warped," as Rowlands says, requires argument. The claim that their motivations are obviously immoral rather than pathological, made in support of the subject/agent distinction, lacks support.

61 Rowlands, "Moral Subjects," 471.

62 In "Responsibility and the Limits of Evil," Watson showed that Strawson's theory implies the paradox that evil counts as its own exemption condition (holding responsible requires moral address, which requires a potential moral interlocutor). On the one hand, we have standing to blame evil wrongdoers; on the other hand, heartless murderers such as Robert Harris do not seem capable of heeding our demands, and so cannot be morally addressed—we lack a shared framework of values. The alternative is to deny (*pace* Strawson) that responsibility requires membership in the moral community. According to Michael McKenna, while Harris is *not* a member of the moral community, he has the *capacity* to participate in it, which explains his responsibility ("The Limits of Evil and the Role of Moral Address"). It is not that he does not understand our values; he repudiates them.

A final point concerns Rowlands's appeal to parity.<sup>63</sup> Rowlands takes this sort of case to confirm the logical independence of moral motivation and responsibility. However, why should our attitudes to children carry over to other species? There may be pragmatic reasons to appraise the boys' motivations that will not apply to animals, such as the need for social order, plaintiffs' legal claims, or scaffolding practices of moral education.<sup>64</sup> We may separate moral evaluation from the fact of holding responsible and yet think that the former is functionally justified by responsibility practices. We turn children into members of the moral community by evaluating their motivations before they can even be held responsible. None of those facts apply to other animals. Hence, even if we concede that the boys' motivations were evil, it does not follow that moral subjecthood applies outside the context at hand.

The second case Rowlands discusses is that of Adolf Hitler in a world of hard determinism, where no one is morally responsible, "which may or may not be the actual world":

We might ... justifiably ... refuse to blame or hold him responsible for what he does. But refusing to classify his motivations as even falling into the category of the moral is highly counterintuitive.<sup>65</sup>

Granted, Hitler's moral motivations are abhorrent even under hard determinism. After all, we can see psychopaths' motivations as vicious while (sometimes) refraining from holding them responsible. But in what sense exactly are deterministic Hitler's motivations of the moral kind? Rowlands implies that determinism precludes control, including over one's motivations, and therefore responsibility, but that those motivations do not presuppose control to be morally appraisable. However, not only is this a controversial claim in the responsibility literature, but rejecting the control condition leads naturally to a view like Quality of Will and therefore the liberal horn of the dilemma. In the above cases, moral responsibility and moral motivation stand or fall together. Such cases make the distinction between moral subjecthood and moral agency intuitively plausible but cannot establish it without further argument. In contrast, Quality of Will has the theoretical virtue of harmonizing our judgments about the cases. Its simplicity is not decisive, but it explains the appearance that Thompson, Venables, and Hitler deserve blame even under circumstances that would normally count as exempting conditions.

63 Rowlands, "Moral Subjects," 473.

64 McGeer, "Co-reactive Attitudes and the Making of Moral Community"; and Vargas, *Building Better Beings*.

65 Rowlands, "Moral Subjects," 471.

To recap, Rowlands's dilemma was either accepting that animals can be morally responsible (the liberal horn) or deflating the import of moral subjecthood (the conservative horn). The dilemma arises from the combination of moral subjecthood with Quality of Will. In the remainder of the paper, I zoom out and lay the groundwork to make the prospect of animal responsibility less threatening. Once we understand what it does not entail, perhaps we will no longer need a subject/agent distinction and could take the liberal horn without worrying.

### 3. ANIMALS OFF THE HOOK

Each of the two “ways out” I will consider consists in protecting animal morality from the upward pressure of responsibility, or at least its practical implications: degrees of responsibility and aspects or “faces” of responsibility.

#### 3.1. *Off the Hook, First Pass: Degrees of Responsibility*

The idea that responsibility and blameworthiness can be a matter of degree is no longer controversial. Reactive attitudes should be sensitive to the degree to which an agent is responsible (i.e., competent and/or free from coercion or other responsibility-canceling influences) and the degree to which their action expresses the relevant ground of responsibility.

Some authors who have argued that animals can be moral agents have been careful to stress the significance of degrees. David DeGrazia writes, “the range over which a given being is responsible is determined by the range of action possibilities for which the being can understand a rule of conduct, roughly what its point is, the consequences of breaking it, and so on.”<sup>66</sup> Moreover, because different capacities are involved in its different aspects, agency varies according to which capacities one possesses and to what degree.<sup>67</sup> Whatever the required competence, it is gradable, and responsibility responses should vary accordingly.<sup>68</sup>

Most accounts of responsibility are amenable to degrees of responsibility. For instance, D. Justin Coates and Philip Swenson propose to amend the reasons-responsiveness account according to how receptive and reactive to reasons an agent is.<sup>69</sup> Quality of Will can adjust degrees of blameworthiness to the quality of the reasons for which agents act—namely degrees of good or ill will.

66 DeGrazia, *Taking Animals Seriously*, 203n107.

67 DeGrazia, *Taking Animals Seriously*, 204.

68 See, e.g., Shoemaker, “Qualities of Will.”

69 Coates and Swenson, “Reasons-Responsiveness and Degrees of Responsibility”; and Fischer and Ravizza, *Responsibility and Control*.

Blame, praise, resentment, indignation, or gratitude then vary accordingly.<sup>70</sup> A proponent of animal morality could thus deny that their view entails that animals must be subject to the same attitudes we direct toward moral agents. If animals are just barely competent, and their actions are minimally morally worthy, then their responsibility need not trigger the same responses normal attributions of responsibility do.

Consider affective motivations, a core component of nonreflective and sentimentalist approaches to animal morality.<sup>71</sup> We can describe their content and appraise their quality in a graded fashion: for example, how much concern for or sensitivity to the distress of others a creature's conduct manifests or how reliably responsive to morally significant situations it is. Thus, even if animals were morally responsible, they might not be very blameworthy or praiseworthy, let alone answerable to us. The range of potential moral worth of their actions may be as limited as the range of their quality of will or reasons responsiveness.

Though attractive, this response will not insulate animals from the outward expression of reactive attitudes. Graded responses are difficult to maintain in practice. People often express reactive attitudes toward beings who should be exempt, such as children and mentally disabled people. We also miscalibrate our responses to people with impaired agency, such as addicts and patients with personality disorders, which is why some advocate for "responsibility without blame."<sup>72</sup> Sometimes, there are good reasons for holding some reactive attitudes. Strawson distinguishes between the "objective" attitude—in which we predict, manage, or control others—and the "participant" attitude—in which we hold each other to account. And he notes that

parents and others concerned with the care and upbringing of young children ... are dealing with creatures who are potentially and increasingly capable both of holding, and being objects of, the full range of human and moral attitudes but are not yet truly capable of either. The treatment of such creatures must therefore represent a kind of compromise, constantly shifting in one direction, between objectivity of attitude and developed human attitudes.<sup>73</sup>

70 Tierney, "Quality of Reasons and Degrees of Responsibility."

71 Andrews and Gruen, "Empathy in Other Apes"; de Waal, *The Age of Empathy*; Ferrin, "Nonhuman Animals Are Morally Responsible"; Monsó, "Morality without Mindreading"; Monsó, Benz-Schwarzburg, and Bremhorst, "Animal Morality"; and Rowlands, *Can Animals Be Moral?*

72 Pickard, "Responsibility without Blame" and "Responsibility without Blame for Addiction."

73 Strawson, "Freedom and Resentment," 19.

But because our attitudes are “constantly shifting,” one should expect some involuntary leakage. As with children, so with animals—we might end up blaming moral subjects when we should not. Indeed, moral subjecthood invites us to shed the objective attitude toward animals.

A reply to this concern is that the excesses of our blaming practices are just that—unjustified—and we should seek to correct them by calling for compassion or understanding instead of blame and indignation. For instance, Hanna Pickard argues that we should refrain from blaming drug addicts while keeping them responsible, because it matters for their own sake that we do so.<sup>74</sup> The difference is that holding animals responsible does not benefit them the way it does people whose agency is impaired; it is not guided by the end of recovery or rehabilitation.

In sum, degrees of responsibility do not dissolve the dilemma. Either animals are moral subjects because they can act for moral reasons, but then they are morally responsible or we will, in practice, be tempted to express some responsibility responses; or animals can only be responsible to a benign degree, but then the content of their motivations is shallower than we might have thought. Can we mitigate the implications of taking the liberal horn of the dilemma by drawing some finer-grained distinctions?

### 3.2. *Off the Hook, Second Pass: Faces of Responsibility*

Start with a distinction between *being* responsible and *holding* responsible, or between reasons to judge that a creature is responsible and reasons, in practice, to hold them responsible. Angela Smith makes the distinction to argue that our attributions of responsibility should not be sensitive to the same considerations that count for or against responding in certain ways, typically with reactive attitudes, to someone being responsible.<sup>75</sup> The question of whether a creature is responsible is distinct from whether it would be fair or appropriate to blame her, even if blameworthiness is conceptually tied to responsibility. There is also a difference between judging someone to be blameworthy and expressing blame, let alone punishing.

Taking degrees of responsibility and the distinction between responsibility *judgments* and *responses*, we might avoid the implication that we should express much by way of reactive attitudes toward animals for their morally good or bad deeds. As Watson notes, “holding people responsible . . . also involves a social setting in which we demand (require) certain conduct from one another and

74 Pickard, “Responsibility without Blame for Addiction.”

75 Smith, “On Being Responsible and Holding Responsible.”

respond adversely to one another's failures to comply with these demands."<sup>76</sup> Since Strawson, the moral responsibility literature has echoed the idea that responsibility responses presuppose a capacity to participate in interpersonal relationships and the moral community.<sup>77</sup> If animals are not implicated in this social setting, then we need not hold them responsible, even if they are. Remember the emphasis on intraspecies accountability by Behdadi, Bekoff and Pierce, and Ferrin. Rowlands might argue that the objectivity of the moral facts that moral subjects are tracking allows us to appraise moral subjects even without a shared social setting. It is also plausible that the shared-social-setting requirement applies to responsibility but not subjecthood. Either way, more needs to be said about the ethical standards that should inform our appraisal of animals of different species.

Consider another helpful distinction between *attributability* and *accountability*. Attributability reflects what Watson calls the aretaic face of responsibility (from the Greek *arete*, meaning excellence), whereas accountability (to others) involves reactive attitudes, holding responsible, which implies believing and acting as if the responsible person is accountable to us or others.<sup>78</sup>

Suppose (*pace* Watson) that we can engage in the aretaic appraisal of animals, morally appreciating their character, their excellences and defects, their virtues and vices. This implies judging them as the authors of their conduct—that their actions are attributable to them. A dog could be foolish or courageous, and it could be appropriate for us to express our approval or disapproval of their behavior but not appropriate to hold them responsible—to *demand* that they answer to us or the moral community. For, as Watson explains, the intelligibility of demanding presumes the interlocutor's understanding. The reactive attitudes are "incipiently forms of communication" or "moral address." But young children and animals are incapable, the argument goes, of understanding "the basic demand."<sup>79</sup>

Thus, the dog's conduct could reflect well or poorly on them, they could be a moral subject, but we may not infer that they are responsible—praiseworthy or blameworthy—for their conduct. The distinction could honor the distinct category of moral subjects. The question is, again, whether we can maintain, in practice, a clear demarcation between those different kinds of judgments.

Watson draws a clear line. He denies that animals are susceptible to "aretaic appraisal," which applies to "one's purposes, ends, choices, concerns, cares,

76 Watson, "Two Faces of Responsibility," 229.

77 McKenna, *Conversation and Responsibility*.

78 Watson, "Two Faces of Responsibility," 231.

79 Watson, "Responsibility and the Limits of Evil."

attachments, and commitments” and hence “presupposes moral capacity, the capacity for adopting and pursuing ends.”<sup>80</sup> John Martin Fischer and Neal A. Tognazzini, in their own “physiognomy of responsibility,” concur: “By asking whether the agent is open to, or is a ‘sensible target’ of, aretaic appraisal, we are asking whether the agent exercised the capacities required to make the agent *the sort of creature whom it might make sense to appraise aretaically*,” which excludes dogs. A dog’s “viciousness” is not moral viciousness and not attributable to her, because she cannot intend to hurt or manifest a lack of moral concern for others, unlike “certain psychopaths, who can indeed have specifically moral intentions.”<sup>81</sup>

There is, however, evidence that chimpanzees, orcas, and bottlenose dolphins can intentionally hurt each other, perhaps manifesting negative moral emotions such as cruelty, envy, or resentment.<sup>82</sup> If Rowlands is correct, moral subjects possess the required capacities. They are capable of flexible, intentional behavior and moral emotions that reliably track morally relevant features of situations. If so, we should accommodate nonhuman moral subjects within our “physiognomy of responsibility.”

It is plausible that humans and animals are exempt on different grounds—psychopaths because they cannot respond to moral reasons, although their actions are still “attributable to them in an aretaic sense”; animals because they cannot entertain moral reasons.<sup>83</sup> Unlike psychopaths, and perhaps like children, moral subjects’ conduct and motivations are presumed to be *responsive to reasons*.<sup>84</sup> We can, using Rowlands’s tracking strategy, reconstruct rational standards for moral subjects’ conduct. In contrast, psychopaths can cognitively grasp moral reasons but fail to properly respond to them. This reveals a tension: psychopaths could be appraised aretaically, but animals could not, even though animals can respond to moral reasons.

More plausibly, barring excuses or justification, moral subjecthood gives us standing to hold, if not express, some reactive attitudes toward some animals. Moral subjects should earn from us more than the objective stance of the

80 Watson, “Two Faces of Responsibility,” 244–45.

81 Fischer and Tognazzini, “The Physiognomy of Responsibility,” 384. Cf. Wolf, *Freedom within Reason*, 63.

82 Monsó and Wrage, “Tactful Animals,” 16–17.

83 Fischer and Tognazzini, “The Physiognomy of Responsibility,” 387.

84 Does the dilemma also apply to children? Are they either less moral or more responsible than we think? If young children can act for moral reasons, then we can draw the distinctions discussed above or say that children are at least partially responsible. We could also take one horn (conservative) for animals and another for children (liberal). Either way, we may need to recalibrate our attitudes if moral subjecthood is how we operationalize our evaluations of children and animals.

ethologist yet less than the participant stance of ordinary responsibility practices. If our responses are likely to misfire, moral subjecthood calls for revisions to our responsibility model. Still, moral subjects *are* responsible in some sense (aretaic attributability). One could also endorse a practice-based, agency cultivation model of responsibility in which animals play no part, but even those views presuppose a quality-of-will account of blame.<sup>85</sup> And if our practices purport to nurture and develop responsible agents, the question becomes what we should make of moral subjects. Our ordinary practices involve different but pervasive kinds of interspecies interactions, from companion animals to currently and formerly farmed animals to animals in the wild. We cannot just assume that no context gives rise to responsibility responses. As Vargas notes, “distinct forms of acculturation provide agents with differential capacities to recognize and respond to moral considerations in different contexts.”<sup>86</sup> The question then becomes one of “moral ecology”:

Once we look beyond intrinsic features of agents to the wider set of relations that structure the various capacities of interest to us, we find that moral ecology matters. . . . The circumstances that support and enable exercises of agency in ways that respect and reflect a concern for morality.<sup>87</sup>

Before concluding, I would like to briefly consider a final way to avert the dilemma. Recent work on normativity suggests that several species of primates possess normative competence.<sup>88</sup> The range of norms includes norms of obedience, reciprocity, care, social responsibility, and solidarity of various forms. In chimpanzees, norm compliance is not external and accidental but is internalized and rests on norm-sensitive motivations. The evidence is growing more generally that normative behavior extends far beyond apes, cetaceans, and elephants to canids, corvids, and rodents.<sup>89</sup> This literature suggests that (some) animals respond to normative reasons, but it also offers an alternative: animals could be *normative* without being *moral*. If animals can respond to norms, and their motivations form part of their excellences, they may qualify for nonmoral

85 McGeer, “Co-reactive Attitudes and the Making of Moral Community”; and Vargas, *Building Better Beings*.

86 Vargas, *Building Better Beings*, 245.

87 Vargas, *Building Better Beings*, 246.

88 Andrews, “Naïve Normativity”; Fitzpatrick, “Chimpanzee Normativity”; Monsó and Andrews, “Animal Moral Psychologies”; and Vincent, Ring, and Andrews, “Normative Practices of Other Animals.”

89 Monsó and Andrews, “Animal Moral Psychologies”; Monsó, “Morality without Mindreading”; and Monsó and Wrage, “Tactful Animals.”

aretaic appraisal. This would involve replacing the category of moral subjects with that of normative animals.

These various replies defuse, to an extent, concerns about the liberal horn of the dilemma. I started writing this article inclined to see these replies as making ad hoc distinctions against a backdrop of continuity between humans and other animals that motivated moral subjecthood. I am now inclined to embrace them. After all, theorists of responsibility believe that different senses of responsibility track important facts about responsibility. We could avail ourselves of these distinctions and conclude that according to Quality of Will, animals can be apt targets of aretaic appraisal but should not be held accountable for their actions. And perhaps that is how it should be. Aretaic appraisals are less burdensome than accountability and seem less morally risky. If that is how it should be, then we might defuse the liberal horn of the dilemma after all. And we could do this while granting my working assumption that responsibility is burdensome. A broader concern about our psychology subsists, though: we often shift within the multifarious physiognomy of responsibility unwittingly, especially when our norms are ill defined, as they are with animals. Deep facts about responsibility notwithstanding, we should tread carefully when it comes to moral subjecthood.

#### 4. CONCLUSION

Knowing whether animals can be moral agents is morally important: moral agents have interests in exercising their moral agency and may have obligations. Some argue that there is a middle ground between mere agency and moral agency: moral subjects, who can act for moral reasons without being morally responsible. Others argue that animals can be responsible but only within their communities. I have put pressure on both views to generate a dilemma: on the liberal horn, the demarcation between moral subjecthood and responsibility dissipates; on the conservative horn, insulating animals from responsibility deflates the significance of moral subjecthood. By drawing finer-grained distinctions, I have sketched a few ways to let animals off the hook—praise and blame asymmetry, degrees and faces of responsibility, and normativity without morality—to clarify the possibilities and identify areas where more conceptual work is needed. Whether or not animals are moral, we owe them credit where it is due, but only there.<sup>90</sup>

*College of Charleston*  
*delonn@cofc.edu*

90 Thanks to Diane Michelfelder and her students and colleagues at Macalester College, Louise Daoust and her students at Eckerd College, and the audience and organizers of the Normative Animals conference at York University. Susana Monsó provided valuable

## REFERENCES

- Anderson, Rajen A., Molly J. Crockett, and David A. Pizarro. "A Theory of Moral Praise." *Trends in Cognitive Sciences* 24, no. 9 (September 2020): 694–703.
- Andrews, Kristin. "Naïve Normativity: The Social Foundation of Moral Cognition." *Journal of the American Philosophical Association* 6, no. 1 (Spring 2020): 36–56.
- Andrews, Kristin, and Lori Gruen. "Empathy in Other Apes." In *Empathy and Morality*, edited by H. L. Maibom, 193–209. Oxford: Oxford University Press, 2014.
- Andrews, Kristin, and Brian Huss. "Anthropomorphism, Anthropectomy, and the Null Hypothesis." *Biology and Philosophy* 29 (September 2014): 711–29.
- Argetsinger, Henry. "Blame for Me and Not for Thee: Status Sensitivity and Moral Responsibility." *Ethical Theory and Moral Practice* 25 (April 2022): 265–82.
- Arpaly, Nomy. *Unprincipled Virtue: An Inquiry into Moral Agency*. Oxford: Oxford University Press, 2003.
- Arruda, Caroline T., and Daniel J. Povinelli. "Two Ways of Relating to (and Acting for) Reasons." *Mind and Language* 33, no. 5 (November 2018): 441–59.
- Back, Youngsun. "Are Animals Moral? Zhu Xi and Jeong Yakyong's Views on Nonhuman Animals." *Asian Philosophy* 28, no. 2 (May 2018): 97–116.
- Bartal, Inbal Ben-Ami, Jean Decety, and Peggy Mason. "Empathy and Pro-Social Behavior in Rats." *Science* 334, no. 6061 (2011): 1427–30.
- Behdadi, Dorna. "A Practice-Focused Case for Animal Moral Agency." *Journal of Applied Philosophy* 38, no. 2 (December 2021): 226–43.
- Bekoff, Marc, and Jessica Pierce. *Wild Justice: The Moral Lives of Animals*. Chicago: University of Chicago Press, 2009.
- Birch, Jonathan. "Animal Cognition and Human Values." *Philosophy of Science* 85, no. 5 (December 2018): 1026–37.
- . "Animal Sentience and the Precautionary Principle." *Animal Sentience* 16, no. 1 (2017).
- Burgis, Jennifer Lynn. "Making Covenants with Brute Beasts: Making Room for Non-Human Animals in a Contractualist Framework." Ph.D. diss., University of Miami, 2018. <https://scholarship.miami.edu/esploro/outputs/doctoral/>

---

feedback on early drafts. I am also indebted to several anonymous referees' comments and encouragements. Ideas related to this paper were discussed at the Rocky Mountain Ethics Congress at the University of Colorado, Boulder, and at New College of Florida. The author received financial support from the Animal Welfare Fund of the Center for Effective Altruism. Alicia Bednar offered valuable assistance as part of the 2021 Cross-College Alliance's Environmental Discovery Awards Program.

- Making-Covenants-with-Brute-Beasts-Making/991031447234502976.
- Byrne, Richard, P. C. Lee, N. Njiraini, J. H. Poole, K. Sayialel, S. Sayialel, L. A. Bates, and C. J. Moss. "Do Elephants Show Empathy?" *Journal of Consciousness Studies* 15, nos. 10–11 (2008): 204–25.
- Clark, Stephen R. L., *The Nature of the Beast: Are Animals Moral?* Oxford: Oxford University Press, 1984.
- Clement, Grace. "Animals and Moral Agency: The Recent Debate and Its Implications." *Journal of Animal Ethics* 3, no. 1 (Spring 2013): 1–14.
- Coates, D. Justin, and Philip Swenson. "Reasons-Responsiveness and Degrees of Responsibility." *Philosophical Studies* 165, no. 2 (September 2013): 629–45.
- Cova, Florian. "Two Kinds of Moral Competence: Moral Agent, Moral Judge." In *What Makes Us Moral? On the Capacities and Conditions for Being Moral*, edited by Bert Musschenga and Anton van Harskamp, 117–30. Dordrecht: Springer Netherlands, 2013.
- DeGrazia, David. *Taking Animals Seriously: Mental Life and Moral Status*. Cambridge: Cambridge University Press, 1996.
- Delon, Nicolas. "Animal Agency, Captivity, and Meaning." *Harvard Review of Philosophy* 25 (2018): 127–46.
- . "Animal Morality and Epistemic Risks." Unpublished manuscript.
- De Waal, Frans B. M. *The Age of Empathy: Nature's Lessons for a Kinder Society*. Toronto: McClelland and Stewart, 2009.
- . *Primates and Philosophers: How Morality Evolved*. Princeton: Princeton University Press, 2006.
- Dixon, Beth. *Animals, Emotions, and Morality*. New York: Prometheus Books, 2008.
- Douglas-Hamilton, Iain, Shivani Bhalla, George Wittemyer, and Fritz Vollrath. "Behavioural Reactions of Elephants towards a Dying and Deceased Matriarch." *Applied Animal Behaviour Science* 100 (October 2006): 67–102.
- Ferrin, Asia. "Good Moral Judgment and Decision-Making without Deliberation." *Southern Journal of Philosophy* 55, no. 1 (March 2017): 68–95.
- . "Nonhuman Animals Are Morally Responsible." *American Philosophical Quarterly* 56, no. 2 (April 2019): 135–54.
- Fischer, John Martin, and Mark Ravizza. *Responsibility and Control: A Theory of Moral Responsibility*. Cambridge: Cambridge University Press, 1998.
- Fischer, John Martin, and Neal A. Tognazzini. "The Physiognomy of Responsibility." *Philosophy and Phenomenological Research* 82, no. 2 (March 2011): 381–417.
- Fitzpatrick, Simon. "Animal Morality: What Is the Debate About?" *Biology and Philosophy* 32, no. 6 (December 2017): 1151–83.
- . "Chimpanzee Normativity: Evidence and Objections." *Biology and*

- Philosophy* 5, no. 4 (2020): 1–28.
- Flack, Jessica C., and Frans B. M. de Waal. “‘Any Animal Whatever’: Darwinian Building Blocks of Morality in Monkeys and Apes.” *Journal of Consciousness Studies* 7, no. 1 (January 2000): 1–29.
- Gruen, Lori. “Death as a Social Harm.” *Southern Journal of Philosophy* 52 (September 2014): 53–65.
- Hockings, Kimberly J., James R. Anderson, and Tetsuro Matsuzawa. “Road Crossing in Chimpanzees: A Risky Business.” *Current Biology* 16, no. 17 (September 2006): 668–70.
- Jamieson, Dale. “Animal Agency.” *Harvard Review of Philosophy* 25 (2018): 111–26.
- Johnson King, Zoë. “Accidentally Doing the Right Thing.” *Philosophy and Phenomenological Research* 100, no. 1 (January 2020): 186–206.
- King, Barbara J. *How Animals Grieve*. Chicago: University of Chicago Press, 2013.
- Korsgaard, Christine M. “Morality and the Distinctiveness of Human Action.” In de Waal, *Primates and Philosophers*, 98–119.
- Mackenzie, Catriona. “Culpability, Blame, and the Moral Dynamics of Social Power.” *Aristotelian Society Supplementary Volume* 95, no. 1 (July 2021): 163–82.
- Markovits, Julia. “Acting for the Right Reasons.” *Philosophical Review* 119, no. 2 (April 2010): 201–42.
- McGeer, Victoria. “Co-reactive Attitudes and the Making of Moral Community.” In *Emotions, Imagination and Moral Reasoning*, edited by Robyn Langdon and Catriona Mackenzie, 299–326. New York: Psychology Press, 2012.
- McKenna, Michael. *Conversation and Responsibility*. Oxford: Oxford University Press, 2012.
- . “The Limits of Evil and the Role of Moral Address: A Defense of Strawsonian Compatibilism.” *Journal of Ethics* 2, no. 2 (1998): 123–42.
- Mikhalevich, Irina. “Experiment and Animal Minds: Why the Choice of the Null Hypothesis Matters.” *Philosophy of Science* 82, no. 5 (December 2015): 1059–69.
- Monsó, Susana. “Empathy and Morality in Behaviour Readers.” *Biology and Philosophy* 30 (2015): 671–90.
- . “Is Predation Necessarily Amoral?” In *Crisis and Critique: Philosophical Analysis and Current Events*, edited by A. Siegetleitner, A. Oberprantacher, M. L. Frick, and U. Metschl, 367–82. Berlin: De Gruyter, 2021.
- . “Morality without Mindreading.” *Mind and Language* 32, no. 3 (June 2017): 338–57.
- Monsó, Susana, and Kristin Andrews. “Animal Moral Psychologies.” In *The*

- Oxford Handbook of Moral Psychology*, edited by Manuel Vargas and John M. Doris, 388–420. Oxford: Oxford University Press, 2022.
- Monsó, Susana, Judith Benz-Schwarzburg, and Annika Bremhorst. “Animal Morality: What It Means and Why It Matters.” *Journal of Ethics* 22, no. 3 (December 2018): 283–317.
- Monsó, Susana, and Antonio J. Osuna-Mascaró. “Death Is Common, So Is Understanding It: The Concept of Death in Other Species.” *Synthese* 199 (December 2021): 2251–75.
- Monsó, Susana, and Brite Wrage. “Tactful Animals: How the Study of Touch Can Inform the Animal Morality Debate.” *Philosophical Psychology* 34, no. 1 (January 2021): 1–27.
- Musschenga, Albert W. “Moral Animals and Moral Responsibility.” *Les ateliers de l'éthique/The Ethics Forum* 10, no. 2 (Summer 2015): 38–59.
- Nelkin, Dana. “Responsibility and Rational Abilities: Defending an Asymmetrical View.” *Pacific Philosophical Quarterly* 89, no. 4 (December 2008): 497–515.
- Piazza, Jared, Justin F. Landy, and Geoffrey P. Goodwin. “Cruel Nature: Harmfulness as an Important, Overlooked Dimension in Judgments of Moral Standing.” *Cognition* 131, no. 1 (April 2014): 108–24.
- Pickard, Hanna. “Responsibility without Blame: Empathy and the Effective Treatment of Personality Disorder.” *Philosophy, Psychiatry, and Psychology* 18, no. 3 (September 2011): 209–23.
- . “Responsibility without Blame for Addiction.” *Neuroethics* 10, no. 1 (April 2017): 169–80.
- Plotnik, Joshua M., and Frans de Waal. “Asian Elephants (*Elephas maximus*) Reassure Others in Distress.” *PeerJ Life and Environment* e278 (2014). <https://peerj.com/articles/278/>.
- Pluhar, Evelyn B. *Beyond Prejudice: The Moral Significance of Human and Non-human Animals*. Durham, NC: Duke University Press, 1995.
- Railton, Peter. “The Affective Dog and Its Rational Tale: Intuition and Attunement.” *Ethics* 124, no. 4 (July 2014): 813–59.
- Rowlands, Mark. *Can Animals Be Moral?* New York: Oxford University Press, 2012.
- . “Moral Subjects.” In *The Routledge Handbook of Philosophy of Animal Minds*, edited by Kristin Andrews and Jacob Beck, 469–74. New York: Routledge, 2017.
- Sapontzis, Stephen F. *Morals, Reason, and Animals*. Philadelphia: Temple University Press, 1987.
- Scanlon, Thomas M. *Moral Dimensions: Permissibility, Meaning, Blame*. Cambridge, MA: Harvard University Press, 2008.

- Sebo, Jeff. "Agency and Moral Status." *Journal of Moral Philosophy* 14, no. 1 (2017): 1–22.
- Shapiro, Paul. "Moral Agency in Other Animals." *Theoretical Medicine and Bioethics* 27, no. 4 (October 2006): 357–73.
- Shoemaker, David. "Qualities of Will." *Social Philosophy and Policy* 30, nos. 1–2 (Winter 2013): 95–120.
- Shupe, Eli. "Punishing Moral Animals." *Canadian Journal of Philosophy* 51, no. 5 (July 2021): 351–66.
- Smith, Angela. "On Being Responsible and Holding Responsible." *Journal of Ethics* 11, no. 2 (December 2007): 465–84.
- Strawson, Peter. "Freedom and Resentment." In *Freedom and Resentment and Other Essays*, 1–25. London: Methuen and Co., 1974.
- Thomas, Natalie. *Animal Ethics and the Autonomous Animal Self*. London: Palgrave Macmillan, 2016.
- Tierney, Hannah. "Quality of Reasons and Degrees of Responsibility." *Australasian Journal of Philosophy* 97, no. 4 (2019), 661–72.
- Vargas, Manuel. *Building Better Beings: A Theory of Moral Responsibility*. Oxford: Oxford University Press, 2013.
- Vilhauer, Benjamin. "Free Will and the Asymmetrical Justifiability of Holding Morally Responsible." *Philosophical Quarterly* 65, no. 261 (October 2015): 772–89.
- Vincent, Sarah, Rebecca Ring, and Kristin Andrews. "Normative Practices of Other Animals." In *The Routledge Handbook of Moral Epistemology*, edited by Aaron Zimmerman, Karen Jones, and Mark Timmons, 57–83. New York: Routledge, 2019.
- Watson, Gary. "Responsibility and the Limits of Evil: Variations on a Strawsonian Theme." In *Perspectives on Moral Responsibility*, edited by John Martin Fischer and Mark Ravizza, 119–48. Ithaca, NY: Cornell University Press, 1987.
- . "Two Faces of Responsibility." *Philosophical Topics* 24, no. 2 (1996): 227–48.
- Wilcox, Mark. "Animals and the Agency Account of Moral Status." *Philosophical Studies* 177, no. 7 (Fall 2020): 1879–99.
- Wolf, Susan. "Asymmetrical Freedom." *Journal of Philosophy* 77, no. 3 (March 1980): 151–66.
- . *Freedom within Reason*. New York: Oxford University Press, 1993.

## THE CHALLENGE FOR CORONAVIRUS VACCINE TESTING

*Bastian Steuwer*

FROM the early days of the COVID-19 pandemic, vaccines were considered the safest and most sustainable way out of the health and economic crisis of the pandemic. Researchers, policymakers, and bioethicists debated ways in which vaccine development could be expedited. One suggestion was human challenge trials in which volunteers are infected with the pathogen after having received either the candidate vaccine or a placebo or alternative control treatment.<sup>1</sup> The idea behind challenge trials is that because researchers do not need to wait for participants to be naturally infected (if ever), challenge trials promise faster results.<sup>2</sup> In a pandemic that induced much suffering, even small gains in time can be highly beneficial. Decision makers hesitated and opted for field trials first. When challenge trials started belatedly in the United Kingdom, safe and efficacious vaccines had already been developed. Was this hesitation justified?

The question is not only of retrospective interest. Pandemic preparedness has received renewed attention due to the salience and visibility of COVID-19, but also due to advances in biotechnology that some fear make pandemics more likely.<sup>3</sup> The question of the permissibility of challenge trials is then also a question of pandemic preparedness in our ethical frameworks and regulations. My argument, which focuses on COVID-19, has, therefore, lessons for future pandemics, too.

One key concern about accelerated testing was the risks to participants. I argue that challenge trials can be justified even on a framework for research ethics that is strongly protective of research subjects. Philosophical arguments for challenge trials have been made both on broadly consequentialist and anti-paternalistic grounds. These arguments were often critical of research

- 1 Eyal, Lipsitch, and Smith, "Human Challenge Studies to Accelerate Coronavirus Vaccine Licensure"; and Plotkin and Caplan, "Extraordinary Diseases Require Extraordinary Solutions."
- 2 Some scientists were less optimistic about the time advantage, pointing to the need to develop a strain of the virus that is not needed in field experiments. See Kahn et al., "For Now, It's Unethical to Use Human Challenge Studies for SARS-COV-2 Vaccine Development."
- 3 Pannu et al., "Strengthen Oversight of Risky Research on Pathogens."

ethics practice.<sup>4</sup> My argument develops a somewhat more sympathetic line of research ethics that not only permits challenge trials but also points to new options in vaccine research that could be useful, especially in pandemics with pathogens more dangerous to individuals than COVID-19. The argument also shows how nonconsequentialist and broadly contractualist moral theory can be an appealing way of thinking about the regulation of risk in medical research. Last, it highlights the connections between the risks to study participants, risks to study cohorts, and the benefits to nonparticipants.

In section 1, I start by discussing and developing ethical standards for clinical research risks. Applying these standards in sections 2 and 3, I argue that challenge trials can meet these standards. I also explain how a low-dosage challenge can render challenge trials permissible that initially appear too risky.<sup>5</sup> Section 4 turns to the question of post-challenge safety testing. I argue that a proposal for accelerated post-challenge safety testing is no more problematic than the established testing procedure.<sup>6</sup> Sections 5 and 6 discuss how and when benefits to nonparticipants can justify risks to participants of clinical research.

#### 1. WHEN ARE RISKS JUSTIFIABLE TO STUDY PARTICIPANTS?

The key concern about challenge studies is that they are overly risky for research subjects.<sup>7</sup> To take an extreme example, it would clearly be impermissible to subject willing volunteers to very high risks of death to find a cure for a minor cosmetic condition that affects only a few people worldwide. Research ethics expresses this idea with the requirement of a favorable risk–benefit ratio.<sup>8</sup> A favorable risk–benefit ratio is a necessary condition that must be met if clinical research is to be permissible. Clinical research is justified if the favorable risk–benefit ratio is satisfied alongside various non-risk-related conditions

4 Savulescu and Wilkinson, “Extreme Altruism in a Pandemic,” focuses on anti-paternalism. Eyal, “Is There an Upper Limit on Risks to Study Participants?” focuses on a broadly consequentialist approach highlighting large stakes. Other arguments like Chappell, “Pandemic Ethics and Status Quo Risk,” are not necessarily consequentialist but challenge the distinction between harms arising from research and harms arising from the pandemic.

5 See Steuwer, Jamroziak, and Eyal, “Prioritizing Second-Generation SARS-COV-2 Vaccines through Low-Dosage Challenge Studies.”

6 See Eyal, Gerhard, and Strom, “Strengthening and Accelerating SARS-COV-2 Vaccine Safety Surveillance.”

7 Deming et al., “Accelerating Development of SARS-COV-2 Vaccines”; and Shah et al., “Ethics of Controlled Human Infection to Address COVID-19.”

8 Emanuel, Wendler, and Grady, “What Makes Clinical Research Ethical?” 2705–6; Rid and Wendler, “A Framework for Risk-Benefit Evaluations in Biomedical Research.”

(informed consent, fair participant selection, etc.). I assume the latter conditions are satisfied in order to focus on the question of research risks.

However, the idea of a favorable risk–benefit ratio is in need of further details. When do benefits outweigh the risks? Which benefits should we take into consideration? How much priority should we give to reducing the risk to participants at the expense of forgoing benefits to nonparticipants? I now argue for three standards that fulfill the favorable risk–benefit ratio.

At times, the favorable risk–benefit ratio is interpreted as a requirement to provide favorable prospects to the participants of clinical research.<sup>9</sup> In other words, research is permissible only when undergoing the research is in the rational self-interest of the participants. It is highly controversial whether research must meet such a high bar, but it is easy to see why clinical research that meets this standard would be permissible. Researchers would be acting no differently from physicians who recommend to patients what they believe is in the best interest of the patient.<sup>10</sup> There can be little doubt about the ability of individuals to give informed consent to such research, just as there can be little doubt about the ability of individuals to give informed consent to medical procedures. Once the participants understand that the gamble is in their self-interest, they will typically consent to it. This is the *favorable prospect standard*.<sup>11</sup>

To see if any less demanding standard is justified, consider the role of informed consent in the aforementioned argument. Informed consent both licenses the risks associated with the research and licenses the necessary intrusions into one's body and privacy. Vaccines require access to a person's body; monitoring requires at least access to medical records. A second interpretation of the risk–benefit ratio limits the role of consent to this latter role. It asks, "Would the risk imposition be permissible if it could be done without invading the person's body and privacy?" Since informed consent means that individuals

- 9 For a good discussion on both the presence of this idea and how it conflicts with important parts of research practice, see Wikler, "Must Research Benefit Human Subjects If It Is to Be Permissible?"
- 10 This corresponds to what Rid and Wendler call the "informed clinician test" ("A Framework for Risk-Benefit Evaluations in Biomedical Research," 158–59). Rid and Wendler also point out that the right comparison is whether the testing is beneficial as compared to already existing interventions and not as compared to no intervention (157–59). This distinction raises interesting questions about the correct comparator. What about citizens of developing countries without access because other countries are hoarding vaccine supply? Unfortunately, I need to sidestep this question here.
- 11 A related idea is "clinical equipoise," which refers to the situation in which the researcher-clinician does not judge either option (participating/nonparticipating) to be better than the other. While the prospects are not favorable in such a case, they are not unfavorable either. See Weijer, "The Ethical Analysis of Risk."

waive their moral objection against these intrusions, we should ask whether the risk imposition by itself is justifiable. If the risk is such that we could have imposed it without consent, then there can be no objection against risks of this kind involved in research. This standard is not equivalent to the favorable prospect standard. Avoiding at all times all net risks of harm to individuals is an impossibly stringent requirement that would lead to paralysis. Many daily activities impose risks on others without any compensating benefit. In many of these activities, risks for some can be justified by benefits to others. For example, when we call an ambulance for an injured person, risks to bystanders potentially hurt by a car accident with the ambulance can be justified by benefits to the injured person.<sup>12</sup> But, of course, there are limits to the extent to which some can be put at risk of harm in order to provide benefits to others.

Therefore, we should focus on the question whether the risk of harm would wrong any individual participating in the research. The focus on wronging individuals also explains why the risking of active harm counts more heavily than the failure to prevent harm due to the COVID-19 pandemic. We need to take care not to wrong anyone. But as long as no individual is wronged by the risk imposition, we are permitted to impose risks with the aim of benefiting others.

The point can also be expressed in the language of rights. Any rights violation necessarily wrongs an individual. I am less certain whether every act of wronging an individual also constitutes a rights violation. In the case of risks, however, it does seem plausible that there is a right against the imposition of some risks.<sup>13</sup> If put in the language of rights, the earlier point is even clearer. Rights act as side constraints to our actions in pursuit of the social good, but if the side constraints are respected, we are free to pursue important social aims.

To ensure that our act of risking harm does not wrong any individual or violate their rights, we must ensure that our action can be justifiable to each of

12 The ambulance example does not involve net risks if we allow for so-called intrapersonal aggregation. That is, we would consider the costs and benefits of living with a principle that generally licenses an act or risk imposition. This renders more actions beneficial for all. T.M. Scanlon appeals to intrapersonal aggregation of this sort (*What We Owe to Each Other*, 197–202, and “Contractualism and Justification,” 24–25, 38–40). Moral theory would be understood as “legalist” in the sense that moral principles are seen as the equivalent of laws that generally govern human behavior. Liam Murphy discusses how contractualism ties up with legalist moral theory in “Nonlegislative Justification.” The question that the favorable prospect standard asks is, however, different. It asks whether, in this particular instance, the risk imposition is in the interest of the agent. This standard of not allowing net risks at any point in time surely leads to paralysis. I thank a reviewer for pressing me to highlight the distinction between net risks of principles allowing risky activities and net risks of individual risky actions.

13 This is argued for by John Oberdiek in *Imposing Risk*, ch. 4. It is also supported by Stephen Perry, “Harm, History, and Counterfactuals,” 1306–9.

them separately. If no single person can raise a valid complaint against the risk imposition, then we have ensured that the risk imposition is justifiable to each and every one. Consequently, the risk imposition does not wrong any single individual.<sup>14</sup> I call this the *justifiable risk standard*. Importantly, the justifiable risk standard requires the presence of benefits to nonparticipants in order to justify the research risks. This sets it apart from the favorable prospect standard and will become important in sections 4 and 5.

The connection to rights against being subjected to risks suggests a third way for clinical research to be ethical. This third way, as the justifiable risk standard, relies on the benefits to nonparticipants to justify research risks, but it does so in a different manner. If individuals can waive their rights to bodily integrity and privacy for the purposes of research, then why can they not waive their right against being subjected to risk, too? Individuals who participate in studies justified under the justifiable risk requirement do so for reasons other than their self-interest. Participants in research, in fact, often report being motivated by considerations other than their self-interest. Participants might be motivated by the desire to help others, to do their part in fighting a disease, or to do something meaningful with their lives. The motives of participants are important here not to evaluate the participants' conduct but rather because of the different justifications that researchers are able to give to individuals depending on whether or not the research is in the participant's clinical interests. If individuals are permitted to waive some rights in pursuit of altruistic motivations, then why should clinical research prevent them from waiving their rights against being subjected to risk in pursuit of them? Indeed, health care systems already accept the idea that individuals can waive their rights against being subjected to (substantial) risk. Around the world, health care systems accepted volunteers during the COVID-19 pandemic, knowing that volunteering to help exposed these individuals to additional risks they would not otherwise face. A good example is volunteers in emergency medical services who, in rural areas, are often exposed for a substantial time to the risk of infection while transporting suspected cases. These health care systems allowed volunteers to be exposed to risks that were only justifiable because individuals consented to these risks. There is no good reason, in principle, why research subjects should

14 This account of the wrongfulness of risk imposition resonates well with contractualist ideas of justifiability. See Scanlon, *What We Owe to Each Other*. Frances Kamm has argued that contractualism is intimately connected with the question of whether our actions wrong individuals. See Kamm, *Intricate Ethics*, 456–68. Oberdiek, who argued in favor of a right against the imposition of risk, refers to contractualism as an answer for how to specify the scope of such a right (*Imposing Risk*, ch. 5).

be treated differently. We should accept a parity between putting consenting individuals at risk outside and inside the research context.<sup>15</sup>

Even when individuals are allowed to waive their right not to be put at risk, that does not imply that all risks are permissible. Critics of research ethics frameworks sometimes suggest that risk-benefit protections for willing and consenting volunteers are motivated by paternalism. They analogize medical research to risky activities like free climbing El Capitan. In the case of free climbing El Capitan, the only plausible justification for restricting autonomous agents from doing so is paternalism.<sup>16</sup>

But medical research is not like free climbing El Capitan. The question is not whether we should prevent individuals from doing something they otherwise would and could do on their own. The question is whether we are permitted to solicit and encourage people to let us do something on them that they otherwise would and could not do on their own. Let me unpack two of these differences.

First, researchers solicit, encourage, and induce volunteers to take part in the study. In studies that cannot be justified under the favorable prospect standard, they need to appeal to the volunteer's altruistic motivations. If researchers ask volunteers to take on additional risks, they need to ensure that they can justify asking for these sacrifices. This means that the risks must be necessary for the proposed research. Recruiting additional volunteers without expecting any scientific benefit could not be justifiable to them. The researchers could not appeal to the altruistic motivations of these subjects. The presence of these additional subjects would not help anyone. Their contributions would be pointless sacrifices. A similar observation holds for cases in which the social benefits are sufficiently trivial that we cannot justify encouraging individuals to take up great risks.

Second, researchers facilitate the risks, and their facilitation is done to serve ends other than those of the risk-taker. Facilitation is different from nonintervention. Anti-paternalism can ask for nonintervention. In the case of El Capitan, that is all that is needed. But in the case of research risks, we

15 See Chappell and Singer, "Pandemic Ethics." Similar comparisons with nonresearch contexts are made by Alex London (see "Reasonable Risks in Clinical Research" and "Clinical Research in a Public Health Crisis"). One reason to treat research risks differently is possible externalities to nonparticipants. The most discussed externality is distrust in vaccines. Vaccine hesitancy arguments raise a variety of complicated empirical and moral questions, so I will largely sidestep these. See, however, the discussion in note 28 below.

16 In the context of COVID-19, see Savulescu and Wilkinson, "Extreme Altruism in a Pandemic." More generally, see Miller and Wertheimer, "Facing Up to Paternalism in Research Ethics"; and Shaw, "The Right to Participate in High-Risk Research."

go beyond nonintervention. Without the research trial, the risk would not exist. The researcher facilitates, in the sense of making possible the risk to the research subject.

This becomes important in cases of excessive heroism. Consider the following case. There is a burning building, perhaps a skyscraper, with very many people inside. There is a small possibility of putting out the fire in the basement and saving these lives if someone runs into the building. The person running into the building would risk almost certain death. No one inside the building can reach the basement. Even if, in this case, we believe it impermissibly paternalistic to prevent a person from running into the building, it is quite another matter for us to facilitate this and give the person the means to do so. Clinical trials with excessive risks do not simply fail to prevent individuals from signing up out of a sense of faint heroism; they actively make this faint heroism possible. The example suggests a kind of proportionality condition that rules out facilitating excessive sacrifices.

The third standard then holds that clinical research is permissible if the additional risks taken up by the participants are neither excessive nor pointless. It rules out extreme acts of altruism and self-sacrifice. I call this the *moderate sacrifice standard*.

To summarize my argument so far, I have argued that morally permissible clinical research must meet one of the three standards I have set out. An important distinction exists between the favorable prospect standard and the other two standards. According to the favorable prospect standard, we can justify research without invoking social benefits. The research is justified the same way as clinical interventions are—purely by reference to the participants' self-interest. The latter two standards—the justifiable risk and moderate sacrifice standards—require, in different ways, social benefits to justify the research.

## 2. ARE CHALLENGE TRIALS EXCESSIVELY RISKY?

As mentioned earlier, challenge trials involve deliberately exposing consenting volunteers to the SARS-COV-2 virus to observe whether the vaccine protects against infection. Importantly, this means that even volunteers in the control arm need to be infected. Opponents of challenge trials believe that the risk is too high. In terms of my framework, these opponents believe that challenge trials do not meet the moderate sacrifice standard and qualify as excessive risks.

Some challenge trials can fend off this challenge. Proposals for human challenge trials typically rely on selecting participants already at low risk from the virus. For young and healthy volunteers, participating is a moderate sacrifice. Proponents of challenge trials have often invoked comparisons with live kidney

donations.<sup>17</sup> The risks involved in kidney donations are clearly proportional to the aim of extending a kidney recipient's life. They also are proportional to the gains that challenge trials could bring.

More difficult are cases in which the mortality risk is high.<sup>18</sup> Consider the risk the virus poses to an octogenarian with multiple preexisting conditions and a weakened immune system. Can this be justified? For this, we need to have a closer look at the benefits of challenge trials. There is a chance that challenge trials will not yield any benefits at all. This might be, first, because the tested vaccine is a dead end. Second, this might be because challenge trials with younger and healthier volunteers would have been similarly informative. Third, field trials might have yielded a similarly fast resolution. Field trials were much faster than proponents of challenge trials feared, and the development of an artificial strain of the virus takes time. The expected value of challenge trials is, then, to some extent, driven by the fact that there is a smaller chance of very large gains. For *if* challenge trials with high-risk participants are not subject to any of the three limitations, then many harms due to the pandemic can be averted.

If the proportionality condition, which determines which risks count as excessive, is read purely in terms of expected value, then this could provide an endorsement even for challenge trials with high-risk participants. But this seems too extreme. Suppose researchers believed that if they experimented on a live lung that is removed from a patient, they might find the resolution to the pandemic immediately. They admit the chance is very, very small, and they admit the patient is almost certain to die. The potential benefits are enormous, so the expected value may appear to be proportional. But we should not succumb to such fanaticism. The moderate sacrifice standard should not be read as simply comparing the prospect of the patient with the expected value. High risks to a patient are excessive if there are only small chances of benefit from the research. This connects with the burning building analogy that I used earlier. What seems objectionable about the example is not the expected value—after all, very many people could be saved. What seems objectionable is the small chance of survival for those entering.

While this argument rejects challenge trials on high-risk patients, the earlier point stands that the risks to healthy and young volunteers are within the margin of moderate sacrifices.

17 Eyal, Lipsitch, and Smith, "Response to Cioffi"; and Jayaram, Sparks, and Callies, "Justifying the Risks of COVID-19 Challenge Trials."

18 This more radical proposal is raised by Savulescu and Wilkinson, "Extreme Altruism in a Pandemic."

## 3. LOW- VERSUS HIGH-DOSAGE CHALLENGES

Things might be different in the next pandemic. In the following, I specify a way for challenge trials to be adopted even if the risk appears initially as too high. Challenge trials are preceded by a dose escalation study that determines how much of a pathogen—what dosage—should be used to infect the participants. A low-dosage challenge trial is a challenge trial that uses a lower dosage than is conventionally used for such challenge trials. For example, consider a trial that uses a dosage corresponding to half of the conventional infection risk. Halving the risk of infection would already reduce the risk of serious harms by half. Without infection, no disease and no harm. But there is a second factor at play. For some diseases, the amount of virus that one is infected with has an impact on the severity of the ensuing disease. There is some evidence that SARS-CoV-2 is among these viruses, although the matter is still subject to scientific dispute.<sup>19</sup> Even if we discount for the provisional nature of this evidence, we should discount the risk by a factor of a little bit more than what is achieved through reductions in the infection risk alone. In the example used, the risk is then reduced by a bit more than half. But the low-dosage challenge could be run at an even lower dosage. In principle, we could reduce the risk as much as is needed to ensure that the moderate sacrifice standard is met.

The low-dosage challenge trial reduces the risk to participants by relying on an exposure that is less likely to infect individual participants. To yield statistically meaningful results, the low-dosage challenge trial needs a larger number of participants. Because a lower proportion of people will be infected in the control arm, researchers need a larger number of people in the control arm (and therefore also in the treatment arm). Nevertheless, this means that each participant faces a lower risk in the trial.

The low-dosage challenge trial should be distinguished from a volunteer lottery in which researchers randomize among volunteers before regular dosage exposure and take the odds prior to randomization as relevant.<sup>20</sup> Such a lottery can trivially reduce risks judged from the standpoint before participant selection. The key difference consists in the way the risk reduction is tied to the exercise of the researcher's agency. In a volunteer lottery, the researcher is performing an equally risky action at the time of exposure. The dice of the previous lottery have already been rolled, and the volunteers who are being exposed

19 For discussion, see Spinelli et al., "Importance of Non-Pharmaceutical Interventions in Lowering the Viral Inoculum"; Trunfio et al., "Lowering SARS-CoV-2 Viral Load Might Affect Transmission but Not Disease Severity in Secondary Cases," as well as Spinelli, Rutherford, and Gandhi, "Authors' Reply."

20 Steel, "Risk Dilution."

receive no safety benefit from the earlier lottery. However, in a low-dosage challenge, researchers perform a less risky action at the time of exposure. The intervention of the researcher is less risky. Exposure is the dice roll, and every volunteer stands a better chance of avoiding harm in the trial.

A volunteer lottery could be designed in order to select and expose at the same time.<sup>21</sup> The trial could be selecting participants at the same time as delivering the vial. But transforming a volunteer lottery into a simultaneous process serves no purpose other than to avoid moral liability. Joining the two processes runs together the risks from infection upon exposure and the risk of being selected for exposure. One risk, which is inherent to the treatment, is joined with another risk that is artificially created by the agent.<sup>22</sup> The low-dosage challenge is different. Exposure to the virus is inherently a chancy process. The only factor explaining the risk in the low-dosage challenge is the exposure to the virus. In other words, in the case of the volunteer lottery, we run a lottery to determine who receives a very risky treatment. By contrast, in a low-dosage challenge, we give a much less risky treatment to more people.

Reflections on the low-dosage challenge thus reveal two points. First, the low-dosage challenge is easier to justify than a regular challenge trial. Second, some form of challenge trial, a suitably low-dosage one, can be justified on grounds of the moderate sacrifice standard.

The risk reduction comes, however, at some price. In order to infer comparably good information from the trial, the number of infections needs to remain more or less constant.<sup>23</sup> With infections being commensurate to a high-dosage challenge, what might matter is the likelihood that there will be harms in the trial cohort. This depends on the exact increase in the number of volunteers and on the extent to which the risk of harm is decreased by a lesser exposure

- 21 I owe this challenge to a reviewer who pressed me to clarify how a low-dosage challenge differs from a volunteer lottery.
- 22 Johann Frick similarly argues for the decomposition test according to which what matters are the odds at each stage of agential intervention. He also highlights that artificially running together different stages by using a surrogate for agential intervention is a way to undermine the test, not to meet it. See Frick, "Contractualism and Social Risk," 210–12. See also Kamm, *Morality, Mortality*, 2:303.
- 23 Does this show that the low-dosage challenge serves no purpose after all? No. Each individual has been subjected to a much lower risk than they would have been in a high-dosage challenge. The comparison between the two is like a scenario in which a harm has to be distributed. A high-dosage challenge concentrates the risk of harm in few individuals; a low-dosage challenge disperses the risk of harm across more individuals. This makes the distribution of risks fairer. See Broome, "Fairness"; and Daniels, "Can There Be Moral Force to Favoring an Identified over a Statistical Life?" It can also be seen as a risky analog to Larry Temkin's "Disperse Additional Burdens View" (*Rethinking the Good*, ch. 3).

dosage. In a pessimistic scenario, a low-dosage challenge does little (if anything) to reduce overall harm. In the next section, I will look at a more extreme version of such a trade-off in which reducing the risk to individuals comes at the cost of increasing the risk that there will be harm in the cohort. If my arguments in favor of a restricted rollout are sound, then they also respond to any concern about the increased cohort size in a low-dosage challenge.

#### 4. FROM SAFETY TESTING TO RESTRICTED ROLLOUT

At whichever level of dosage the challenge trial is performed, I endorsed the (near) consensus that challenge trials should be performed only with low-risk participants. However, if the vaccine is not tested on members of high-risk demographics, we have incomplete information about vaccine safety and need a bridge safety study. The problem is how to generalize from our test population to our target populations. There is no clear rule for how to deal with this generalization. For example, some countries like India insist that trials have to be performed on the local population before being released. For most other countries, the trial data from, for example, Brazil was deemed sufficient. Given that the effects of COVID-19 were quite different in different age groups, it is reasonable that testing on older people was needed to solidify our evidence of vaccine safety. A common and uncontroversial protocol for such bridge safety testing is the following:

*Safety Testing:* A vaccine that has proven to be efficacious in a challenge study will be tested on persons from previously excluded groups under close safety monitoring. Assume testing will include approximately three thousand elderly persons. If the tests are successful, the vaccine will be rolled out universally.

Safety Testing requires the informed consent of all participants. No one doubts that Safety Testing, an established procedure for establishing that vaccines are safe before release, is permissible. But which standard does it meet? This question is important, as we shall see, because it determines whether social benefits are necessary in justifying the procedure. Can Safety Testing be justified on grounds of the favorable prospect standard, or, as in the case of challenge trials, do we need to appeal to social benefits? Safety Testing includes risks of harm to the individuals participating in the study. These are harms caused by either vaccine toxicity, increased exposure to SARS-COV-2, or SARS-COV-2 exposure with the background of a faulty vaccine that enhances disease severity. However, given that the test candidate has already been shown safe and efficacious in human studies, these risks are reduced. Moreover, participants would face increased risks of a SARS-COV-2 infection in the absence of the test vaccine. The individual

benefits of the test vaccine are the possibility of longer protection from SARS-CoV-2. One might, therefore, believe that Safety Testing meets the favorable prospect standard and is in the rational self-interest of elderly volunteers.

However, this judgment is disputable. The risks are still partially uncertain, and the vaccine is still experimental. Safety information has not been gathered yet for older persons. Given these concerns about the still experimental vaccine, I will proceed on the assumption that the justification for Safety Testing needs to appeal to the social benefits to nonparticipants. In other words, Safety Testing meets the moderate sacrifice standard or the justifiable risk standard. Later, in section 5, I will explain how my argument changes if we do not need to appeal to social benefits to justify Safety Testing because Safety Testing meets the favorable prospect standard. The social benefits at stake include the eventual protection of large populations from the virus. The earlier the vaccine is ready and can reduce transmission rates, the greater the social benefits from the trial. The social benefits of the test vaccine, if successful and properly distributed, thus include thousands of saved lives.

A controversial alternative to Safety Testing that would cut time in the release of the vaccine is the following protocol:

*Restricted Rollout:* A vaccine that has proven to be efficacious in a challenge study will be released to a restricted, yet large group of consenting persons under conditions of registration and close monitoring. The restricted rollout includes previously excluded groups. Assume a rollout to one million persons with approximately three hundred thousand elderly persons. If the monitoring is successful in that large group, the vaccine will be rolled out universally.

Restricted Rollout would make the vaccine available to a large population by declaring the vaccine “conditionally approved.” Comparisons can be drawn with data from population-wide health care providers like the National Health Service or via samples from the nonvaccinated population.<sup>24</sup>

I now turn to my argument for the moral equivalence between Safety Testing and Restricted Rollout. In terms of the risks and benefits to individual participants, Restricted Rollout imposes the same or almost the same individual risks of harm on participants as Safety Testing. The vaccine itself is neither more nor less dangerous to individuals in either of the two protocols. The only relevant risk factors that may change are that Safety Testing provides for a better opportunity to teach participants about minimizing risks and that Safety Testing provides for

24 Eyal, Gerhard, and Strom, “Strengthening and Accelerating SARS-COV-2 Vaccine Safety Surveillance,” 3456.

better monitoring and timely detection of adverse effects of the vaccine. Monitoring means that adverse effects are more likely to be detected in participants, and possible interventions can be taken. These interventions can range from the suspension of follow-up vaccine shots to hospitalization. This reduces the risks to participants. Close monitoring is easily feasible in Safety Testing, given the small number of individuals involved. Operationally, monitoring the larger participant group is more difficult in Restricted Rollout than in Safety Testing.

However, Restricted Rollout fares *better* than Safety Testing in another respect that is relevant for the timely detection of adverse effects. Because of its larger size, Restricted Rollout generates better safety information than Safety Testing, and that information can be used to intervene in the trial when necessary. The larger trial size allows researchers to detect rare vaccine side effects. Writing about drug safety, Brian Strom points out that traditional drug safety protocols typically do not detect adverse effects that occur in frequencies of 1 in 1,000. Even larger trials like Safety Testing would struggle to detect adverse outcomes that have frequencies of less than 1 in 3,333.<sup>25</sup> Such adverse side effects are not uncommon for drugs or vaccines. For comparison, the high-profile example of blood clots following coronavirus vaccinations that have paused vaccinations in various countries has been estimated at the time at a frequency of 1 in 100,000.<sup>26</sup>

So, while Restricted Rollout has less individual monitoring, it also has the capacity to detect a greater variety of adverse effects that can allow researchers to intervene and minimize the risk to participants. On balance, it is therefore not clear that Restricted Rollout produces greater risks to participants than Safety Testing. A reasonable estimate on which I will proceed is that the two ways in which Safety Testing and Restricted Rollout reduce risks even out. It is straightforward that Restricted Rollout also provides the same individual benefits to each participant as Safety Testing does. In the case of an efficacious and safe vaccine, both protocols offer earlier added protection from SARS-COV-2 to each participant.

The two protocols differ, however, in the social benefits to nonparticipants as well as in the number of participants. The social benefits are larger in Restricted Rollout for two reasons. The first is related to the earlier point about the increased power to detect rare adverse effects. The vaccine, when tested in Restricted Rollout, will, therefore, be safer upon eventual release than a vaccine tested by means of Safety Testing. In addition, just as researchers are better equipped to observe

25 Strom, "How the US Drug Safety System Should Be Changed," 2072. The second calculations are based on Eyal, Gerhard, and Strom, "Strengthening and Accelerating SARS-COV-2 Vaccine Safety Surveillance," 3455.

26 Cines and Bussel, "SARS-COV-2 Vaccine-Induced Immune Thrombotic Thrombocytopenia," 2255.

rare side effects, they are better equipped to gain information about the effects of the vaccine on more fine-grained demographics. The increased abilities for such fine-grained observations can make the vaccine safer in the long run by providing more detailed safety information. In addition, Restricted Rollout produces larger social benefits by cutting short the pandemic and preventing loss of life and misery from the direct and indirect effects of the pandemic.

In Safety Testing, the benefits are jointly produced by the efforts of a smaller group. In Restricted Rollout, larger benefits are jointly produced by the efforts of a larger group. The right comparison for the benefits to nonparticipants is the *marginal* social benefit—that is, the benefits that are possible due to increased participation in the testing process. This is in line with what the moderate sacrifice standard demands. To assess whether we can justify exposing additional volunteers to the risk, we need to know whether including more volunteers will also lead to sufficient benefits or whether their contribution would be superfluous or an instance of faint heroism. We could not appeal to altruistic reasons to justify the inclusion of additional test candidates if we could achieve the same (or almost the same) altruistic benefit without these test candidates. A similar argument holds for the justifiable risk standard. An individual could raise a complaint against a risk imposition if that risk imposition does not produce enough marginal benefits to nonparticipants. In such a case, the complaint against the additional risk imposition could outweigh any possible complaint against the withholding of benefits to nonparticipants. In other words, individuals may be required to bear the burden of a risk imposition for the sake of greater benefits to nonparticipants, but without such benefits to nonparticipants, there is no consideration that justifies the risk imposition.

Even if we focus only on the marginal social benefits of the wider release in Restricted Rollout as opposed to Safety Testing, the marginal social benefits will be very substantial in Restricted Rollout. The increased size of the testing population produces, if the vaccine is safe and efficacious, greater benefits in terms of earlier reduction of transmission rates and delivers more fine-grained information about adverse effects from vaccine usage. We need additional participants to produce these benefits. The protocol of Restricted Rollout can make a large impact on transmission rates and generate better safety information only because it releases the vaccine to a large group. Even though estimating the social benefits is difficult, it is reasonable to assume that the benefits produced by the added test population meet the moderate sacrifice standard or perhaps the justifiable risk standard.

The real difference between the two protocols is then neither the individual effects on participants nor the marginal benefits to nonparticipants. The real difference is simply the scale of the risk imposition (and, thus, the scale of *total*

social benefits to nonparticipants). Safety Testing subjects a smaller number to the risk; Restricted Rollout subjects a larger number to the risk. Should this difference matter?

I believe it should not. We can give the following argument why the scale of the risk imposition does not matter in itself. The risk imposition on one group of three thousand volunteers is justifiable to each of them. The risks are outweighed by benefits to themselves and by benefits to nonparticipants. Either this means that the risk imposition itself is not morally problematic and justifiable to them, or this means that we can permissibly appeal to their altruistic motivations. In either case, no one person in the group of three thousand volunteers would be wronged by the risk imposition. Now, take a second group of three thousand volunteers. Here, too, the risk imposition is justifiable to them either in virtue of their self-interest or our appeal to their altruistic motivations. The risks are independent; testing both the first and second groups creates no adverse effect on any one person. The argument repeats until we reach all three hundred thousand volunteers.

If the risk imposition to the first three thousand volunteers was justifiable to each one of them and wronged none of them, then it has to be justifiable to all other persons who are affected in the very same way. If no one would be wronged if the risk was only imposed on their group of three thousand, then who is wronged in the larger group? Individual objections to the risk imposition cannot depend on the fact that the decision-maker is doing something to *other* people, which is perfectly permissible.

This argument can be generalized. In its essence, it holds that the permissibility of risks in clinical research is invariant to scale. As long as scaling up produces the same individual risks, individual benefits, and marginal social benefits, it is permissible to perform the research on the larger group as well. I will call this the *scale invariance argument*. The scale invariance argument can also explain why the cohort effect for low-dosage challenges is not problematic. Scale invariance means that the number of participants can be increased as long as the marginal social benefits are sufficiently high. This is the case for a low-dosage challenge trial.

##### 5. RISKS AND BENEFITS TO PARTICIPANTS AND NONPARTICIPANTS

It is helpful to compare my argument to a similar argument made by Johann Frick, among others.<sup>27</sup> According to this argument, risks that are in an

27 Frick, "Contractualism and Social Risk," 186–88; Dougherty, "Aggregation, Beneficence, and Chance," 8–11; Hare, "Should We Wish Well to All?" 455–67.

individual's self-interest can permissibly be scaled up. If the risk is in the rational self-interest of various persons and it would be permissible to impose the risk on each person taken separately, then it should also be permissible to impose the risk on all persons taken together. My scale invariance argument is structurally similar but differs insofar as it does not require that the risks are in the rational self-interest of each individual. Instead, my argument holds that if the individual risk is justifiable because of considerations of self-interest *and* marginal social benefits taken together, then it is justifiable to impose the risks all at once. The difference between the two arguments is important for two reasons. First, it more satisfactorily explains why we permit the risks of experimental vaccines in controlled testing environments. The fact that we regard the vaccines as experimental indicates that we are not convinced there is a large-scale self-interested argument in favor of the vaccines. Second, and relatedly, the combination of self-interest and marginal social benefit can explain why my argument need not necessarily imply an even more radical option that dispenses with the post-challenge safety study.

*Unrestricted Rollout:* A vaccine that has proven to be efficacious in a challenge study will be released to any person who wishes to be vaccinated.

Restricted Rollout already achieves very large social benefits in terms of shortening the COVID-19 pandemic by several months. The proposed protocol contemplates a universal rollout once short-term outcomes have been analyzed. Unrestricted Rollout, or skipping the safety bridge study, makes safety monitoring very difficult. This could have additional benefits if everything goes well, but it also comes with corresponding risks. This shows that my revised argument that relies on marginal social benefits is sensitive to the scale of the risk imposition in one sense. The argument is sensitive to considerations regarding the necessity of imposing these risks. Scaling up the risk does not guarantee that social benefits will be scaled up at the same rate. Only when scaling up the risk means that all relevant factors can be scaled up is it permissible to proceed with the risk imposition for the larger group.

What if, contrary to my assumption so far, such benefits are not, in fact, necessary? Perhaps it is the case that the individual benefits outweigh the individual risks. In future pandemics, there might be some vaccines or medicines for which this is the case. These trials would meet what I described as the favorable prospect standard. Does my scale invariance argument, together with the assumption that the favorable prospect standard is met, entail that Unrestricted Rollout is morally permissible (or even required) for such trials?

Meeting the favorable prospect standard means that taking the vaccine is in the rational self-interest of those who wish to take it. Unrestricted Rollout

means that everyone for whom this is the case is allowed to access the vaccine. Therefore, provided the favorable prospect standard is met, any objection to Unrestricted Rollout cannot rest on paternalism. Any objection to Unrestricted Rollout would have to be based on effects on nonparticipants. There are two such pertinent considerations. Both of these are empirical, and their strength will depend on the details of the vaccine in question.

The first consideration is possible effects that such a policy can have on public trust in vaccines. Dispensing with a safety bridge study, as Unrestricted Rollout effectively does, deviates markedly from our ordinary process of vaccine certification. It is possible that this is acceptable to the public, given the unusual conditions of a pandemic. But there is also a danger that this undermines trust in vaccine and drug certification. If the latter is the case, then Unrestricted Rollout would be causing long-term harm for short-term gains.<sup>28</sup> The second effect on nonparticipants is the possible risk of increased pathogen exposure. Some vaccine candidates have the reverse effect of increasing exposure to the pathogen. The protocol of Restricted Rollout registers participants and monitors. It also makes it easier to isolate those participating in the rollout as much as possible from the rest of society. Unrestricted Rollout does not and can, therefore, create additional risks for nonparticipants due to an increased spread of pathogens.

Again, whether these effects are actually present in the case of any given candidate vaccine depends on circumstances and difficult empirical questions. (The same holds, as I outlined earlier, for similar concerns about Restricted Rollout.) For some candidate vaccines, these considerations will be weighty enough. For others, they will not. If neither of these adverse effects is weighty enough and if the vaccine meets the favorable prospect standard, then my argument entails that the more radical option of Unrestricted Rollout is permissible. But in such a case, it is also hard to see what would be wrong with this implication. This would be a vaccine that is in the rational self-interest of many persons and does not create negative externalities for nonparticipants. What possible reason could we have for depriving some individuals of taking a medical intervention that is in their best interest without harming third parties?

28 Richard Yetter Chappell doubts this argument on grounds that trust considerations would mean refraining from aiding innocent people now for the sake of protecting others—namely, those who distrust vaccines—from self-inflicted harm (“Pandemic Ethics and Status Quo Risk,” 69–70). However, as I put it in the main text, the crux of the argument is that decline in trust in vaccines has long-term consequences. Trust in vaccines often stems from a trust in regulatory mechanisms and institutions. If trust in vaccines and the medical establishment generally declines, then everyone loses out because infectious diseases can spread more easily.

## 6. SCALE INVARIANCE AND COHORT RISKS

One further aspect changes with the scale of the risk imposition. It becomes more likely that there will be harm *ex post* the risk imposition in the proposed protocol. Although the risk to each individual is the same whether she is in the smaller or larger protocol, in the smaller protocol, there is a lower likelihood that a participant will be harmed. Some critics of challenge studies appear to be concerned mainly with the risk to the trial cohort.<sup>29</sup> Does this constitute an objection to my argument that the permissibility of risk impositions should be invariant to mere scaling up?

A critic might argue that my argument has only established that no individual would be wronged by either the low-dosage challenge or Restricted Rollout. But this critic would go on to argue that whether an action wrongs any one individual is not sufficient to establish that the action is not wrong. Actions can be wrong without wronging any single individual. One way to spell this out is by embracing pluralism about moral rightness. The argument that I have given so far captures one important wrong-making feature of an act. My argument has shown that this wrong-making feature is not present in the cases I discussed. But loss of aggregate well-being could be another wrong-making feature of an act. The low-dosage challenge or Restricted Rollout might be wrong for this reason. Promoting aggregate well-being, under this understanding, is a *pro tanto* reason in favor of an action.<sup>30</sup>

A second way to spell this out gives a less prominent role to aggregative and impersonal considerations. According to this way, in almost all cases of interpersonal morality, the question whether or not an action is justifiable to each person determines the moral permissibility of the action. Interpersonal morality can be defined as governing those cases in which only the effects on persons are morally relevant. Only in some cases of interpersonal morality can this be overridden by exceptional circumstances. A great loss of life could be such an exceptional circumstance.<sup>31</sup> This objection would most naturally focus on the fact that the risks in the low-dosage challenge and Restricted Rollout appear to be positively correlated. In the unlikely scenario of great toxicity, this

29 See Corey et al., “A Strategic Approach to COVID-19 Vaccine R&D”; Deming et al., “Accelerating Development of SARS-COV-2 Vaccines.”

30 This way of spelling out pluralism is broadly in line with Johann Frick’s pluralism about rightness. Frick does not use the language of “wronging.” However, he makes clear that both wrong-making features are parts of “interpersonal morality,” which deals with our duties to other persons; see Frick, “Contractualism and Social Risk,” 218–23.

31 This is the view that I tentatively favor. For an excellent discussion of the tension between personal and impersonal considerations, see Nagel, *Mortal Questions*, ch. 5.

would affect a great number of individuals at once. The worst-case scenario is worse if we adopt these protocols.

But to move from the fact that in these protocols there is a higher likelihood that volunteers will be harmed to the conclusion that these protocols are wrong (even if not wronging any one) is too quick. The reason is that both the low-dosage challenge and Restricted Rollout save the lives of many nonparticipants compared to the slower, established protocols. The quicker release of the vaccine means that the COVID-19 pandemic will be shortened, and many lives will be saved. The net effect is going to be one of more statistical lives saved rather than lost. The moral catastrophe of a large number of persons dying is already happening in an ongoing pandemic, putting concerns about prioritizing the worst-case scenario in perspective.

The objection to the proposed protocols would have to be that it is more likely that lives of *participants* will be lost. It is not unusual in research ethics to be especially concerned with the risk to trial participants, largely because this risk is actively caused by the researchers. This is particularly evident in the case of challenge trials in which researchers deliberately infect, but this is also the case in field trials in which there is the risk that the vaccine administered by researchers enhances the severity of the existing disease.

There are indeed good grounds for special concern with research subjects. The most natural concern is that the risk imposition wrongs an individual or violates their rights. If we can save a larger number only by violating the rights of a smaller number of people, then we may not do so. This explains why active harm caused by researchers is prioritized heavily over harm researchers passively allow. But my whole argument rejects the view that any single person is wronged. I have not made the simple consequentialist argument that challenge trials avert more harm than they cause. Rather, I argued that the risk impositions inherent in the low-dosage challenge and the Restricted Rollout do not wrong any individual, nor do they violate any of their rights. The objection currently under consideration is different. It relies on the idea that individualized and interpersonal morality does not capture everything of relevance. The objection pushes us to consider collectivized and impersonal morality. Our concern may be, for example, the loss of aggregate well-being. But if *this* is our concern, then we do not have any good reason to ignore the effects on nonparticipants. There cannot be an objection that, for example, Restricted Rollout compromises aggregate well-being when it, in fact, saves more lives than Safety Testing would.

The central point of my argument is not limited to the COVID-19 pandemic. It also applies to other health emergencies or future pandemics in which faster testing protocols would avert great harms to public health. There are ways to

avert such harm and save many lives without compromising the value of each individual or sacrificing some for the sake of the greater good. It is one of the cases in which deontological and consequentialist considerations do not pull in opposite directions.<sup>32</sup>

Ashoka University  
 bastian.steuwer@ashoka.edu.in

## REFERENCES

- Broome, John. "Fairness." *Proceedings of the Aristotelian Society* 91, no. 1 (June 1991): 87–101.
- Chappell, Richard Yetter. "Pandemic Ethics and Status Quo Risk." *Public Health Ethics* 15, no. 1 (April 2022): 64–73.
- Chappell, Richard Yetter, and Peter Singer. "Pandemic Ethics: The Case for Risky Research." *Research Ethics* 16, nos. 3–4 (July–October 2020): 1–8.
- Cines, Douglas B., and James B. Bussel. "SARS-COV-2 Vaccine-Induced Immune Thrombotic Thrombocytopenia." *New England Journal of Medicine* 384, no. 23 (June 2021): 2254–56.
- Corey, Lawrence, John R. Mascola, Anthony S. Fauci, and Francis S. Collins. "A Strategic Approach to COVID-19 Vaccine R&D." *Science* 368, no. 6494 (May 2020): 948–50.
- Daniels, Norman. "Can There Be Moral Force to Favoring an Identified over a Statistical Life?" In *Identified versus Statistical Lives: An Interdisciplinary Perspective*, edited by I. Glenn Cohen, Norman Daniels, and Nir Eyal, 110–23. Oxford: Oxford University Press, 2015.
- Deming, Meagan E., Nelson L. Michael, Merlin Robb, Myron S. Cohen, and Kathleen M. Neuzil. "Accelerating Development of SARS-COV-2 Vaccines—The Role for Controlled Human Infection Models." *New England Journal of Medicine* 383, no. 10 (September 2020): e63.
- Dougherty, Tom. "Aggregation, Beneficence, and Chance." *Journal of Ethics and Social Philosophy* 7, no. 2 (May 2013): 1–19.
- Emanuel, Ezekiel J., David Wendler, and Christine Grady. "What Makes Clinical Research Ethical?" *Journal of the American Medical Association* 283, no. 20 (May 2000): 2701–11.

32 Earlier versions of this paper were presented at the Virtual Ethics Work-in-Progress Group and seminars at Rutgers University and Princeton University. I am grateful to the audiences for helpful comments and criticism. I am also grateful to Nir Eyal, Todd Karhu, Kai Spiekermann, as well as multiple anonymous reviewers for their thoughts and suggestions.

- Eyal, Nir. "Is There an Ethical Upper Limit on Risks to Study Participants?" *Public Health Ethics* 13, no. 2 (July 2020): 143–56.
- Eyal, Nir, Tobias Gerhard, and Brian L. Strom. "Strengthening and Accelerating SARS-COV-2 Vaccine Safety Surveillance through Registered Pre-approval Rollout after Challenge Tests." *Vaccine* 39, no. 26 (June 2021): 3455–58.
- Eyal, Nir, Marc Lipsitch, and Peter G. Smith. "Human Challenge Studies to Accelerate Coronavirus Vaccine Licensure." *Journal of Infectious Diseases* 221, no. 11 (May 2020): 1752–56.
- . "Response to Cioffi." *Journal of Infectious Diseases* 222, no. 1 (July 2020): 169–70.
- Frick, Johann. "Contractualism and Social Risk." *Philosophy and Public Affairs* 43, no. 3 (Summer 2015): 175–223.
- Hare, Caspar. "Should We Wish Well to All?" *Philosophical Review* 125, no. 4 (October 2016): 451–72.
- Jayaram, Athmeya, Jacob Sparks, and Daniel Callies. "Justifying the Risks of COVID-19 Challenge Trials: The Analogy with Organ Donation." *Bioethics* 36, no. 1 (January 2022): 100–106.
- Kahn, Jeffrey P., Leslie Meltzer Henry, Anna C. Mastroianni, Wilbur H. Chen, and Ruth Macklin. "For Now, It's Unethical to Use Human Challenge Studies for SARS-COV-2 Vaccine Development." *Proceedings of the National Academy of Sciences* 117, no. 46 (October 2020): 28538–42.
- Kamm, F. M. *Intricate Ethics: Rights, Responsibilities, and Permissible Harm*. Oxford: Oxford University Press, 2007.
- . *Morality, Mortality*. Vol. 2, *Rights, Duties, and Status*. New York: Oxford University Press, 1996.
- London, Alex John. "Clinical Research in a Public Health Crisis: The Integrative Approach to Managing Uncertainty and Mitigating Conflict." *Seton Hall Law Review* 39, no. 4 (2009): 1173–202.
- . "Reasonable Risks in Clinical Research: A Critique and a Proposal for the Integrative Approach." *Statistics in Medicine* 25, no. 17 (September 2006): 2869–85.
- Miller, Franklin G., and Alan Wertheimer. "Facing Up to Paternalism in Research Ethics." *Hastings Center Report* 37, no. 3 (May–June 2007): 24–34.
- Murphy, Liam. "Nonlegislative Justification." In *Principles and Persons: The Legacy of Derek Parfit*, edited by Jeff McMahan, Tim Campbell, James Goodrich, and Ketan Ramakrishnan, 247–74. Oxford: Oxford University Press, 2021.
- Nagel, Thomas. *Mortal Questions*. Cambridge: Cambridge University Press, 1979.
- Oberdiek, John. *Imposing Risk: A Normative Framework*. Oxford: Oxford

- University Press, 2017.
- Pannu, Jaspreet, Megan J. Palmer, Anita Cicero, David A. Relman, Marc Lipsitch, and Tom Inglesby. "Strengthen Oversight of Risky Research on Pathogens." *Science* 378, no. 6625 (December 2022): 1170–72.
- Perry, Stephen R. "Harm, History, and Counterfactuals." *San Diego Law Review* 40, no. 4 (2003): 1283–313.
- Plotkin, Stanley A., and Arthur Caplan. "Extraordinary Diseases Require Extraordinary Solutions." *Vaccine* 38, no. 24 (May 2020): 3987–88.
- Rid, Annette, and David Wendler. "A Framework for Risk-Benefit Evaluations in Biomedical Research." *Kennedy Institute of Ethics Journal* 21, no. 2 (June 2011): 141–79.
- Savulescu, Julian, and Dominic Wilkinson. "Extreme Altruism in a Pandemic." *Journal of Medical Ethics Blog*, April 23, 2020. <https://blogs.bmj.com/medical-ethics/2020/04/23/extreme-altruism-in-a-pandemic/>.
- Scanlon, T. M. "Contractualism and Justification." In *Reasons, Justification, and Contractualism: Themes from Scanlon*, edited by Markus Stepanians and Michael Frauchinger, 17–43. Berlin: De Gruyter, 2021.
- . *What We Owe to Each Other*. Cambridge, MA: Belknap Press, 1998.
- Shah, Seema K., Franklin G. Miller, Thomas C. Darton, Devan Duenas, Claudia Emerson, Holly Fernandez Lynch, Euzebiusz Jamrozik, et al. "Ethics of Controlled Human Infection to Address COVID-19." *Science* 368, no. 6493 (May 2020): 832–34.
- Shaw, David. "The Right to Participate in High-Risk Research." *Lancet* 383, no. 9921 (March 2014): 1009–11.
- Spinelli, Matthew A., David V. Glidden, Efsthathios D. Gennatas, Michel Bielecki, Chris Beyrer, George Rutherford, Henry Chambers, et al. "Importance of Non-Pharmaceutical Interventions in Lowering the Viral Inoculum to Reduce Susceptibility to Infection by SARS-COV-2 and Potentially Disease Severity." *Lancet Infectious Diseases* 21, no. 9 (September 2021): e296–301.
- Spinelli, Matthew A., George Rutherford, and Monica Gandhi. "Authors' Reply." *Lancet Infectious Diseases* 21, no. 7 (July 2021): 915–16.
- Steel, Robert. "Risk Dilution: Or, How to Run a Minimal-Risk HIV Challenge Trial." *Journal of Applied Philosophy* 40, no. 1 (February 2023): 133–49.
- Steuwer, Bastian, Euzebiusz Jamrozik, and Nir Eyal. "Prioritizing Second-Generation SARS-COV-2 Vaccines through Low-Dosage Challenge Studies." *International Journal of Infectious Diseases* 105 (April 2021): 307–11.
- Strom, Brian L. "How the US Drug Safety System Should Be Changed." *Journal of the American Medical Association* 295, no. 17 (May 2006): 2072–75.
- Temkin, Larry S. *Rethinking the Good: Moral Ideals and the Nature of Practical Reasoning*. Oxford: Oxford University Press, 2012.

- Trunfio, Mattia, Andrea Calcagno, Stefano Bonora, and Giovanni Di Perri. "Lowering SARS-COV-2 Viral Load Might Affect Transmission but Not Disease Severity in Secondary Cases." *Lancet Infectious Diseases* 21, no. 7 (July 2021): 914–15.
- Weijer, Charles. "The Ethical Analysis of Risk." *Journal of Law, Medicine and Ethics* 28, no. 4 (Winter 2000): 344–61.
- Wikler, Daniel. "Must Research Benefit Human Subjects If It Is to Be Permissible?" *Journal of Medical Ethics* 43, no. 2 (February 2017): 114–17.

# THE PERSON AS ENVIRONMENTALLY INTEGRATED DEMENTIA, LOSS, AND EXTENDED COGNITION

*Matilda Carter*

How could I possibly be disintegrating as a human being, solely due to the slow deterioration of my brain?

Christine Bryden, *Will I Still Be Me?*

IN A 2022 REPORT, the World Health Organization described addressing dementia as “one of the greatest health challenges of our generation.”<sup>1</sup> This phrase neatly captures the terms in which public discussion about the condition tends to proceed: dementia is a health issue, and as with all health issues, we should primarily be concerned about prevention, early detection, and effective treatment. While these are certainly urgent demands, there are also socio-political dimensions to this issue that ought not to be neglected—namely the ways in which institutions and individuals treat people living with dementia.

For dementia self-advocate Christine Bryden, key among these concerns is the dominant narrative of dementia as a process that irreversibly sets those who live with it on a path to the destruction of their personal identities and of their personhood.<sup>2</sup> When presented with her diagnosis, she felt intense personal anguish that she attributes to this view, describing an “overwhelming fear of future non-being.”<sup>3</sup> Reflecting also on the experiences of others, she rejects what can be termed the *loss narrative* as both stigmatizing and oppressive.<sup>4</sup> The primary aim of this paper is to validate and philosophically bolster these claims.

In section 1, I highlight three widely disseminated distortions in the public understanding of dementia that reflect an implicit acceptance of the idea that it is a condition fundamentally characterized by loss such that those who live with it will inevitably lose their personal identities and their personhood. I then argue that this idea acts as a legitimating ideology for the stigma that people

1 World Health Organization, *A Blueprint for Dementia Research*, v.

2 Bryden, *Will I Still Be Me?* 16.

3 Bryden, *Will I Still Be Me?* 39.

4 Bryden, *Will I Still Be Me?* 62, 120.

living with dementia face, while contributing to their oppression by marginalization and cultural imperialism.

Personal identity and personhood, however, are distinct concepts, the losses of which have distinct implications. Even if it were true that people living with dementia lost their personal identities throughout the progression of the condition, such that they became metaphysically different people than they were at its onset, they would still have strong claims to be free from stigma and oppression. As I demonstrate in section 2, on the other hand, losing their personhood would remove them from the scope of justice altogether, such that any harmful effects engendered by social arrangements would only press on our relative moral concern. Challenging the loss narrative must begin, then, with a defense of the personhood of people living with dementia.

With this aim in mind, in section 3 I develop and defend an account of the person as environmentally integrated. Making novel links between feminist care ethics and the extended mind thesis, this account conceives of personhood as a relational attribute that is held by all those who share an environment of cognitive extension. As the progression of dementia does not threaten this status, all people living with dementia are persons and thus entitled to liberation from the social injustices entailed by the loss narrative.

#### 1. STIGMA, OPPRESSION, AND THE LOSS NARRATIVE

Whether they are living with Alzheimer's disease, vascular dementia, frontotemporal dementia, or any of the other conditions that fall under the umbrella term "dementia," all who live with this condition experience a progressive deterioration in cognitive function.<sup>5</sup> As most of us value our cognitive function, alongside the capacities that depend on it, it is reasonable to assume that most of us would experience the development of dementia as personally costly. It would therefore seem philosophically irresponsible to deny any connection between dementia and loss.

It is one thing, however, to note that dementia involves the deterioration of cognitive function and quite another to claim that these losses, in Dan Brock's terms, "ultimately destroy personal identity and personhood in the patient."<sup>6</sup> This is the essence of the loss narrative, denounced by Bryden as stigmatizing and oppressive: it depicts dementia as a condition that withers away at fundamental features of who we are, such that those who develop it are irreversibly set on a path toward becoming indistinct human objects. It thus renders a life

5 World Health Organization, "Dementia."

6 Brock, "Justice and the Severely Demented Elderly," 73.

lived with dementia one that is *fundamentally* characterized by loss—not just of specific capacities but of personal identity and personhood.

In this section, I bolster the political case developed in Bryden's self-advocacy work against the loss narrative. I begin by outlining three key distortions in public understanding around dementia and the lives of those who live with it, each of which reflects the firm grip the loss narrative holds on public imagination. I then demonstrate the contribution of these distortions to a global stigma about dementia, arguing that the loss narrative acts as a legitimating ideology for the attitudes that underpin it. I then link these stigmatizing attitudes to structural injustices faced by people living with dementia worldwide, concluding that the loss narrative, both directly and indirectly, contributes to oppression.

### 1.1. Three Distortions

It is a well-established methodological norm within egalitarian political philosophy, particularly when dealing with questions concerning the position of social groups of which the speaker is not a member, to adopt some of the tenets of standpoint epistemology. Even if they do not sign up to the wider framework of understanding all knowledge claims as socially situated, it seems to me uncontroversial to think that egalitarians at the very least ought to accept that persons who experience structural injustice are likely to have insights that outsiders do not.<sup>7</sup> In this light, the mere fact that it comes from Bryden, who identifies the public attitudes it engenders as an equal contributor to her “constant struggle” with dementia as the condition's symptoms themselves, gives us reason to take seriously the claim that the loss narrative is socially dominant.<sup>8</sup>

To leave the discussion here, however, would be to risk circularity. The aim of this section is after all to demonstrate that people living with dementia *are* a group who experience structural injustices (in part) *because* of the loss narrative. To begin such an argument from the claim that we should view the insights of dementia self-advocates as privileged due to their experience of structural injustice—a claim that is itself under contention—would be to veer dangerously close to begging the question. In order to bolster Bryden's argument, then, it is necessary to provide some evidence of the loss narrative in action. While to my knowledge no reputable studies on public adherence to this narrative exist, there is ample evidence of widespread distortions in understanding that reflect an underlying if often only implicit commitment to it. Here I explore three such distortions, providing sufficient evidence to begin analyzing its political effects.

7 This is a norm that can be traced back to Alcoff, “The Problem of Speaking for Others,” 5–32.

8 Bryden, *Will I Still Be Me?* 62.

The first of these distortions consists in a widespread negativity bias toward dementia. Both researchers and the public, as Bryden notes, overwhelmingly focus on what is lost during the progression of dementia, with little serious discussion of what might be gained.<sup>9</sup> In research, this manifests through a historical focus on biomedical analyses of capacity loss at the expense of qualitative research on the experience of living with dementia.<sup>10</sup> Despite some conscious movement toward the language of “living well” with dementia in public policy circles, alongside growing calls for the greater inclusion of those who live with the condition in research, the dominance of deficit-focused research remains for the most part intact.<sup>11</sup> Likewise, studies on attitudes toward lives lived with dementia consistently report pervasive negative characterizations, particularly among those who lack knowledge about the condition.<sup>12</sup>

To be clear, the mere observation of negative elements of living with dementia is not what is at issue. Rather, what makes this phenomenon a distortion in understanding is the excessive focus on these aspects of the condition, reflecting the idea that the condition is fundamentally characterized by inevitable and eventually total loss. Under such a paradigm, there is no need to pay any serious attention to the benefits or improved capacities that a person may accrue throughout the progression of dementia; indeed, these are rendered trivial if not definitionally impossible. There have been no extensive studies, for instance, into the extent to which people living with dementia, in tandem with a decline in memory, experience an increase in what Bryden calls a “sense of the present time, the sense of ‘now,’ of how to live each moment and treasure it as if it were the only experience to look and wonder at.”<sup>13</sup> If the growing popularity of meditation and mindfulness practices is anything to go by, this is something that many people value and strive for, but it is rarely if ever thought of or publicly presented as a potential benefit of dementia in the present context, which is typified by widespread negativity bias.

Even where changes appear to have made a person living with dementia happier, moreover, they are often interpreted through a second distortion in understanding: the denial of authenticity. Where this occurs, the person before onset is presented as the authentic self, with subsequent changes in values, preferences, and personality traits viewed as suspect, potentially inauthentic

9 Bryden, *Will I Still Be Me?* 47–48.

10 Dewing, “Participatory Research,” 11–12.

11 Webb, Williams, Gall, and Dowling, “Misfitting the Research Process.”

12 Chang and Hsu, “Relationship between Knowledge and Types of Attitudes towards People Living with Dementia,” 3777.

13 Bryden, *Dancing with Dementia*, 11.

manifestations of the underlying condition. In other words, such changes are understood as *losses* of aspects of the authentic self as opposed to the kind of changes in our identities that we all make throughout our lives.

This distortion conflicts with testimony from those who live with dementia about their sense of identity. A majority of respondents in a 2011 study of UK adults with early-stage dementia, for example, reported little change in their identities as a whole, despite changes to their personalities that had occurred since the development of the condition.<sup>14</sup> This is corroborated by Bryden, who, despite the advanced state of her condition, reports experiencing a continuous sense of self.<sup>15</sup> However, as is apparent in survey data, the denial of authenticity is a view that has been widely disseminated. In a 2011 study on adults in Northern Ireland, for instance, 75 percent of respondents agreed with the statement “once they have dementia the person you knew eventually disappears.”<sup>16</sup> Likewise, many participants in a 2019 study of caregivers in the United States described dementia as a shameful condition, referring to those who live with it with phrases like “a shell of themselves,” “losing control,” “becoming like a child,” and “losing their mindset.”<sup>17</sup>

As with the first distortion, it is important to clarify the boundaries of the concern I am raising. Whether or not a person living with dementia is metaphysically the same person as they were at onset is not at issue here. While there is a lot at stake in the answer to that question, including the moral and legal force of advance decisions to refuse treatment and the norms that should govern the permissible continuance of intimate relationships they had before its onset, it does not bear directly on the question of *authenticity of changes* nor on the question of whether a person living with advanced dementia has a personal identity at all.<sup>18</sup> A person, after all, could metaphysically be a different person than the one they were at onset yet nevertheless still possess a personal identity.

The core of this second distortion, rather, is the imposition of a hierarchical relationship between the person at onset and the person throughout the course of dementia, whereby the former is prized as the authentic self, such that the latter’s differences are understood in terms of loss rather than change. Embracing this distortion reflects an implicit commitment to the idea of destruction

14 Caddell and Clare, “I’m Still the Same Person,” 379–98.

15 Bryden, *Will I Still Be Me?* 121.

16 McManus and Devine, *Dementia*.

17 Lopez, Rose, Kenney, Sanborn, and Davis, “Managing Shame,” 183.

18 I have discussed the moral and legal force of advance decisions to refuse treatment elsewhere. See Carter, “Advance Directives,” 32–41. For an illuminating discussion regarding the norms that should govern the continuance of intimate relationships, see Kukla, “A Nonideal Theory of Sexual Consent,” 274–78.

of personal identity posited by the loss narrative, in the sense that it constructs dementia as a process that eats away at the authentic self, with every change in personal identity dragging the person who lives with the condition away from the person they really are and toward something hollowed out and inauthentic. Thus even small changes are subject to suspicion around their authenticity, representing under this distortion steps toward the inevitable destruction of personal identity involved in the progression of dementia.

This leads neatly to the third widespread distortion about dementia: fatalism. Those laboring under the assumptions of this distortion are unable to conceive of a life lived with dementia as worth living, reflecting an implicit commitment to the loss narrative's depiction of dementia as an irreversible descent into the loss of personal identity and personhood. Alongside denying or downplaying the benefits people may accrue through the development of dementia and questioning the authenticity of changes in their personalities and values, there is a tendency in both research and public discourse to deny or downplay the ability of people living with the condition to overcome the challenges posed by the physical deterioration of their neurological matter.

In research, this fatalism can be observed in the allocation of funding. Between 2011 and 2016, over 95 percent of research funding for dementia by G7 countries was allocated toward cures and disease-modifying treatment, with only the small fraction leftover allocated to research on improving the lives of those that live with the condition.<sup>19</sup> Among the public, we can observe the far-reaching dissemination of this idea through survey data. In the 2023 Dementia Attitudes Monitor conducted by Alzheimer's Research UK, for example, only 12 percent of respondents considered improvements in quality of life a top priority for research, in comparison to the combined 63 percent who favored prioritizing research on cures, prevention, and medication to stop the development of the condition.<sup>20</sup>

It is no doubt important to conduct biomedical research of this kind; to consider this a priority is not to have a distorted understanding of dementia. However, because this research is unlikely to benefit the majority of persons living with dementia today, placing such great emphasis on it is effectively to abandon attempts to improve their lives or to empower them to act. Fatalism of this kind represents a distortion in understanding because it proceeds from premises that are plainly false. The very existence of self-advocates like Bryden who are able to engage in written and spoken advocacy work while living with

19 Pickett and Brayne, "The Scale and Profile of Global Dementia Research Funding," 1888–89.

20 Alzheimer's Research UK, *Dementia Attitudes Monitor*, 55.

dementia stands as evidence against this idea.<sup>21</sup> Moreover, as Bryden notes, it conflicts with our broader understanding of the effects of social context on capabilities and the well-established phenomenon of lifelong neuroplasticity.<sup>22</sup> It also directly conflicts with the research data we *do* have on improving the quality of life of people living with dementia, which suggests that social relationships and social engagement correlate with better outcomes and that appropriate social organization can improve functional abilities.<sup>23</sup>

The loss narrative, in sum, is evident in three distortions in public understanding of dementia: negativity bias, denial of authenticity, and fatalism. The first reflects the idea that a life lived with dementia is fundamentally characterized by loss, the second that dementia involves a destruction of personal identity, and the third that dementia sets a person irreversibly on a path to the loss of personal identity and personhood. As each of these distortions is widely disseminated, we have reason to think of the loss narrative as socially dominant and to take seriously Bryden's claims that it is stigmatizing and oppressive.

### 1.2. *The Loss Narrative and Stigma*

A recent survey by Alzheimer's Disease International (ADI) suggests that the stigma of dementia is a pressing global problem, with 85 percent of respondents who live with the condition reporting the experience of it in at least one aspect of their daily lives.<sup>24</sup> This is corroborated by the accounts of dementia self-advocates like Bryden, who describes herself as being "surrounded by negative views of dementia," and Rukiya Mukadam, who reports a "very strong, very powerful" taboo about dementia within the British Kashmiri community to which she belongs and among Asian communities more broadly.<sup>25</sup>

To conduct the survey, the ADI adopted a four-part model developed by Nicolas Rüsçh, Matthias C. Angermeyer, and Patrick W. Corrigan to explain the stigma of mental illness. On this account, stigma occurs when people with relative power internalize negative views about some group (stereotypes), which manifest as negative emotional responses (prejudice) and behavioral responses (discrimination).<sup>26</sup> Accordingly, to calculate the global prevalence of stigma, the ADI report uses self-reported experiences of discrimination attributable

21 Alongside Bryden, see Swaffer, *What the Hell Happened to My Brain?*; and Taylor, *Alzheimer's from the Inside Out*.

22 Bryden, *Will I Still Be Me?* 46–47.

23 Martyr et al., "Living Well with Dementia," 2136; and Poulos et al., "A Comprehensive Approach to Reablement in Dementia," 450–58.

24 Alzheimer's Disease International, *World Alzheimer Report 2019*, 14–15.

25 Bryden, *Will I Still Be Me?* 62; and Mukadam, "Time to Break the Taboo," 234–41.

26 Rüsçh, Angermeyer, and Corrigan, "Mental Illness Stigma," 529–39.

to prejudice grounded in stereotypes held by people with social power over others living with dementia.<sup>27</sup>

For descriptive purposes such as these, the four-part model functions reasonably well. It is not obvious, however, that it offers a concept of stigma that is distinct enough for a philosophical analysis. What is being tracked after all is wrongful discrimination; the mere fact that it has a particular root cause does not by itself require the adoption of a separate concept of stigma. Indeed, it might be thought that the route from power to prejudice is the root cause in the overwhelming majority of wrongful discrimination cases, weakening further the case for describing it in any other terms. This model, in other words, lacks a distinguishing feature that tells us *what stigma is* and why it is of particular concern.

Elizabeth Anderson's approach to stigma fills this gap by introducing the idea of *legitimation*. In her terms, stigmatized people are presented as the "*proper* objects of dishonor, contempt, disgust, fear, or hatred on the basis of their group identities."<sup>28</sup> Stigma, then, consists not merely in powerful people holding stereotypes about a particular group that lead to prejudice and discrimination but in those ideas being held within a social context that causes them to be perceived as legitimate and the actions that follow from them justified. Wrongful discrimination resulting from stigma is, in this sense, either socially approved of or perceived by its perpetrators as such.

This insight can be incorporated into the model provided by Rüsich et al. via the addition of a fifth element: a *legitimizing ideology*. Stigma can then be said to occur when people with relative power propagate, adhere to, and reinforce a socially dominant set of ideas about a group (a legitimizing ideology), from which they derive negative views (stereotypes) that are socially approved of, which manifest as negative emotional responses (prejudice) and behavioral responses (discrimination) that are considered, under the prevailing set of social ideas, justifiable.

With the addition of this fifth element, wrongful discrimination resulting from stigma can be distinguished from cases that lack a clear legitimizing ideology. Here I have in mind those that result from unconscious bias rather than conscious commitment to particular stereotypes and prejudicial attitudes. This line is not always clear cut; wrongful discrimination against minority racial groups, for example, can involve both unconscious biases and commitment to tenets of white supremacy (whether the actors understand them as such or not). Not all such cases, however, are obviously the result of conscious adherence to a socially dominant legitimizing ideology. Consider, for instance,

27 Alzheimer's Disease International, *World Alzheimer Report 2019*, 17–33.

28 Anderson, "Equality," 43 (emphasis added).

employment discrimination against candidates with Northern English accents in the United Kingdom. Accent-based stereotypes and prejudices are well-observed phenomena in the United Kingdom, yet there is no widespread overt support for a set of ideas that would render legitimate the refusal to hire, for example, a person from Manchester for a middle-management role in an office merely because of their accent.<sup>29</sup> In such cases, we can say that there are stereotypes, prejudice, and discrimination, but they are not the result of stigma.<sup>30</sup>

Armed with this distinct concept of stigma, we are better equipped to understand the relationship between the loss narrative and the wrongful discrimination against people living with dementia identified by the ADI. Many of the stereotypes mentioned in the report—including the belief that all people living with dementia are a burden to their families and the health care system, that they are incompetent, and that they are unable to contribute to society—clearly proceed from the widespread distortions in understanding that themselves involve implicit commitment to the idea of the loss narrative.<sup>31</sup> Accordingly, the same can be said of the prejudicial and discriminatory behaviors included in the report. If one believes that people living with dementia are passive and will become decreasingly legitimate representatives of their own interests as the condition progresses, then one might feel justified in not taking their opinions seriously or in denying them choices.<sup>32</sup> Even behaviors of aversion, like shunning or otherwise avoiding people living with dementia, make a certain kind of sense when coupled with the distortions of negativity bias, denial of authenticity, and fatalism;<sup>33</sup> persons might wonder why it is so important to maintain a social relationship with someone who, according to the legitimating ideology of the loss narrative, is irreversibly disintegrating in front of them.

Bryden is, in this sense, right to connect the stigma faced by people living with dementia to the loss narrative. The idea that people living with dementia are irreversibly set on a path to the destruction of their personal identity and

29 Sharma, Levon, and Ye, “50 Years of British Accent Bias.”

30 This may not be true for other British accents, especially where they more directly interact with the legitimating ideologies of white supremacy and Anglocentrism. Multicultural London English accents, for instance, seem quite clearly stigmatized due to their association with minority racial groups, discrimination against whom is afforded legitimacy by the socially dominant ideology of white supremacy. Likewise, working-class Glaswegian accents are plausibly understood as stigmatized because the discrimination that follows from widely held associations with violence and poor education is plausibly understood as socially sanctioned under the legitimating ideology of Anglocentrism—a set of ideas that places England and Englishness at the head of a hierarchy of nations and national identities within the UK.

31 Alzheimer’s Disease International, *World Alzheimer Report 2019*, 17.

32 Alzheimer’s Disease International, *World Alzheimer Report 2019*, 24.

33 Alzheimer’s Disease International, *World Alzheimer Report 2019*, 24.

personhood, alongside the distortions in understanding that assume it, seems to play the role of a legitimating ideology in the wrongful discrimination identified by the ADI report. Without its social dominance, absent another legitimating ideology, such behaviors would be the result of socially disapproved of, unconscious biases. The loss narrative is thus crucial to understanding why people living with dementia are stigmatized rather than merely subject to wrongful discrimination.

### 1.3. *The Loss Narrative and Oppression*

Let us turn now to Bryden's claim that the loss narrative contributes to the oppression of people living with dementia. On Iris Marion Young's influential account, oppression is an umbrella term referring to five distinct but related structural injustices: exploitation, marginalization, cultural imperialism, powerlessness, and violence.<sup>34</sup> While a case could be made that people living with dementia suffer from all five, here I want to focus on the two most clearly connected to the loss narrative: marginalization and cultural imperialism.

Marginalization, on Young's account, consists of a process by which a whole group of people are "expelled from useful participation in social life" such that they are blocked from exercising their capacities in "socially defined and recognised ways."<sup>35</sup> The ADI report, which is the largest global survey to date on attitudes toward dementia, does not address this face of oppression directly. Nevertheless, by observing the types of discriminatory behavior the respondents reported suffering from, a clear pattern of exactly this sort of expulsion can be discerned.

A significant number of respondents, for example, reported (i) having their rights and responsibilities taken away, (ii) being treated unfairly in their social life, and (iii) being avoided or shunned. One respondent, regarding the first category, told the researchers that he had lost "the right to work and at times to think for [himself]."<sup>36</sup> Another, regarding the second category, reported having been "shunned [in their] effort to help volunteer to prepare and serve [a] meal" at a clubhouse function.<sup>37</sup> Regarding the third, respondents reported no longer being called by close friends and being ostracized by faith communities.<sup>38</sup>

This type of social exclusion is at the heart of Young's concept of marginalization: the relegation of a social group to the margins of a society such that they are denied opportunities to contribute in a way that grants them social

34 Young, *Justice and the Politics of Difference*, 40–41.

35 Young, *Justice and the Politics of Difference*, 53–54.

36 Alzheimer's Disease International, *World Alzheimer Report 2019*, 32.

37 Alzheimer's Disease International, *World Alzheimer Report 2019*, 27.

38 Alzheimer's Disease International, *World Alzheimer Report 2019*, 31.

recognition. Though it may not always be accompanied by material deprivation, especially when people living with dementia have substantial assets to fall back on, this form of social deprivation is a significant impediment to social equality. As Young argues when discussing the elderly as a marginalized group, “even if marginals were provided a comfortable material life within institutions that respected their freedom and dignity, injustices of marginality would remain in the form of uselessness, boredom, and lack of self-respect.”<sup>39</sup>

No doubt the development of dementia is itself a significant impediment to exercising capacities in socially recognized ways. Yet as each of these examples illustrates, this natural barrier is often reinforced with social barriers. There is no reason why people living with dementia, especially in the early stages, ought to be excluded from the workplace entirely, why they cannot help to prepare a meal where assistance is available, nor why they cannot participate in social interactions with their friends and their wider communities. Such oppressive social marginalization clearly reflects widespread dissemination of the fatalist distortion about dementia, grounded in the loss narrative’s idea of irreversibility.

Turning to cultural imperialism, Young defines this face of oppression according to the conflict between the subjective experience of a subordinated group and a dominant group’s interpretation of that experience. Where this is operative, she argues that the subordinated group experiences a “paradoxical oppression” in which their own interpretations of their experiences are rendered partially or fully invisible, but they are “stamped with an essence” consisting of highly visible, widely known stereotypes about their experiences, which are derived from the dominant group’s perspective.<sup>40</sup>

The effects of this face of oppression on people living with dementia are readily apparent in Bryden’s discussion of the loss narrative. The idea that dementia irreversibly sets those who live with it on a path to a loss of personal identity and personhood, she argues, is an “outsider’s view” that does not reflect the “insider’s experience.”<sup>41</sup> While Bryden discusses gains in attentiveness to present time, the loss narrative instead presents dementia as a process that solely removes capabilities.<sup>42</sup> While Bryden claims that she has experienced a continuity of self throughout the progression of her condition, the loss narrative instead presents dementia as a disintegration of self, such that people who live with the condition cease to represent their true and authentic characters.<sup>43</sup>

39 Young, *Justice and the Politics of Difference*, 55.

40 Young, *Justice and the Politics of Difference*, 58–61.

41 Bryden, *Will I Still Be Me?* 11.

42 Bryden, *Dancing with Dementia*, 11.

43 Bryden, *Will I Still Be Me?* 23–25.

While Bryden clear-sightedly explains how she has used technology to adapt to and overcome the effects of dementia on her short-term and working memory, the loss narrative instead presents people living with dementia as passive and helpless in the face of a total annihilation of what it means to be a human.<sup>44</sup>

This hiding of the insider's perspective under a thick cloud of outsider-derived stereotypes is the essence of cultural imperialism, and it has significant effects on those who experience it. As Young argues, because members of groups oppressed in this way are regularly forced to react to the behavior of others influenced by these stereotyped images, they must expend a significant amount of energy to resist internalizing them and to maintain the positive sense of themselves needed to challenge them.<sup>45</sup> Bryden describes exactly this phenomenon when discussing the heavy burden of reactions to her advocacy work, noting "I am thought to lack insight, so it does not matter if I am excluded. But if I do have insight, then I am said to lack credibility as a true representative for people with dementia."<sup>46</sup> Indeed, when she included brain scans in her presentation slides in order to prove her credibility, she was even accused of faking them, reporting in a 2005 interview that she was told "if your brain scans are really yours, you shouldn't be able to speak."<sup>47</sup>

In sum, Bryden's claim that people living with dementia are stigmatized and oppressed by a socially dominant loss narrative is well supported. The widespread dissemination of three distortions in understanding about dementia—negativity bias, denial of authenticity, and fatalism—reflects implicit commitment to the loss narrative's depiction of dementia as a condition that inevitably and irreversibly leads to the loss of personal identity and personhood. These ideas, I have argued, act as a legitimating ideology in the stigma of dementia, as identified by the 2019 ADI report. Further, they are clear causal contributors to the oppression of people living with dementia through marginalization and cultural imperialism. The loss narrative stands therefore as an impediment to social equality for people living with dementia.

## 2. THE SIGNIFYING ROLES OF PERSONHOOD

The previous section bolstered the political case, originating in Bryden, against the loss narrative. These arguments, however, are not conclusive, because they are sufficient to motivate action to challenge the loss narrative only if all of

44 Bryden, *Will I Still Be Me?* 27–41.

45 Young, *Justice and the Politics of Difference*, 60.

46 Bryden, *Dancing with Dementia*, 40.

47 Rix, "I Live in a Little Cloud."

the claims underpinning it are wrong. In making these arguments, I offered evidence that contradicts the idea that losses incurred by dementia are irreversible and suggested that people living with dementia can possess personal identities no less authentic than those they possessed at onset (whether or not they are metaphysically the same person). I did not, however, address the claim that people living with dementia are at risk of losing their personhood. In part, I chose to omit discussion of this aspect of the loss narrative because it was possible to demonstrate the social dominance and stigmatizing, oppressive character of this set of ideas without doing so. Primarily, however, I chose to bracket this concern because it is of such consequence for the overall argument that it ought to be dealt with separately.

If it were true that dementia eventually causes a person that lives with it to lose their personhood, then the political case I raised in the previous section would be significantly undermined. It would remain the case that the loss narrative leads to distortions in understanding and carries a stigmatizing, oppressive character, but the appropriate response to the narrative would shift from disavowal to reform; it would be no kindness, after all, to deceive people living with the early stages of dementia into thinking that they would not experience this fundamental loss.

More significantly, if there is a significant subclass of people living with dementia who are no longer persons, and all those who live with dementia are irreversibly set on a path toward joining them, then the idea that the harms engendered by the loss narrative represent *injustices* would be significantly undermined. This is so because the concept of personhood carves out the normative landscape via three significant signifying roles such that our moral duties toward those who do not possess it differ significantly from those who do. Falling on the wrong side of it, as I demonstrate in this section, would weaken the ability of people living with dementia, particularly those with advanced dementia, to make moral and political claims on others.

These three signifying roles are well captured by Eva Feder Kittay, who describes personhood as a concept that “marks the moral threshold above which equal respect for the intrinsic value of an individual’s life is required and the requirements of justice are operative and below which only relative interest has moral weight.”<sup>48</sup> To claim that some being is a person is accordingly to signify one or more of the following three things about them: (1) that we owe stronger moral duties toward that being than those that are not persons (i.e., that they cross the relevant threshold), (2) that that being has the same moral status as other beings that are persons (namely, the status that entitles them to equal

48 Kittay, “At the Margins of Moral Personhood,” 101.

respect for the intrinsic value of their lives), and (3) that the duties we owe to that being are duties of social justice rather than those of basic moral concern.

Personhood often plays signifying role 1 in discussions about animal ethics.<sup>49</sup> A common argument, which can be traced back to the work of John Locke, goes something like this: while we may have some duties of moral consideration toward nonhuman animals, they are not *persons*, so they do not have the kind of strong claims against, for example, breeding for medical testing that persons do.<sup>50</sup> Scholars adopting such an argument will typically bolster this point by specifying a set of cognitive features necessary for personhood and then demonstrate that most if not all nonhuman animals do not possess them.<sup>51</sup>

Such conceptions of personhood, however, do not neatly divide human beings and nonhuman animals. Many arguments in favor of abortion rights, for instance, have been mounted in exactly these terms: that human fetuses do not possess the requisite cognitive capacities to meet the threshold of personhood.<sup>52</sup> More troublingly for the subject of this paper, strong cognitive criteria have the effect of excluding a significant number of cognitively disabled persons, including those living with advanced dementia, thereby relegating them to a lower moral status than other humans. Indeed, Brock argues that it is instructive to compare the mental capacities of at least some nonhuman animals and people living with advanced dementia, concluding that neither meet the threshold necessary for personhood.<sup>53</sup>

The serious consequences of such an exclusion are apparent in discussions that evoke personhood's second signifying role. Though sometimes taken to be a relatively thin concept, the moral equality of persons—understood as the idea that the interests of persons are of equal importance—is foundational to contemporary political philosophy, forming what Ronald Dworkin describes as a “kind of plateau.”<sup>54</sup> Beings who are not persons are not typically afforded this status, even when they are subject to the same political institutions as persons, meaning their interests can permissibly be disregarded or overruled without the strong justifications to which persons are entitled.

This distinction reflects relatively well our present-day consensus in Western societies on animal rights issues. While animal welfare concerns are taken

49 See Chan and Harris, “Human Animals and Nonhuman Persons,” 304–6.

50 See Locke, *An Essay Concerning Human Understanding*, 188.

51 Beauchamp, “The Failure of Theories of Personhood,” 310–18.

52 See Himma, “A Dualist Analysis of Abortion,” 48–55; Little, “Abortion and the Margins of Personhood,” 331–38; and Tooley, *Abortion and Infanticide*, 40–104.

53 Brock, “Justice and the Severely Demented Elderly,” 86.

54 Dworkin, “Comment on Narveson,” 25. On the “thin concept” of moral equality of persons, see Nathan, “What Is Basic Equality?” 7–8.

seriously, they are typically taken to be different in kind and importance from those of human welfare. So while there tends to be strong public disapproval of needless cruelty, there is general acceptance (subject to minority dissent, of course) on the permissibility of practices that would clearly violate the moral equality of persons if applied to them, such as those involved in livestock farming and animal testing. The implication of excluding some human beings from the status of personhood, however, is that they too do not possess strong claims of equal consideration, potentially legitimating practices that violate their interests for the sake of others. Kittay strongly rejects the questioning of the personhood of people with severe cognitive disabilities because of exactly these kinds of consequences—consequences that would too befall people living with advanced dementia if they were so excluded.<sup>55</sup>

That nonpersons do not have the same moral claims on us that our fellow persons do of course does not mean that they have no claims at all: the case of animal welfare demonstrates this. Nevertheless, in discussions in which personhood plays signifying role 3, these are claims that are limited in *type* as well as strength. While we can be cruel or inhumane toward a nonperson, we cannot treat them *unjustly*: the requirements of social justice are operative only between persons.

Anderson adopts this type of argument when defending her second-person method of justifying principles of justice, stating that “a claim of justice is essentially expressible as a demand that a *person* makes on an agent whom the speaker holds accountable.”<sup>56</sup> Her primary interlocutor, G. A. Cohen, whom she charges with using a third-person form of justification—justification according to normative and factual premises to which the identity of the person making the argument and the audience are irrelevant—also seems to imply that persons, and persons alone, are the subjects of justice, framing the goal of egalitarian arguments about social justice as an inquiry into the currency it “requires *people* to have equal amounts of.”<sup>57</sup>

If the personhood of people living with dementia were genuinely threatened by the condition, then the stigmatizing and oppressive effects of the loss narrative would not always equate to injustices. Even where they did, moreover, the urgency of addressing them would be blunted by the reinforcement of one of the key distortions the loss narrative engenders: fatalism. If all people living with dementia are irreversibly set on a path to losing the very status that entitles

55 Kittay, “At the Margins of Moral Personhood,” 106–7.

56 Anderson, “The Fundamental Disagreement between Luck Egalitarians and Relational Egalitarians,” 23 (emphasis added).

57 Cohen, “On the Currency of Egalitarian Justice,” 906 (emphasis added).

them to considerations of justice, then any injustices they face are guaranteed to end without the need to mobilize resources and energy against them. This would not absolve the society that failed to address or prevent them, but it would provide grounds to justify, on an all-things-considered basis, prioritizing other pressing issues of justice in circumstances of scarcity.

As with those victims of injustice that have long since died, there may remain a concern of historical injustice for those living with dementia that have lost their personhood. Consideration under this kind of framework, however, falls short of consideration as a standard subject of justice. The dead are not the direct beneficiaries of action to address historical injustice; in cases of individual injustice, it is their surviving loved ones who claim and receive restitution on their behalf, while in cases of group-based injustice, this role falls to surviving members of the social group. Losing their personhood likewise prevents those living with dementia from being direct beneficiaries of action to address those injustices they faced before their condition had progressed. And while loved ones and surviving members of the social group might be motivated to benefit those now nonpersons on whose behalf they have made claims of historical injustice, they would be under no moral obligation to do so. Worse still, historical injustice would cover only that which had occurred before the loss of personhood, so any continuing harms caused by these phenomena would not generate claims of injustice and would pull on only our relative moral concern.

It should be noted that some theorists do think of justice as applying to nonpersons, such that they use the concept of personhood in a way that does not invoke signifying role 3. Even when this is the case, however, a distinction between persons and nonpersons is made, such that justice for nonpersons is of a different priority or of a different kind to justice for persons. Richard Arneson, for instance, has recently stated that principles of justice apply to persons “and other beings as well” but considers the latter a “complication” that needs to be “set aside.”<sup>58</sup> Likewise, Martha Nussbaum dedicates a chapter of her influential monograph *Frontiers of Justice* to the issue of “Justice for Non-Human Animals,” providing a set of arguments that use the same theoretical framework as those that apply to persons but are distinct in their conclusions.<sup>59</sup>

Justice arguments of this kind, if coupled with a conception of personhood that validated this component of the loss narrative, would still present a problem for the political case I raised in section 1. If justice for nonpersons were of a different kind than that of persons, then it is possible that the propagation of stigmatizing and oppressive language would not represent as serious as an

58 Arneson, “Responsibility and Distributive Justice,” 412.

59 Nussbaum, *Frontiers of Justice*, 325–405.

injustice for the former as for the latter (or may not represent an injustice at all). We do not after all tend to describe negative language used toward animals as stigmatizing or otherwise unjust unless we have a prior commitment to them holding the same moral status and, as such, being worthy of the same level of respect as persons.<sup>60</sup>

In sum, because personhood plays three crucial signifying roles in carving out the moral landscape, the political case against the loss narrative would be severely blunted if it were true that people living with dementia were at risk of losing it. Because persons are typically taken to have a higher moral status than nonpersons, the interests of those people living with dementia who had lost their personhood would be of lower weight. Because only persons are typically taken to be moral equals, they would not be entitled to equal consideration in the formation of political and social structures. Finally, because only persons are typically taken to be subjects of justice (and even when nonpersons are included, they are not typically included in the same way as persons), the urgency of tackling the stigma and oppression engendered by the loss narrative would decrease significantly. Successfully rejecting the loss narrative, then, requires defending the personhood of all people living with dementia.

### 3. THE PERSON AS ENVIRONMENTALLY INTEGRATED

Defending people living with dementia against the idea that the condition inevitably and irreversibly threatens their personhood will not on its own suffice to overcome the social dominance of the loss narrative entirely. It is nevertheless a necessary first step toward that goal. Of the available accounts of personhood in the literature, however, there are few that are promising for making it. Evidently, most accounts that rely on some cognitive “performance criterion,” to use Michael Bérubé’s term, will support the loss narrative because of the cognitive deterioration involved in dementia; persons living with advanced dementia are unlikely to meet Jeff McMahan’s criterion of having a “rich and complex mental life,” for instance.<sup>61</sup> Perhaps more surprisingly, as Bryden notes, social accounts such as dementia studies pioneer Tom Kitwood’s view of personhood by social bestowal can have a similar effect.<sup>62</sup> Where such accounts render one’s personhood reliant on recognition by others—recognition that people living with dementia are at great risk of not receiving due to the social

60 For a representative argument of the latter kind, see Milburn and Cochrane, “Should We Protect Animals from Hate Speech?” 1149–72.

61 Bérubé, “Equality, Freedom, and/or Justice for All,” 100; and McMahan, *The Ethics of Killing*, 45.

62 Bryden, *Will I Still Be Me?* 22–23.

injustices they face—they render it, in the words of Hojjat Soofi, “unreasonably socially contingent.”<sup>63</sup>

In light of this, I aim in this section to develop and defend an account of personhood that is inclusive of all people living with dementia and that presents them as no more at threat of losing it than any other person. On this account of the person as environmentally integrated, personhood is a relational attribute, possessed by all those who share an environment of cognitive extension. In building this view, I make novel links between the extended mind thesis and feminist care ethics. The section begins therefore with a summary of the former, after which I flesh out my account by reference to the latter. I then close by considering some objections.

### 3.1. *The Extended Mind Thesis: A Primer*

In an influential 1998 article, Andy Clark and David Chalmers defend two distinct but related claims about the nature of human cognition. First, drawing on phenomena such as the use of calculators and writing tools, they argue that aspects of the external environment are often intimately involved in our cognitive processes in such a way that they both support and enhance them. Second, they argue that under specific conditions, these aspects of the external environment are rightly considered constituent parts of our minds.<sup>64</sup>

The first of these claims, at least for those of us who find ourselves increasingly reliant on technology to complete cognitive tasks, ought to be intuitive. Our social world abounds with cognition-supporting artefacts (among other extra extracranial elements), without which certain processes would be difficult or even impossible to engage in. Few but the most gifted mathematicians, for instance, could intracranially complete the kind of complex calculations supported by calculators, and even then, it is unlikely they would be able to do so at speed. Likewise, few but the most gifted musicians would be able to compose and arrange a complex piece of music for multiple instruments without the aid of music manuscript paper. It is the second claim, however, that distinguishes the *extended* mind thesis from cognate theories about cognitive scaffolding, stirring significant controversy along the way.<sup>65</sup>

In developing their argument, Clark and Chalmers ask us to consider the following two cases:

63 Soofi, “Normative Force of Appeals to Personhood in Dementia Care,” 888.

64 Clark and Chalmers, “The Extended Mind,” 9–12.

65 For a representative argument in favor of cognitive scaffolding but against cognitive extension, see Sterelny, “Minds,” 465–81.

Inga hears from a friend that there is an exhibition at the Museum of Modern Art and decides to go see it. She thinks for a moment and recalls that the museum is on 53rd Street, so she walks to 53rd Street and goes into the museum. . . . Otto suffers from Alzheimer's disease, and like many Alzheimer's patients, he relies on information in the environment to help structure his life. Otto carries a notebook around with him everywhere he goes. When he learns new information, he writes it down. When he needs some old information, he looks it up. . . . Today, Otto hears about the exhibition at the Museum of Modern Art and decides to go see it. He consults the notebook, which says the museum is on 53rd Street, so he walks to 53rd Street and goes into the museum.<sup>66</sup>

Otto's notebook, Clark and Chalmers argue, plays the same role as Inga's biological memory in retrieving the address of the museum.<sup>67</sup> It is thus subject to what has since been termed the *parity principle*, expressed by Clark in solo-authored work as: "if, as we confront some task, a part of the world functions as a process which, were it done in the head, we would have no hesitation in recognizing as part of the cognitive process, then that part of the world is (for that time) part of the cognitive process."<sup>68</sup>

The extended mind thesis evidently represents a radical reconceptualization of the metaphysics of cognition. It should be noted, however, that the second claim is not intended to entirely subsume the first; some—likely most—extracranial elements are genuinely merely supportive. This is so because, as Clark has since emphasized, the "no hesitation in recognizing" condition of the parity principle is fairly stringent, requiring that the extracranial element (a) is reliably and typically invoked, (b) contains information that is more or less automatically endorsed, and (c) contains information that is easily accessible as and when required.<sup>69</sup> While Otto's notebook meets these criteria, other artefacts that might be used to retrieve the address of the museum, such as a rarely-consulted book or a device with access to the internet, would not.

Its limited range of application nevertheless has not spared the second claim of the extended mind thesis from criticism. Fred Adams and Ken Aizawa, for example, have argued that it rests on a "coupling-constitution fallacy" that inappropriately labels all elements coupled with the mind as part of it and that its proponents have failed to demonstrate that external elements like Otto's notebook are genuinely part of the mind because they have not demonstrated that they

66 Clark and Chalmers, "The Extended Mind," 12–13.

67 Clark and Chalmers, "The Extended Mind," 13.

68 Clark, "Memento's Revenge," 44.

69 Clark, "Memento's Revenge," 46.

bear the “mark of the cognitive.”<sup>70</sup> Likewise, Keith Butler has dismissed it on the grounds that the “final locus of computational and cognitive control resides in the head of the subject.”<sup>71</sup> While there is insufficient space to do justice to these debates here, it is worth briefly summarizing the responses Clark has made to these critics, as they help to clarify the shape and boundaries of the thesis.

Regarding Adams and Aizawa, Clark has responded that the thesis does not involve the claim that external elements can be by themselves cognitive. The point rather is that they can sometimes be properly considered parts of a cognitive system, consisting of both internal and external resources. So in the case of Otto, it is not that the notebook itself believes that the address of the Museum of Modern Art is at 53rd Street but that the cognitive system of which both Otto’s brain and the notebook are a part holds that belief. Consequently, the “mark of the cognitive,” whatever it consists in, is borne by the whole system, not its constituent parts.<sup>72</sup>

Regarding Butler, Clark has disputed the very idea that the “final locus of computation and cognitive control” determines the boundaries of the agent. Long-term memory stores after all play no more of a part in Inga’s final choosing than Otto’s notebook does, but both contain information that significantly influences their bearers’ identities as agents. While it might be appropriate to identify the locus of final choosing with consciousness or subjectivity, then, Clark argues that to do so with the cognitive agent is to “shrink the mind and self beyond recognition, reducing [its bearer] to a mere bundle of control processes targeted on occurrent mental states.”<sup>73</sup>

While it is not without its critics, then, the extended mind thesis should not be caricatured. It does not imply that all cognition-supporting extracranial elements are parts of their users’ minds—only those that meet conditions of deep integration. It does not imply that external elements are capable of bearing mental states (by themselves)—only that they can be parts of a cognitive system that bears those states. Finally, it posits only that cognition can be (and often is) extended: related but distinct concepts like consciousness and subjectivity may very well be entirely intracranial.

### 3.2. *Sharing an Environment of Cognitive Extension*

If the extended mind thesis is right, it requires a significant shift in the way we think about cognition. More broadly, it also requires a significant shift in what

70 Adams and Aizawa, “Defending the Bounds of Cognition,” 67–68.

71 Butler, *Internal Affairs*, 205.

72 Clark, “Coupling, Constitution, and the Cognitive Kind,” 96–97.

73 Clark, “*Memento’s* Revenge,” 55–56.

we think of as core human capabilities. Clark anticipates this in solo-authored work, arguing that human beings are primed to both seek cognitive support from external sources and integrate those elements into our cognitive systems, making us “natural-born cyborgs.”<sup>74</sup>

It would be tempting, given both the centrality of dementia to Clark and Chalmers’s case for the extended mind thesis and Bryden’s extensive use of technology to enable her to write, to ground the conception of personhood necessary to challenge the loss narrative in this status.<sup>75</sup> “Persons as natural-born cyborgs” would be an account that included people living with dementia and that presumably encouraged reverence of the extracranial elements they rely on to fulfil cognitive functions. It might thus be a politically useful account to rally around when advocating on the behalf of those living with dementia.

It is not obvious, however, that such an account could successfully play the three signifying roles typically played by conceptions of personhood to carve out the moral landscape. The ability to extend one’s mind (or indeed, having a mind that is capable of being extended) is certainly impressive, but it seems morally arbitrary—akin to flight or other species-specific abilities. Without additional information, it is difficult to see how the ability could possibly be used to justify a moral hierarchy between those beings that have it and those that do not. Furthermore, it seems like a capacity that some are able to exercise better than others, calling into question its ability to ground the moral equality of persons and, by implication, its suitability for determining the scope of justice—the issues with using scalar properties in this way are the subject of extensive debate in the literature on moral equality.<sup>76</sup>

This is a significant problem, because any account of personhood I develop here needs to be conceptually robust enough to replace those in common usage. If not, then it is unlikely to gain traction as an alternative, dampening its ability to contribute to overcoming the loss narrative. Instead of developing a conception of personhood grounded in the mere capacity for cognitive extension, then, I want to further complicate—and hopefully strengthen—this picture by introducing an additional element: the moral significance of human relationships.

Theorists working in feminist care ethics have long criticized the idea of the independent rational agent, highlighting the unique extent and length of human juvenile care needs and the complex ways in which human beings depend on

74 Clark, *Natural-Born Cyborgs*, 31.

75 Bryden, *Will I Still Be Me?* 27–34.

76 For a helpful summary of the relevant literature, see Sher, “Why Are We Moral Equals?” 17–29.

one another to meet their needs as adults.<sup>77</sup> It is in this spirit that Kittay rejects theories of personhood that would exclude her cognitively disabled daughter, arguing that our moral duties toward one another originate in and depend on for their character “a matrix of relationships embedded in social practices through which the relations acquire meanings.”<sup>78</sup> Similar arguments have been made by Hilde Lindemann, who draws attention to the social practices we engage in to initiate other humans into personhood and hold them there through recognition and social identity shaping.<sup>79</sup> Personhood can be understood in this light as a *relational attribute*: conferred through relationships and given meaning through the social practices within which they are embedded.

While there are great strengths to this view, the process by which the relationships gain such significant moral meanings—significant enough to ground the kind of moral hierarchy engendered by the three signifying roles of personhood—is a little hazy. In Kittay’s description of the process, no causal link is offered to explain the conferring of moral meanings onto social relations by social practices, leaving a key component of the care ethics view of the moral significance of human relationships unspecified. Similarly, while Lindemann’s account of the relational composition and reinforcement of personal identity is sociologically persuasive, it is not immediately clear we should think that these practices are a source of moral *value*. This lacuna can be filled, I propose, by introducing the view of human cognition implied by the extended mind thesis.

As cognition is strongly environmentally determined, so too is our collective ability to generate, develop, and adhere to moral concepts. Social practices accordingly can be said to confer social relationships with moral meaning because of the way they build and develop the environment into which our minds must extend to be able to engage in moral reasoning. Productive practices, for instance, generate material goods, some of which become objects of moral reasoning—about what they are for and who should receive them—and some of which, such as writing implements, support it by enabling persons to work through and communicate their solutions to complex problems. Likewise, cultural practices, such as those from which languages develop, support us in thinking abstractly and communicating moral ideas to others. In short, such relationships can be said to confer personhood because they are embedded in the social processes from which this moral concept, along with the rest of morality itself, emerges—a quality that is not morally arbitrary and is thus suitable for grounding the moral distinction between persons and nonpersons.

77 Held, *The Ethics of Care*, 10.

78 Kittay, “At the Margins of Moral Personhood,” 111.

79 Lindemann, *Holding and Letting Go*, 1–30.

Note that there is no need here to invoke the kind of performance criteria that can scupper the ability of an account of personhood to justify the moral equality of persons and in turn the ability for such an account to ground justice considerations. While the environment of cognitive extension from which moral concepts are derived is developed through the performance of social practices, it is not active participation that confers personhood but *presence*. To be a person on this view is merely to be integrated into an environment in which these practices are operative, such that one is both shaped by it and contributes to shaping it merely through one's presence as a particular kind of being in a particular kind of relationship that is itself part of the overarching matrix.<sup>80</sup>

All people living with dementia, at any stage of progression, can be a part of this matrix of overlapping relationships merely by being situated in the right sort of relationships. While it is certainly possible that someone might be cruelly cut off from relating in this way, the kind of extreme actions necessary to do so could just as easily be targeted at persons without dementia. To fully remove such a person from the matrix, all those with whom they have significant relationships would need to be removed from it, all records that indicate that they are a specific person who has related to the overarching environment in specific ways over time would need to be destroyed, and they would need to be so completely cut off from human contact that they would have no opportunities to form relationships that would reconnect them to an environment of cognitive extension. As well as being a plausible conception of personhood, then, the idea of the person as environmentally integrated also provides grounds to challenge the loss narrative.

### 3.3. *Three Objections*

I have sketched out a conception of personhood that combines insights from care ethics and the extended mind thesis in order to include people living with dementia and to challenge the loss narrative. Fully exploring its wider consequences is far beyond the scope of this paper and the limited space available. Nevertheless, before closing I want to respond to three pertinent objections that could be made, as each will help to further clarify the account. The three objections pertain to over-inclusion, over-exclusion, and misguidedness.

The first two of these objections are standard fare for any account of personhood: such accounts need to draw the line somewhere and so will inevitably

80 Note that this presence is active and continuing, not static. The continued existence of persons living with dementia in an environment related in specific ways to other persons and artefacts leads to continual reshaping of the meaning-making practices that arise within it, just as does the presence of other persons. For an exploration of these sorts of dynamics, see Chapman, Philip, and Komesaroff, "Towards an Ecology of Dementia," 209–16.

face backlash for including certain beings and excluding others. At this stage, I can anticipate two such arguments that this account may face: first, that by making the key criterion for personhood relational, it risks including beings or even objects that some eccentric persons report relationships with; and second, that by doing so it excludes human beings we are not related with yet we ought to recognize as persons—namely, so-called uncontacted peoples who live without sustained contact with the international community.

Regarding the first, it ought to be noted that care-ethics-style reasoning requires a two-way relationship; a person is not in the relevant sort of relationship with their car merely because they profess to love it, for instance.<sup>81</sup> Accordingly, though it is presence and not performance that determines whether a being is a person, that presence depends on the existence of very minimal capacities to relate to other humans—that is, those elements of social cognition that involve detecting and responding to other humans, which appear to be preserved to at least a minimal degree in the progression of all dementias.<sup>82</sup> While this might rule inanimate objects out, some may yet be concerned that it captures at least some nonhuman animals kept as pets, such as dogs and cats, all of whom may turn out to have such capacities. If such a conclusion is entailed, however, it is not necessarily so bizarre as to undermine the account; our practices suggest that we do in fact regard certain nonhuman animals as possessing a significant moral status when we are related to them in particularly meaningful ways, so it would not be too much of a stretch from common-sense moral reasoning to describe them as persons.

Regarding the second argument, it ought to be noted that such peoples, if they are in the right kinds of relationships with each other, do share an environment of cognitive extension. It is not *our* environment, but that does not preclude them from possessing personhood—that is, being persons *to each other*. While what being a person means may differ across matrices of overlapping relationships, we can recognize from within our own the practices of meaning-making and moral reasoning that take place in others, especially among other humans whom we can reasonably assume would be persons to us were we related to them in the right sort of ways. The account therefore does not entail that we have no moral duties toward outsiders or reasons to treat them as persons merely because they do not share *our* environment of cognitive extension.

81 Noddings, “Care Ethics and ‘Caring’ Organizations,” 77–79.

82 See Heitz et al., “Cognitive and Affective Theory of Mind in Dementia with Lewy Bodies and Alzheimer’s Disease,” 373–81; Harciarek and Cosentino, “Language, Executive Function and Social Cognition in the Diagnosis of Frontotemporal Dementia Syndromes,” 178–96; and Batra, Sullivan, Williams, and Geldmacher, “Qualitative Assessment of Self-Identity in People with Advanced Dementia,” 1260–78.

The last potential criticism concerns the wisdom of the entire enterprise. Kittay has raised significant concerns about the practice of making moral distinctions between human beings—a practice with a dark and bloody history.<sup>83</sup> Along these lines, a critic might object that we would be better off rejecting the concept of personhood altogether due to the political risks it may pose to human beings (or nonhuman animals) who might fall—or be falsely thought to fall—on the wrong side of the threshold. Indeed, it has been argued that a focus on personhood is not even particularly helpful in the specific case of people living with dementia.<sup>84</sup>

To this I make two (tentative) responses. First, it might seem that we need a way of distinguishing morally between human beings in order to resolve conflicts around abortion rights, the status of anencephalic children, the moral permissibility of embryonic research, and cognate issues. And even if objectors do not accept that it is philosophically necessary, they ought to consider why the use of a conception of personhood might be *politically* necessary. The political context is one in which the term “personhood” is used and widely invoked. Postponing the defense of the personhood of people living with dementia to engage in the much more arduous enterprise of extinguishing it from our moral vocabulary entirely comes with a significant opportunity cost that is likely to serve members of this group badly in the short to medium term.

The account of the person as environmentally integrated, then, can be used to defend the personhood of people living with dementia and can resist some of the concerns raised here. Because it is not based on a morally arbitrary attribute, it is suitable for playing the signifying role of establishing a moral hierarchy between persons and nonpersons. Because it is not based on a performance criterion, it is suitable for grounding the idea that persons are morally equal, which in turn can ground the idea of social justice. By combining insights from care ethics and the extended mind thesis, the idea of loss of personhood present in the loss narrative can be rejected.

#### 4. CONCLUSION

In this paper I have bolstered the political case made by self-advocate Christine Bryden against a key contributor to social injustices faced by people living with dementia: the loss narrative. By examining the distortions it engenders and its contribution to stigma and oppression, I have argued that there are strong and urgent reasons to challenge it. In order to do so, however, it is necessary

83 Kittay, “At the Margins of Moral Personhood,” 114–26.

84 Higgs and Gilleard, “Interrogating Personhood,” 773–80.

to develop an account of personhood that is inclusive of people living with dementia, such that the idea of the loss of personhood contained within the narrative can be rejected. I have therefore set out an account of the person as environmentally integrated.

University of Glasgow  
matilda.carter@glasgow.ac.uk

#### REFERENCES

- Adams, Fred, and Ken Aizawa. "Defending the Bounds of Cognition." In Menary, *The Extended Mind*, 67–80.
- Alcoff, Linda. "The Problem of Speaking for Others." *Cultural Critique*, no. 20 (1991): 5–32.
- Alzheimer's Disease International. *World Alzheimer Report 2019: Attitudes to Dementia*. London: Alzheimer's Disease International, 2019.
- Alzheimer's Research UK. *Dementia Attitudes Monitor: Wave 3*. November 2023. <https://www.dementiastatistics.org/attitudes/Dementia%20Attitudes%20Monitor%20-%20Wave%203%20report.pdf>.
- Anderson, Elizabeth. "Equality." In *The Oxford Handbook of Political Philosophy*, edited by David Estlund, 40–57. New York: Oxford University Press, 2012.
- . "The Fundamental Disagreement between Luck Egalitarians and Relational Egalitarians." In *Distributive Justice and Access to Advantage*, edited by Alexander Kaufman, 21–39. Cambridge: Cambridge University Press, 2015.
- Arneson, Richard. "Responsibility and Distributive Justice." In *The Oxford Handbook of Moral Responsibility*, edited by Dana K. Nelkin and Derk Pereboom, 412–30. New York: Oxford University Press, 2022.
- Batra, Sadhvi, Jacqueline Sullivan, Beverly R. Williams, and David S. Geldmacher. "Qualitative Assessment of Self-Identity in People with Advanced Dementia." *Dementia* 15, no. 5 (2016): 1260–78.
- Beauchamp, Tom L. "The Failure of Theories of Personhood." *Kennedy Institute of Ethics Journal* 9, no. 4 (1999): 309–24.
- Bérubé, Michael. "Equality, Freedom, and/or Justice for All." In *Cognitive Disability and Its Challenge to Moral Philosophy*, edited by Eva F. Kittay and Licia Carlson, 97–110. Oxford: Wiley-Blackwell, 2010.
- Brock, Dan W. "Justice and the Severely Demented Elderly." *Journal of Medicine and Philosophy* 13, no. 1 (1988): 73–99.
- Bryden, Christine. *Dancing with Dementia: My Story of Living Positively with Dementia*. London: Jessica Kingsley Publishers, 2005.

- . *Nothing about Us, without Us! 20 Years of Dementia Advocacy*. London: Jessica Kingsley Publishers, 2015.
- . *Will I Still Be Me? Finding a Continuing Sense of Self in the Lived Experience of Dementia*. London: Jessica Kingsley Publishers, 2018.
- Butler, Keith L. *Internal Affairs: A Critique of Externalism in the Philosophy of Mind*. Dordrecht: Kluwer, 1998.
- Caddell, Lisa S., and Linda Clare. “I’m Still the Same Person: The Impact of Early-Stage Dementia on Identity.” *Dementia* 10, no. 3 (2011): 379–98.
- Carter, Matilda. “Advance Directives: The Principle of Determining Authenticity.” *Hastings Center Report* 52, no. 1 (2022): 32–41.
- Chan, Sarah, and John Harris. “Human Animals and Nonhuman Persons.” In *The Oxford Handbook of Animal Ethics*, edited by Tom L. Beauchamp and R. G. Frey, 304–31. New York: Oxford University Press, 2011.
- Chang, Chia-Yu, and Hui-Chuan Hsu. “Relationship between Knowledge and Types of Attitudes towards People Living with Dementia.” *International Journal of Environmental Research and Public Health* 17, no. 11 (2020): 3777.
- Chapman, Michael, Jennifer Philip, and Paul Komesaroff. “Towards an Ecology of Dementia: A Manifesto.” *Journal of Bioethical Inquiry* 16, no. 2 (2019): 209–16.
- Clark, Andy. “Coupling, Constitution, and the Cognitive Kind: A Reply to Adams and Aizawa.” In Menary, *The Extended Mind*, 81–99.
- . “Memento’s Revenge.” In Menary, *The Extended Mind*, 40–66.
- . *Natural-Born Cyborgs: Minds, Technologies, and the Future of Human Intelligence*. New York: Oxford University Press, 2003.
- Clark, Andy, and David Chalmers. “The Extended Mind.” *Analysis* 58, no. 1 (January 1998): 7–19.
- Cohen, G. A. “On the Currency of Egalitarian Justice.” *Ethics* 99, no. 4 (July 1989): 906–44.
- Dewing, Jan. “Participatory Research: A Method for Processing Consent with Persons Who Have Dementia.” *Dementia* 6, no. 1 (2007): 11–25.
- Dworkin, Ronald. “Comment on Narveson: In Defense of Equality.” *Social Philosophy and Policy* 1, no. 1 (Autumn 1983): 24–40.
- Harciarek, Michał, and Stephanie Cosentino. “Language, Executive Function and Social Cognition in the Diagnosis of Frontotemporal Dementia Syndromes.” *International Review of Psychiatry* 25, no. 2 (2013): 178–96.
- Heitz, Camille, Vincent Noblet, Clélie Phillipps, Benjamin Cretin, Natacha Vogt, Nathalie Philippi, Jennifer Kemp, et al. “Cognitive and Affective Theory of Mind in Dementia with Lewy Bodies and Alzheimer’s Disease.” *Alzheimer’s Research and Therapy* 8, no. 1 (2016): 373–81.
- Held, Virginia. *The Ethics of Care: Personal, Political, and Global*. Oxford: Oxford

- University Press, 2006.
- Higgs, Paul, and Chris Gilleard. "Interrogating Personhood." *Personhood, Identity and Care in Advanced Old Age* 20, no. 8 (2016): 773–80.
- Himma, Kenneth E. "A Dualist Analysis of Abortion: Personhood and the Concept of Self *qua* Experiential Subject." *Journal of Medical Ethics* 31, no. 1 (2005): 48–55.
- Kittay, Eva F. "At the Margins of Moral Personhood." *Ethics* 116, no. 1 (October 2005): 100–31.
- Kitwood, Tom. *Dementia Reconsidered: The Person Comes First*. Buckingham: Open University Press, 1997.
- . "On Being a Person." In *Dementia Reconsidered, Revisited: The Person Still Comes First*, edited by Dawn Brooker, 2nd ed., 6–16. New York: Open University Press, 2019.
- Kukla, Quill R. "A Nonideal Theory of Sexual Consent." *Ethics* 131, no. 2 (January 2021): 270–92.
- Lindemann, Hilde. *Holding and Letting Go: The Social Practice of Personal Identities*. New York: Oxford University Press, 2014.
- Little, Margaret O. "Abortion and the Margins of Personhood." *Rutgers Law Journal* 39, no. 2 (Winter 2008): 331–48.
- Locke, John. *An Essay Concerning Human Understanding*. New York: Oxford University Press, 1964.
- Lopez, Ruth P., Karen M. Rose, Lauren Kenney, Victoria Sanborn, and Jennifer D. Davis. "Managing Shame: A Grounded Theory of How Stigma Manifests in Families Living with Dementia." *Journal of the American Psychiatric Nurses Association* 26, no. 2 (2020): 181–88.
- Martyr, Anthony, Sharon M. Nelis, Catherine Quinn, Yu-Tzu Wu, Ruth A. Lamont, Catherine Henderson, Rachel Clarke, et al. "Living Well with Dementia: A Systematic Review and Correlational Meta-Analysis of Factors Associated with Quality of Life, Well-Being and Life Satisfaction in People with Dementia." *Psychological Medicine* 48, no. 13 (2018): 2130–39.
- McMahan, Jeff. *The Ethics of Killing: Problems at the Margins of Life*. New York: Oxford University Press, 2002.
- McManus, Maria, and Paula Devine. *Dementia: Public Knowledge and Attitudes*. Access Research Knowledge, 2011. <https://www.ark.ac.uk/ARK/sites/default/files/2018-08/update77.pdf>.
- Menary, Richard, ed. *The Extended Mind*. Cambridge, MA: MIT Press, 2010.
- Milburn, Josh, and Alasdair Cochrane. "Should We Protect Animals from Hate Speech?" *Oxford Journal of Legal Studies* 41, no. 4 (2021): 1149–72.
- Mukadam, Rukiya. "Time to Break the Taboo." In *People with Dementia Speak Out*, edited by Lucy Whitman, 234–41. London: Jessica Kingsley Publishers,

- 2016.
- Nathan, Christopher. "What Is Basic Equality?" In *Do All Persons Have Equal Moral Worth? On "Basic Equality" and Equal Respect and Concern*, edited by Uwe Steinhoff, 1–16. New York: Oxford University Press, 2015.
- Noddings, Nel. "Care Ethics and 'Caring' Organizations." In *Care Ethics and Political Theory*, edited by Daniel Engster and Maurice Hamington, 72–84. New York: Oxford University Press, 2015.
- Nussbaum, Martha C. *Frontiers of Justice: Disability, Nationality, Species Membership*. Cambridge, MA: Harvard University Press, 2006.
- Pickett, James, and Carol Brayne. "The Scale and Profile of Global Dementia Research Funding." *Lancet* 394, no. 10212 (2019): 1888–89.
- Poulos, Christopher J., Antony Bayer, Lauren Beaupre, Linda Clare, Roslyn G. Poulos, Rosalie H. Wang, Sytse Zuidema, and Katherine S. McGilton. "A Comprehensive Approach to Reablement in Dementia." *Alzheimer's and Dementia: Translational Research and Clinical Interventions* 3, no. 3 (2017): 450–58.
- Rix, Juliet. "I Live in a Little Cloud." *Guardian*, April 26, 2005. <https://www.theguardian.com/books/2005/apr/26/booksonhealth.lifeandhealth>.
- Rüsch, Nicolas, Matthias C. Angermeyer, and Patrick W. Corrigan. "Mental Illness Stigma: Concepts, Consequences, and Initiatives to Reduce Stigma." *European Psychiatry* 20, no. 8 (2005): 529–39.
- Sharma, Devyani, Erez Levon, and Yang Ye. "50 Years of British Accent Bias." *English World-Wide* 43, no. 2 (2022): 135–66.
- Sher, George. "Why Are We Moral Equals?" In *Do All Persons Have Equal Moral Worth? On "Basic Equality" and Equal Respect and Concern*, edited by Uwe Steinhoff, 17–29. New York: Oxford University Press, 2015.
- Soofi, Hojjat. "Normative Force of Appeals to Personhood in Dementia Care: A Critical Examination of Kitwood's Account of Personhood." *Bioethics* 35, no. 9 (2021): 884–90.
- Sterelny, Kim. "Minds: Extended or Scaffolded?" *Phenomenology and the Cognitive Sciences* 9, no. 4 (2010): 465–81.
- Swaffer, Kate. *What the Hell Happened to My Brain? Living beyond Dementia*. London: Jessica Kingsley Publishers, 2016.
- Taylor, Richard. *Alzheimer's from the Inside Out*. London: Health Professions Press, 2007.
- Tooley, Michael. *Abortion and Infanticide*. New York: Oxford University Press, 1983.
- Webb, Joseph, Val Williams, Marina Gall, and Sandra Dowling. "Misfitting the Research Process: Shaping Qualitative Research 'in the Field' to Fit People Living with Dementia." *International Journal of Qualitative Methods*

19 (2020).

World Health Organization. *A Blueprint for Dementia Research*. Geneva: World Health Organization, 2022.

———. “Dementia” (webpage). Last modified March 15, 2023. <https://www.who.int/news-room/fact-sheets/detail/dementia>.

Young, Iris M. *Justice and the Politics of Difference*. Princeton: Princeton University Press, 1990.

MURDERERS ON THE BALLOT PAPER  
BAD APPLES, MORAL COMPROMISE, AND THE  
EPISTEMIC VALUE OF PUBLIC DELIBERATION  
IN REPRESENTATIVE DEMOCRACIES

*Richard Beadon Williams*

EPISTEMIC DEMOCRATS argue that the legitimacy of democratic authority partially depends on the ability of democratic states to make the correct decisions—the decisions that better promote the common good—more effectively than antidemocratic alternatives. In this paper I argue that epistemic democrats typically overlook the centrality of party politicians to representative democracies. The democratic choice of policy is mediated through the democratic choice of politician.

This paper will defend two core contributions. In a critical direction, the first core contribution is to put the democratic process of politicians competing for votes in elections at the center of political theorizing. Epistemic democrats risk forcing this central characteristic of representative democracy to have an ad hoc fit with their preconceived models of direct democracy. In particular, epistemic democrats overlook the active possibility that a competent public could still yield bad outcomes because of how bad apples behave in the legislature and how conscientious politicians should react. This paper shows that conscientious politicians should compromise with each other in order to gain the political alliances and electoral support necessary to stop the murderers on the ballot paper from winning and wielding political power with killer consequences. So the active possibility of the worst politicians on the ballot paper winning and wielding political power as they wish potentially spoils the epistemic benefits of widespread public competence for the rest of us. Rather than promote the truth, a conscientious politician should compromise the epistemic benefits of widespread public competence with whatever rhetoric, lies, and bullshit will gain the alliances and votes necessary to resist the bad apples, whether the bad apples are inside or outside her political party.

As a moral agent, a party politician with personal integrity may feel compelled to promote her moral convictions, and compromising on those moral convictions may compromise her personal integrity. However, to see only that

moral compromises compromise personal integrity is to overlook whether compromises may cultivate a different type of integrity. A politician, as an elected representative, should cultivate her “democratic integrity” and take responsibility for protecting the material interests of those she represents. An uncompromising politician who prioritizes her personal integrity as a moral agent neglects her democratic integrity as an elected representative.

The bad apples in the legislature provide a powerful reason to construct a less idealized and more realistic model of deliberation. If a model of deliberation is to help show how democracy outperforms antidemocratic alternatives, it must become much more sensitive to the weighty profession-specific obligations of politicians in representative democracies to resist the bad apples.

In a constructive direction, the second core contribution is that deliberation is potentially useful for discovering how to resist the bad apples. This paper will show that a potential epistemic value of public deliberation in *representative* democracies is that it can empower politicians to discover what I call “deliberated compromises.” It allows politicians to persistently know what moral compromises to advocate for in order to resist the bad apples. Deliberation can empower politicians to know which moral compromises will gain the alliances and votes necessary to resist bad apples in light of the constantly changing range and intensity of political sentiments among the public. The bad apples problem provides a powerful reason for more realistic models of deliberation to redirect themselves away from the public promoting truth and toward politicians promoting compromise.

## 1. REPRESENTATIVE DEMOCRACY

### 1.1. *Politician Incompetence*

Epistemic democrats argue that the legitimacy of democratic authority partially depends on the ability of democratic states to produce the correct decisions more effectively than antidemocratic alternatives.<sup>1</sup> Broadly speaking, they argue that aggregating competent votes in elections and voicing diverse views in public deliberations can empower democratic states to discover the correct decisions. As aggregative epistemic democrats, Robert Goodin and Kai Spiekermann have defended the Condorcet Jury theorem, showing that if only competent people

1 Cohen, “An Epistemic Conception of Democracy”; Coleman and Ferejohn, “Democracy and Social Choice”; Anderson, “The Epistemology of Democracy”; Estlund, *Democratic Authority*; Landmore, *Democratic Reason*; Schwartzberg, “Epistemic Democracy and Its Challenges”; Goodin and Spiekermann, *An Epistemic Theory of Democracy*; and Cerovac, *Epistemic Democracy and Political Legitimacy*.

vote, and they vote independently, the competent majority is exponentially more likely to choose the correct decision than a competent minority.<sup>2</sup> As a deliberative epistemic democrat, H el ene Landemore has defended the Diversity Trumps Ability theorem, showing that if a cognitively diverse public spreads the effective problem-solving heuristics scattered among its members during deliberations, the diverse public is more likely to choose the correct decision than a more cognitively able but less cognitively diverse group of experts.<sup>3</sup> Whatever the particular mechanism might be, epistemic democrats provide an attractive ideal that enables them to evaluate the political competence of the public and to aspire toward institutional reforms that should promote their political competence more fully. Against epistemic democracy, epistocrats (those who advocate for expert rule) argue that the public is too politically incompetent to make epistemic democracy a realistic ideal.<sup>4</sup> In defense of epistemic democracy, epistemic democrats argue that the public can and should become politically competent enough to make epistemic democracy a realistic ideal.

In a different direction, I am primarily concerned with a more neglected set of assumptions. The first core contribution of this paper is that epistemic democrats typically overlook the centrality of elected politicians in representative democracies. I will therefore explore whether the assumptions about *politician* competence rather than *public* competence are realistic. Even if epistemic democrats were to assume a fully realistic model of the public, they would still risk a utopian ideal if they assumed a hopelessly optimistic model of elected politicians. The incompetent politicians in the legislature potentially spoil the epistemic benefits of widespread public competence during elections.

In practice, representative democracy rather than direct democracy is typical. Representative democracies typically contain political parties as useful instruments for elected politicians to win elections and govern effectively. Political parties typically unite around common political agendas during elections, but they contain persistent internal divisions with divergent political traditions, policy preferences, and political aspirations among their members. Representative democracy fundamentally changes the type of choices that competent voters must make. They cannot directly choose a policy: they directly choose a politician and indirectly choose a policy. In other words, the democratic choice of policy is mediated through a democratic choice of politician.<sup>5</sup> The political

2 Goodin and Spiekermann, *An Epistemic Theory of Democracy*, 17–36.

3 Landemore, *Democratic Reason*, 89–117.

4 Caplan, *The Myth of the Rational Voter*; Lopez-Guerra, *Democracy and Disenfranchisement*; and Brennan, *Against Democracy*.

5 Even with referendums, the contributions of party politicians seeking to win the next election typically still influence the referendum result.

competence of the public risks pushing against the political incompetence of politicians. Competent voters are less capable of choosing good policies if it is mediated through a choice of bad politicians.

Epistemic democrats typically assume that representative democracy is not fundamentally different from direct democracy.<sup>6</sup> Perhaps epistemic democrats can argue that if a large legislature is constituted correctly, many if not most of the epistemic benefits of public deliberation spill into or are cultivated within the legislature itself.

Epistemic democrats might argue that the epistemic benefits of widespread public competence in representative democracies will probably elect competent politicians. In particular, Goodin and Spiekermann have argued that the difference between representative democracy and direct democracy partially depends on how politicians conceive of their role as elected representatives.<sup>7</sup> First, they argue that competent voters should elect a delegate-style politician who votes in light of the judgements of those she represents rather than a trustee-style politician who votes in light of her personal judgements. In the spirit of Condorcetian democracy, a delegate-style politician is probably a competent politician because the majority judgement of those she represents is exponentially more likely to be the correct judgement than the minority judgement or her personal judgement.

Second, Goodin and Spiekermann have argued that the epistemic benefits of deliberation in the legislature will probably improve the competence of politicians. Even if elections do not always elect competent delegates, deliberation in the legislature will induce competence in otherwise incompetent politicians.<sup>8</sup> Once enough politicians are competent, the aggregation of votes in the legislature probably produces the correct policies, as a competent majority in the legislature is exponentially more likely to be correct than a competent minority.

Landemore has rejected that representative democracy is merely a feasible second best to the unfeasible ideal of direct democracy, arguing that it has particular epistemic advantages.<sup>9</sup> However, Landemore rejects that representative democracy has the elitist epistemic advantage of electing the more capable and competent people to political office. Following political scientist Nadia Urbinati, Landemore has argued that representative democracy has the more egalitarian epistemic advantage of constructing a feedback loop between the

6 Estlund, *Democratic Authority*, 160–61; Landemore, *Democratic Reason*, 4; and Goodin and Spiekermann, *An Epistemic Theory of Democracy*, 254–59.

7 Goodin and Spiekermann, *An Epistemic Theory of Democracy*, 254–46.

8 Goodin and Spiekermann, *An Epistemic Theory of Democracy*, 256–59.

9 Landemore, *Democratic Reason*, 10, 105–6.

people's inputs and the proposals of the representative assembly.<sup>10</sup> This process provides the time necessary for the public and the politicians to revise and refine their judgements, cultivating a reflective type of wisdom regarding the policy preferences of the public and the policy decisions of the politicians.

Contrary to these epistemic democratic expectations, I argue that a critical mass of party politicians in the legislature potentially spoil the epistemic benefits of widespread public competence for the rest of us. Political philosopher Patrick Tomlin has argued that otherwise able groups can become unable to perform collective actions if a critical mass of their membership remains unwilling to contribute enough.<sup>11</sup> He has provided a hypothetical case of one hundred soldiers who need everybody to follow their orders if they are to cross a river. However, there are always three or more soldiers who are able but unwilling to follow their orders. So the few able but unwilling soldiers translate into a willing but unable unit. Similarly, a representative democracy needs enough competent voters and enough competent politicians if they are to produce the correct decisions. However, as explored next, there are potentially more than enough politicians who are able but unwilling to promote the correct decisions. First, the worst politicians on the ballot paper—the bad apples—are typically unwilling to promote the correct decisions. Second, a conscientious politician is not always willing to make the correct decision if it may cost her the next election. Third, a conscientious politician should become willing to compromise on promoting the correct decisions in order to gain the political alliances and electoral support necessary to resist the bad apples. So a critical mass of incompetent politicians in the legislature (including both the bad apples and the conscientious politicians seeking to win the next election and resist the bad apples) potentially translates into an incompetent representative democracy despite widespread public competence.

### *1.2. The Primacy of Electoral Competence*

The ordinary incentives of party politicians competing for votes in representative democracies can significantly blunt the effects of widespread public competence for practical reasons. It is not infeasible for bad politicians to win votes during elections. Good campaigns can elect bad politicians. In order to distinguish between the complex virtues and vices of politicians, it is helpful to distinguish between ethical, epistemic, and electoral competence. Ethically competent politicians are principled and pragmatic enough to do good and avoid harm reliably. Epistemically competent politicians are empirically informed and epistemically

10 Urbinati, *Representative Democracy*.

11 Tomlin, "Should We Be Utopophobes about Democracy in Particular?"

rational enough to know the truth reliably. Electorally competent politicians gain the political alliances and electoral support necessary to win the next election reliably. The circumstances of politics give electorally competent but epistemically or ethically incompetent politicians a competitive advantage over ethically and epistemically competent but electorally incompetent politicians. In particular, hypocrisy often gives politicians a competitive advantage. Politicians frequently do not practice what they preach. First, partisan politicians are prone not to practice what they preach consistently, especially if consistency would disadvantage their political party and inconsistency would advantage their political party. Partisanship has many moral virtues and socially good consequences.<sup>12</sup> Nevertheless, partisanship does risk some moral vices, and hypocrisy is one of them. Second, careerist politicians are disposed to preach the party line and practice whatever they expect to progress their professional careers.<sup>13</sup> Third, Machiavellian politicians are willing to preach virtuous principles and practice whatever they expect to give themselves more political power and personal glory, however ugly.<sup>14</sup> Whatever their motivations might be, electorally competent politicians often preach whatever rhetoric, lies, and bullshit they expect to win critical votes during elections and then practice whatever advances their narrow group or personal interests when in office.

Epistemically incompetent politicians are bad, but ethically incompetent politicians are typically among the worst. An ethically competent but epistemically incompetent politician is typically willing to do good, but she is frequently unable to know how to do good. She often fails to do good, but she is well intentioned. In contrast, an ethically incompetent politician is typically unwilling to do good and willing to do harm. She is willing to promote her own personal good, whatever harm she may do in the process. Ethically incompetent politicians are typically among the worst politicians on the ballot paper. When the infamous bank robber Willie Sutton was asked why he robbed the bank, he is rumored to have answered “because that’s where the money is.” Similarly, some ethically incompetent politicians are on the ballot paper because political office is where the power is. Lacking any significant principled convictions, cult leaders typically advocate for whatever populist policies help to cultivate a personally pleasurable cult of personality. Alternatively, corporatists publicly advocate for whatever populist policies allow them to profit from a kleptocracy or a chumocracy behind closed doors. They redirect significant public resources toward themselves, their family, close friends, or political allies

12 Muirhead and Rosenblum, “The Ethics of Partisanship.”

13 Weber, “The Profession and Vocation of Politics.”

14 Machiavelli, *The Prince*.

to the significant disadvantage of the public. Whatever ethically incompetent politicians might do, they are badly intentioned, and they frequently advocate for harmful policies in cynical campaigns.

It is not always very easy for conscientious politicians to win votes during elections. Good politicians can lose with bad campaigns. Representative democracy has selection effects. In other words, elections select party politicians who are able to gain the alliances and votes necessary to win the next election. Those able to do whatever is necessary to win gain a competitive advantage over those who are unable. Representative democracy also has treatment effects. In other words, elections induce a willingness in politicians to win the next election by any means necessary.<sup>15</sup> They must become willing to cultivate the fragile electoral support and internal alliances within their divided political parties and the fragile electoral support and external alliances with sympathetic voters and politicians across party lines. Those willing to win by any means necessary gain a competitive advantage over the unwilling. The circumstances of politics shape the behavior of conscientious politicians. As explored next, a conscientious politician must cultivate a pragmatic type of sensibility toward how to win the next election.

In order to win critical votes, a conscientious politician must become sensitive to the opinions of a critical mass of her political alliances and electoral support, however ignorant, misinformed, or irrational they might be. In practice, a generally competent public will still contain many ignorant, misinformed, irrational, and otherwise incompetent people, and generally competent people will still have particular knowledge gaps, particular false and irrational beliefs, and other particular incompetencies. So a conscientious politician often does not need to promote empirically informed and epistemically rational judgements about how to do good in order to win the next election. She needs only to confirm whichever ignorant, misinformed, and irrational opinions a critical mass of her political alliances and electoral support accept. Worse, a conscientious politician must occasionally avoid empirically informed and epistemically rational judgements about how to do good in order to win the next election. If she constantly contradicts the incompetent opinions of a critical mass of her political alliances and electoral support, she risks losing the next election. Whatever electoral strategies may win, a conscientious politician cannot consistently prioritize the truth over vote accumulation if she wishes to remain a politician.

It is implausible to presume that there are no conscientious politicians in the legislature and that the bad apples comprise a majority. However, the

15 I assume politicians should continue to obey the law.

legislature potentially lacks a critical mass of conscientious politicians and potentially has a critical mass of bad apples. Too few politicians are Goodin/Spiekermann-style delegates or are willing to participate in Landemore's wisdom-inducing feedback loop. So there are potentially too few conscientious politicians in the legislature and too many bad apples for the epistemic benefits of widespread public competence to spill into the legislature.

## 2. THE BAD APPLES

With the pragmatic sensibilities of party politicians who are focused on winning the next election in the background, I will defend the profession-specific obligation of politicians to protect innocent people from the bad apples in the legislature. Politicians are not purely self-interested vote grabbers. As fellow humans, politicians contain a similarly complex bundle of self-interested and public-spirited motivations as everybody else. However, despite their similar motivations, it has been long recognized that the will of the voters and the will of the politicians frequently differ.<sup>16</sup> Rather than judge the different wills of politicians harshly, I will argue that their wills *should* differ. Politicians should become sensitive to their profession-specific obligations, even if they might be significantly different from the ethical obligations of voters. As an elected representative, a politician can and should cultivate a professional type of sensibility toward how she can protect innocent people from the bad apples in the legislature.

What should conscientious politicians do about the bad apples? I will argue that the active possibility of the bad apples winning and wielding political power as they wish has ripple effects across how all politicians should behave. The circumstances of politics should shape the ethical obligations of politicians. So politicians competing for votes also should significantly blunt the effects of widespread public competence for principled reasons. Perhaps enough conscientious politicians win elections for the epistemic benefits of widespread public competence to potentially spill into the legislature. Nevertheless, politicians are ethically obliged to make it harder for the bad apples in the legislature to significantly harm innocent people. As a consequence, a conscientious politician should prioritize electoral competence over ethical and epistemic competence. Whoever she believes the bad apples are, she should typically prioritize whatever it takes to gain the internal and external political alliances and the electoral support necessary to resist them, even if she must compromise

<sup>16</sup> Rousseau, *The Social Contract and Other Later Political Writings*, 57.

the ability of the epistemic benefits of widespread public competence to spill into the legislature to do it.<sup>17</sup>

Epistemic democrats wish representative democracies to promote the truth. Indeed, the truth is a highly valuable good. Nevertheless, politicians have many weighty ethical obligations, and some of them might conflict with their ethical obligation to promote the truth. No politician is ethically obliged to promote the truth regardless of the consequences. Following political philosopher William Galston, the first priority of politics is to avoid the worst.<sup>18</sup> The public does not need to gain the best outcomes in order to live lives they consider good, but they must avoid the worst outcomes to live lives they consider good. They must avoid civil war, famine, economic collapse, and comparable catastrophes to live good lives. In order to avoid the worst outcomes, party politicians are ethically obliged to resist the bad apples, whether the bad apples are inside or outside their political parties. So, epistemic democrats risk being too insensitive to the conflicting obligations of politicians. A persistent ethical obligation of politicians is to promote whatever moral compromises are necessary to make the bad apples significantly less powerful and to avoid whatever uncompromising truths might fail to gain critical alliances and votes. Among the many weighty reasons to promote moral compromises in politics, the ethical obligation to resist the bad apples is a particularly powerful reason. The ethical obligation to resist the bad apples provides a particularly powerful reason to prioritize whichever compromises will gain the alliances and votes necessary to resist the bad apples and avoid whichever compromises may compromise those compromises.

### 2.1. *Thou Shalt Not Lie!*

Politicians often lie—and frequently out of moral vice rather than for some greater good. However, epistocrat Jason Brennan has shown that it is not always wrong for politicians to lie.<sup>19</sup> It is plausible to presume that lying is generally wrong, but an absolute prohibition against lying is highly implausible. In particular circumstances, it is not wrong to lie. If there is a known murderer at your door, and she asks if you are hiding your neighbor in your house, it is morally permissible to lie to the murderer in order to protect your hidden neighbor. It is not wrong to lie to murderers at the door. In his characteristically colorful style, Brennan has argued by analogy that if there are murderers at the ballot box and they will knowingly vote for badly intentioned policies that will directly kill

17 Whether conscientious politicians should break the law to resist the bad apples exceeds the scope of this paper.

18 Galston, “Realism in Political Theory,” 394.

19 Brennan, “Murderers at the Ballot Box.”

many innocent people, it is morally permissible for politicians to lie to them. In order to save innocent lives, it is morally permissible for politicians to lie about their support for the killer policies in order to win the election and then to oppose the killer policies when in office.

Similarly, Brennan has argued by analogy that if there are ignorant voters at the ballot box and they will unknowingly vote for well-intentioned policies that will nevertheless significantly harm many innocent people, it is still morally permissible for politicians to lie to them. If a conscientious politician honestly and openly opposes the harmful policies during elections, she may risk losing the election to an honest supporter of the harmful policies. In this scenario, her honest choice is a bad choice. The honest choice passively contributes to the harm of many innocent people: the conscientious politician can avoid that harm if she lies during the election. The dishonest choice is the better choice in this case: it actively contributes to avoiding significant harm. The conscientious politician should lie about her support for harmful policies during the election in order to win the votes and then should oppose the policies when in office in order to avoid the harm. She could even lie about her opposition to the harmful policies when in office and preach that circumstances have made her support for the harmful policies ineffective or infeasible. This shows that it is not always wrong for a politician to lie to ignorant voters at the ballot box. If anything, politicians are ethically obliged to lie to ignorant voters at the ballot box if lying is likely to avoid significant harm.

In a different direction, there are other ethical reasons for politicians to compromise the truth in politics. Rather than politicians merely reflecting or reacting to the moral and epistemic vices of the voters, politicians themselves have particular moral and epistemic vices to which the more conscientious among them should react. Suppose there are murderers on the ballot paper, and they knowingly advocate for badly intentioned policies that will directly kill many innocent people. In that case, it is morally permissible for a conscientious politician to preach whatever compromised truths will gain the political alliances and electoral support necessary to resist the murderers on the ballot paper and to save innocent lives.

Similarly, suppose there are ideological politicians on the ballot paper, and they unknowingly advocate for well-intentioned policies that will nevertheless significantly harm many innocent people. The ideologues put the lives, liberties, and happiness of hundreds, thousands, and millions of people at risk with their irresponsible misuse of state power in wars, policing, the courts, prisons, and elsewhere. The ideologues could intensely support wars of aggression with violent blowback, militarized policing, harsh sentences for nonviolent crimes, and do little about prison violence. Alternatively, the ideologues may intensely

oppose defensive wars, defund policing, support soft sentences for violent crimes, and wish to abolish prisons. Whatever the ideologues might wish to do with political power, it is morally permissible for a conscientious politician to preach whatever compromised truths will gain the political alliances and electoral support necessary to resist the ideologues and thereby avoid significant harm.

If a conscientious politician were to honestly and openly support the uncompromised truth during elections, she may risk losing the alliances and votes necessary to resist the ideologues. In other words, the uncompromised truth is a bad choice: choosing the uncompromised truth passively contributes to the empowerment of the ideologues that the conscientious politician may avoid if only she compromises on the truth to gain critical alliances and votes. The conscientious politician should compromise on the truth in order to gain the alliances and votes necessary to resist the ideologues. So it is not always wrong for a politician to compromise on the truth when ideologues are on the ballot paper. Compromising on the truth is the better choice if it actively contributes to resisting the ideologues. If anything, politicians are ethically obliged to compromise on the truth when ideologues are on the ballot paper in order to gain the alliances and votes necessary to resist the ideologues.

Whatever the epistemic democratic analysis of representative democracy might be, epistemic democrats should become much more sensitive to the fact that party politicians have many more ethical obligations beyond the ethical obligation to promote the truth.<sup>20</sup> Even if diverse deliberations and vote aggregation during elections do filter out many of the bad apples from the legislature, not all of the bad apples are filtered out. There are potentially enough partisans, careerists, Machiavellians, cult leaders, corporatists, and ideologues on the ballot paper to allow very harmful policies to succeed, whichever political party might be in government. So a politician should prioritize her ethical obligation to protect innocent people from the bad apples even if she must compromise the epistemic benefits of widespread public competence in order to gain critical alliances and votes.

If and when the ethical obligation to resist the bad apples overpowers the ethical obligation to promote the truth, many if not most of the epistemic benefits of widespread public competence will not spill into the legislature. To

20 Patrick Grim et al. have provided an extensive empirical analysis of epistemic democracy that shows that the epistemic benefits of widespread public competence can survive in representative democracies (“Representation in Models of Epistemic Democracy”). Whatever the plausibility of their analysis might be, it simply does not consider whether profession-specific obligations do or should significantly compromise the ethical and epistemic competence of politicians.

compromise the epistemic benefits of widespread public competence is bad, but to allow the bad apples to do significant harm to many innocent people is worse. So the active possibility of the bad apples winning and wielding political power as they wish is enough to ethically oblige conscientious politicians to compromise the epistemic benefits of widespread public competence in order to gain the alliances and votes necessary to resist the bad apples. As a consequence, epistemic democrats risk an unrealistic model of politicians since the active possibility of bad apples in the legislature should motivate conscientious politicians to spoil the epistemic benefits of widespread public competence for the rest of us.

## 2.2. *Won't Somebody Please Think of the Common Good?*

The avoidance of the worst is not the only consideration in political decision making. The promotion of the common good is also a central consideration in political decision making. Brennan has argued that the political power of democratic citizens voting in elections should become conditional on political competence and that political competence depends on the ability to promote the common good.<sup>21</sup> In other words, the right to vote should depend on the ability to promote the common good. Similarly, perhaps the political power of party politicians in office should also become conditional on the ability to promote the common good. So, if a politician compromises on promoting the common good in order to resist the bad apples, she may become too politically incompetent to hold political office. In the opposite direction, I will argue that if a politician compromises on the obligation to resist the bad apples in order to promote the common good instead, she might become too politically incompetent to hold political office. As explored next, the ethical obligations of the average voter and of the average politician should differ in light of the different stakes involved in their political decisions.<sup>22</sup> The terms and conditions for political competence should become sensitive to the different obligations of the different participants within the democratic process.

The average voter is typically only one out of millions in the electorate. For example, if fifty or so voters out of the fifty million or so voters in an electorate voted for controversial public spending or tax cuts that they judge necessary to

21 Brennan, *Against Democracy*, 144–47.

22 It is generally recognized that the stakes of political decisions affect political behavior. Lomasky and Brennan, “Is There a Duty to Vote?”; Pincione and Tesón, *Rational Choice and Democratic Deliberation*; Hamlin and Jennings, “Expressive Political Behaviour”; Somin, *Democracy and Political Ignorance*; and Brennan, *Against Democracy*. In a similar spirit, the different stakes should affect the ethical obligations of the different participants in the political process.

promote the common good, whatever the unintended political consequences might be, they would not make much if any difference to who wins, how they win, or how they govern. They are only 0.00001 percent of the electorate. Even if five hundred, five thousand, or fifty thousand voters voted for the divisive policies necessary to promote the common good, they would still remain only 0.00001 percent, 0.0001 percent, or 0.001 percent of the electorate. So perhaps it is not unreasonable to expect competent voters to prioritize the controversial policies necessary to promote the common good, despite the active possibility of bad apples in the legislature, because the average voter has very limited political power in the political process. As a consequence, the ability to know how to promote the common good may remain central to the political competence of the average voter.

Conversely, the average politician is typically one out of only hundreds in a legislature. For example, if fifty or so politicians out of the five hundred or so politicians in a legislature advocated for the divisive public spending or tax cuts they judge are necessary to promote the common good, whatever the unintended political consequences might be, they risk failing to gain the political alliances and electoral support necessary to resist the bad apples. They are approximately 10 percent of the legislature. Even if only five politicians uncompromisingly advocated for the common good, they still remain a potentially critical 1 percent of the legislature. So the particular type of political power the average politician has in the political process makes it unreasonable to expect party politicians to prioritize the common good. The average politician is much more able to resist the bad apples than the average voter. With that profession-specific power comes the profession-specific responsibility to resist the bad apples. Consequently, epistemic democrats in particular and political philosophers more generally should become much more sensitive to the fact that politicians bear a profession-specific obligation to resist the bad apples even if they must compromise on the correct but controversial policy decisions necessary to promote the common good in the process.

Independently of how well democracy can promote the common good compared to antidemocratic alternatives, political theorizing should also focus on how well democracy can resist the bad apples compared to antidemocratic alternatives. The moral compromises fundamental to the fragile alliances and support of conscientious politicians can empower them to limit significantly how much harm the bad apples inside or outside their political parties can do. If conscientious politicians gain critical alliances and votes through moral compromise, they can take significant alliances and votes away from the bad apples. So a potential democratic tradeoff is that moral compromises may limit the advocacy of conscientious politicians for the divisive decisions

necessary to promote the common good, but those compromises can also empower them to limit significantly the advocacy of the bad apples for the worst policies available.

### 2.3. *Democratic Integrity*

The external consequences of an action are not the only considerations in political decision making. Following political philosopher Bernard Williams, consequentialist calculuses neglect the moral value of personal integrity.<sup>23</sup> The internal commitments of the agent are also central considerations in political decision making. A minimum-integrity politics is unattractive. If a party politician does whatever she expects to promote better consequences regardless of her principled commitments, she puts the public at risk of very unprincipled behavior in order to gain slightly greater goods, which she is not certain of gaining in return. However, a maximum-integrity politics is also unattractive. If a politician protects her personal integrity regardless of the external consequences, she puts the public at risk of great wrongs in order to avoid the significantly lesser wrong of compromising her personal integrity. Williams therefore defended a medium-integrity politics: integrity must not be valued too little nor be valued too much. When a conscientious politician judges that circumstances compel her to compromise on her principled commitments, she should express a sincere sense of guilt afterward in order to reassure the public that she recognizes the wrong she has done and that she did not and will not compromise her principled commitments unless she is confident it will avoid a significantly greater wrong in return.

As explored next, a conscientious politician should compromise her principled commitments to avoid the significantly greater wrong of allowing the bad apples to win and wield political power as they wish. Political ethicist Edward Hall has argued that party politicians typically acquire competing ethical obligations as moral agents and as political advocates for the interests and values of those they represent.<sup>24</sup> When these competing ethical obligations conflict, politicians are forced to get dirty hands. When the circumstances of politics force a conscientious politician to choose between two wrongs, she should do the significantly lesser wrong to avoid the significantly greater wrong. Nevertheless, the politician does get dirty hands: she is still morally guilty of doing wrong even if it is the better choice all things considered. However, the politician would not have remained morally innocent if she had not done the significantly lesser wrong. She still would have gotten dirty hands but for a different

23 Williams, "A Critique of Utilitarianism."

24 Hall, "Political Compromise and Dirty Hands," 228.

reason. In that case, the politician would have been guilty of failing to avoid the significantly greater wrong. Inaction is not always morally on par with action, but inaction is not always morally innocent, especially if an action would have avoided a significantly greater wrong.

Following Hall, it is useful to recall sociologist Max Weber's distinction between an ethics of conviction and an ethics of responsibility.<sup>25</sup> A conviction politician does not accept a high level of responsibility for the unintended but foreseeably bad consequences of acting on her good convictions. A conviction politician, as a moral agent, prioritizes her personal integrity and promotes her moral convictions. She may therefore be willing to act on her good convictions even when she risks unintended but foreseeably bad consequences. Similarly, a "values advocate" typically prioritizes the moral values that she judges or that those she represents judge are correct, even if she must neglect her ethical obligation to advocate for the material interests of those she represents as a consequence.

In a different direction, a responsible politician accepts a high level of responsibility for the unintended but foreseeably bad consequences of acting on her good convictions. A responsible politician, as an elected representative, prioritizes a different type of integrity—her democratic integrity—and takes responsibility for protecting the material interests of those she represents. Independently of whatever unacquired ethical obligations people as moral agents might have, politicians as elected representatives acquire a professional obligation to protect those they represent.<sup>26</sup> So a responsible politician is willing to act against her good convictions to avoid unintended but foreseeably bad consequences. Similarly, an "interests advocate" typically prioritizes her professional obligation to advocate for the material interests of those she represents even if she must compromise the moral values she or those she represents judge are correct in the process.

Unfortunately, moral compromises often do compromise personal integrity.<sup>27</sup> However, to see only that moral compromises compromise personal integrity is to overlook whether moral compromises may cultivate a different type of integrity. In particular, an interests advocate cultivates her *democratic integrity* as an elected representative. An interests advocate protects those she represents from the bad apples, even if she must compromise her personal integrity as a moral agent in the process of gaining critical political alliances and electoral

25 Weber, "The Profession and Vocation of Politics."

26 A similar type of view is expressed in Mark Philp's 2014 Report for the Committee on Standards in Public Life ("Public Ethics and Political Judgment").

27 Leopra, "On Compromise and Being Compromised"; and Leopra and Goodin, *On Complicity and Compromise*.

support. The moral good of cultivating democratic integrity potentially compensates for the moral bad of compromising personal integrity. Conversely, a values advocate compromises her democratic integrity to conserve her personal integrity. A values advocate promotes those values she judges or those she represents judge are correct, even if she must compromise her democratic integrity as an elected representative and fails to protect those she represents from the bad apples as a consequence.<sup>28</sup> The moral bad of compromising democratic integrity potentially taints the moral good of conserving personal integrity.

Hall has argued that a conscientious politician primarily aims to promote her principled commitments as effectively as her circumstances permit.<sup>29</sup> I will argue that the active possibility of the bad apples winning and wielding political power as they wish unfortunately means that circumstances rarely permit a conscientious politician to promote her principled commitments very effectively. To not resist the bad apples is grossly negligent. If the bad apples win and wield political power as they wish, the risk is that the material interests of the innocent people conscientious politicians represent will be harmed by some of the worst policies available. So a conscientious politician should compromise the moral values that she judges or that those she represents judge are correct and deploy whatever rhetoric, lies, and bullshit will gain the political alliances and electoral support necessary to resist the bad apples. Rather than become Goodin and Spiekermann's competent delegate or some type of uncompromising values advocate, the conscientious politician should become a compromising interests advocate. As an interests advocate, a conscientious politician is guilty of compromising the moral values that she judges or that those she represents judge are correct. However, she would not have remained morally innocent as a values advocate. If she does not compromise those moral values to resist the bad apples, she is guilty of failing to advocate for the material interests of those she represents.

Epistemic democrats might argue that the decision to do the significantly lesser wrong is the correct decision.<sup>30</sup> To compromise the otherwise correct decision in order to resist the bad apples is itself the correct decision in those circumstances. However, in a dirty-hands choice between two wrongs, there is no right. As dirty-hands theorists argue, there is something morally good about avoiding the significantly greater wrong, but there remains something morally bad about doing the lesser wrong.<sup>31</sup> In a dirty-hands choice, decisions

28 Dovi, *The Good Representative*, 164; Philp, "What Is to Be Done?" 479; and Hall, "Political Compromise and Dirty Hands," 221.

29 Hall, "Integrity in Democratic Politics."

30 Estlund, *Democratic Authority*, 163. This type of view is also extensively defended in Nielson, "There Is No Dilemma of Dirty Hands."

31 Hall, "Political Compromise and Dirty Hands," 217.

that could count as correct are not feasible. The only feasible decisions are two wrong decisions. To see doing the lesser wrong as the correct decision is to overlook the deep residual moral bad still fully present in the lesser wrong.

Epistemic democrats should expect that more than enough party politicians potentially spoil the epistemic benefits of widespread public competence. Because of the circumstances of politics, competent voters are prone to vote for electorally competent but ethically and epistemically incompetent politicians. First, competent voters may vote for bad apples. They mistake electoral competence for ethical and epistemic competence. Bad apples may appear principled and knowledgeable during campaigns; the circumstances of politics thus reward the electoral competence of bad apples.

Second, competent voters might vote for a conscientious politician. The conscientious politician is principled and knowledgeable. However, a conscientious politician should prioritize her electoral competence even if she must compromise her ethical and epistemic competence in the process. She should compromise the moral values that she judges or that those she represents judge are correct in order to gain the alliances and votes necessary to protect the material interests of those she represents from the bad apples. In order to resist the bad apples, the conscientious politician may become guilty of failing to respect particular truths, failing to do particular goods, and failing to avoid particular harms. The circumstances of politics force her to do the significantly lesser wrong of compromising her ethical and epistemic competence in order to avoid the significantly greater wrong of allowing the bad apples to win and to wield political power as they wish.

Third, competent voters could vote for a conscientious politician who is unwilling to prioritize electoral competence—in other words, a politician who is uncompromisingly principled. However, the uncompromising politician simply risks losing the next election. The circumstances of politics thus punish the electoral incompetence of a conscientious politician. So she would therefore be guilty of the significantly greater wrong of failing to gain the political alliances and electoral support necessary to resist the bad apples.

It is plausible to presume that there are some uncompromising politicians in the legislature. However, the legislature potentially contains a critical mass of compromising politicians willing to resist the bad apples. So there are potentially too many compromising politicians in the legislature for the epistemic benefits of widespread public competence to spill into the legislature. Whatever the composition of large legislatures might be, the bad apples problem shows that current epistemic democratic aspirations are potentially defective. As elected representatives, party politicians should not primarily aspire to reap the epistemic benefits of widespread public competence. As elected

representatives, politicians should primarily aspire to resist the bad apples, whatever the consequences for the correct decisions might be.

### 3. DELIBERATED COMPROMISES

As explored above, the principled reasons for compromise extend beyond the typical reasons of reciprocity, inclusion, and mutual respect.<sup>32</sup> As elected representatives, party politicians have a profession-specific obligation to protect the material interests of those they represent. So a politician can and should compromise the moral values that she judges or that those she represents judge are correct in order to resist the bad apples. However, politicians need a mechanism that would inform them of which moral compromises will help them gain the political alliances and electoral support necessary to resist the bad apples. In a different direction from epistemic democracy, the next core contribution of this paper is to show that a potential epistemic value of public deliberation in representative democracies is that politicians may use deliberation to inform themselves about which moral compromises to advocate for. Rather than participate in Goodin and Spiekermann's competence-inducing deliberation or Landemore's wisdom-inducing feedback loop, conscientious politicians can and should participate in a compromise-discovering type of deliberation. The epistemic value of public deliberation in representative democracies with incompetent politicians may look very different from that of direct democracies with competent voters.

In between the political ideal of public consensus and the political reality of state coercion is moral compromise.<sup>33</sup> A consensus typically discovers *common* ground that two parties share. It contains principles both parties already accept whatever else they accept. Following political philosopher John Rawls, an overlapping consensus contains moral principles all reasonable people accept for moral reasons.<sup>34</sup> In contrast, a compromise typically discovers a *middle* ground that is close enough to the two parties and not too distant from either political party. It contains principles neither party already accepts but that both parties will accept. A *modus vivendi* conception of compromise contains moral principles that a critical mass of people (reasonable or otherwise) will accept for

32 Gutmann and Thompson, *The Spirit of Compromise*; Weinstock, "On the Possibility of Principled Moral Compromise"; and Wendt, *Compromise, Peace and Public Justification*.

33 Bellamy, Kornprobst, and Reh, "Introduction"; and Spang, "Compromise in Political Theory."

34 Rawls, *Political Liberalism*, 147–48.

pragmatic reasons.<sup>35</sup> A Rawlsian overlapping consensus and a *modus vivendi* compromise are second-best agreements but for different reasons.<sup>36</sup> Nobody accepts them as the correct conception of justice. A Rawlsian overlapping consensus is a second-best agreement because it contains only those moral principles all reasonable people accept for moral reasons. So a consensus second best can look quite similar to the correct first best. Conversely, a *modus vivendi* compromise is a second-best agreement because it primarily contains those moral principles a critical mass of people will accept for pragmatic reasons. So a compromise second best can look very different from the correct first best.

Nobody accepts moral compromises because they judge that they provide the correct decision. Everybody accepts moral compromises because they recognize that people disagree over which decisions are correct. Nobody accepts moral compromises because they are coerced. Everybody accepts moral compromises as second-best or third-rate agreements that forgo the correct decision to avoid an even worse outcome. In private life, people typically commit to conflicting values that must compete against and compromise with each other: internal moral compromise is a common characteristic of private life. Similarly, in public life, radically diverse people typically commit to conflicting values that must compete against and compromise with each other.<sup>37</sup> In other words, interpersonal moral compromise is a familiar feature of public life. *Modus vivendi* compromises do not righteously aim to promote the correct religious, moral, or political values since all of those values are deeply controversial. *Modus vivendi* compromises realistically aim to promote those few common interests most if not all members of a radically diverse political community share. They aim to avoid violent conflict, preserve a peaceful coexistence, and cultivate productive cooperation in a political community with radically divergent religious, moral, and political values.

Political philosophers typically see deliberation as aiming at consensus and see compromise as the product of negotiation.<sup>38</sup> In a different direction, the next core contribution of this paper is to show that deliberation is potentially a *compromise-discovery process*.<sup>39</sup>

35 Gray, *Two Faces of Liberalism*; Horton, "Realism, Liberal Moralism and a Political Theory of Modus Vivendi"; and McCabe, *Modus Vivendi Liberalism*.

36 Vallier, "On Distinguishing Publicly Justified Polities from *Modus Vivendi* Regimes," 219–21.

37 Hall, *Value, Conflict, and Order*, 71–89.

38 Van Parijs, "What Makes a Good Compromise?"; and May, "Compromise in Negotiation."

39 Jones and O'Flynn, "Can a Compromise Be Fair?"; Warren and Mansbridge, "Deliberative Negotiation"; Weinstock, "Compromise, Pluralism, and Deliberation"; and Spang, "Why a Fair Compromise Requires Deliberation."

It is helpful to distinguish between what political theorist Richard Bellamy has called “bargained compromises” and what I call “deliberated compromises.”<sup>40</sup> A *bargained* compromise is primarily self-interested: two parties promote a middle ground to advance their narrow individual and group interests as effectively as possible given their opposing interests. Conversely, a *deliberated* compromise is primarily public-spirited: two parties promote a middle ground to promote their political, moral, or religious values as effectively as possible given their opposing values.<sup>41</sup> A deliberated compromise foregrounds opposing *values* rather than opposing *interests*.<sup>42</sup>

Rather than participate in Landmore’s wisdom-inducing feedback loop, conscientious politicians can and should exploit the feedback loop between the public’s policy preferences and the politicians’ policy decisions to discover which deliberated compromises will empower them to resist the bad apples. Deliberation allows a politician to revise and refine which values she is willing to compromise on in light of what the voters are willing to compromise on. In return, deliberation allows the voters to revise and refine which values they are willing to compromise on in light of what politicians are willing to compromise on. A politician must follow public opinion when judging which deliberated compromises are electorally feasible. However, a politician can also lead public opinion on which deliberated compromises she judges are necessary to resist the bad apples. This shows that the dynamic between public opinion and the policy decisions of politicians is much more interactive and complex than that of vote-hungry politicians blindly following an ignorant public. Party politicians blunt not only the epistemic benefits of widespread public competence but also the epistemic significance of widespread public ignorance. If a politician is less sensitive to public opinion and more sensitive to her profession-specific obligations, public ignorance becomes less of a problem. As explored above, politicians are ethically obligated to protect those they represent. Even if the voters are too ignorant to know how to promote the common good by themselves, they may remain competent enough to help politicians protect their material interests from the bad apples.<sup>43</sup>

40 Bellamy, *Liberalism and Pluralism*. Bellamy has also explored “trimmed” and “segregated” types of compromise, which exceed the scope of this paper.

41 Bellamy prefers what he has called a “negotiated compromise”: they aim to acquire the reciprocal accommodation of opposing interests and values. Presumably, reciprocal accommodation aims to promote the negotiators’ values rather than advance their interests. So negotiated compromises are a special type of deliberated compromise.

42 Benditt, “Compromising Interests and Principles.”

43 Political scientist Gerry Mackie has argued that voters are competent enough to contribute to the mandates of party politicians. Mackie, “Rational Ignorance and Beyond.”

As explored next, informative public deliberation allows party politicians to know which moral compromises to advocate for in order to win critical votes during elections and to gain critical alliances when in office.<sup>44</sup> First, elections incentivize politicians to become willing to seek the vote of the median voter in order to maximize their share of the vote.<sup>45</sup> The median voter provides imperfect protection against polarized political sentiments. The median voter prefers mildly good policies that most do not judge are the best but most do not judge are the worst. However, a politician still needs a mechanism that would inform her of the policy preferences of the median voter. I will show that politicians are able to know the policy preferences of the median voter with informative public deliberation. Deliberation reveals the political judgements of the voters. It makes political judgements publicly known and encourages voters to justify their political judgements to each other in light of opposing judgements.<sup>46</sup> So deliberation can empower politicians to persistently discover the diverse and dynamic political judgements among the voters and to infer the policy preferences of the median voter; without it, politicians are left mostly in the dark about the complex and constantly changing political judgements of the voters. Deliberation can empower politicians to persistently infer which mildly good policies most do not judge are the best nor the worst in order to gain the political alliances and electoral support necessary to resist the bad apples.

For example, deliberation can empower party politicians to persistently infer a level of income redistribution that neither progressive liberals nor market liberals judge is the best nor the worst. Economist Dan Usher has argued that the median voter supports some level of income redistribution: “self-interest can be relied upon in voting about the redistribution of income, narrowing the gap between rich and poor, without removing the gap completely, altering people’s ordering on the scale of rich and poor or destroying incentives to work and save.”<sup>47</sup> However, politicians still need a mechanism to *know* the level of income redistribution the median voter supports. As explored next, politicians are able to know the level of income redistribution the median voter supports with informative public deliberation.

44 Politicians can also use deliberation for a variety of other purposes. In particular, it remains possible that bad apples will use deliberation more effectively than conscientious politicians and that deliberation will reveal moral compromises that empower bad apples rather than help conscientious politicians resist them. This possibility exceeds the scope of this paper.

45 Black, “On the Rationale of Group Decision Making”; and Downs, *An Economic Theory of Democracy*.

46 Elster, “Arguing and Bargaining in Two Constituent Assemblies.”

47 Usher, *The Economics of Voting*, i.

Deliberation makes the acceptable levels of income redistribution among progressive liberals and market liberals known. Progressive liberals typically hold that income redistribution is imperfect: it does not eliminate social inequality. It does not permanently liberate the working class from working under exploitative terms; it only temporarily reduces the exploitation of the working class as the logic of capital accumulation forces the capitalist class to increase the exploitation of the working class over time. Nevertheless, progressive liberals can reveal through deliberation the lowest level of income redistribution they reflectively judge necessary to significantly reduce the economic power of the capitalist class over the working class and consequently reduce the capacity of capital to exploit labor.

Conversely, market liberals typically hold that income redistribution is less than perfect for a different reason: it reduces economic freedom. It is not the case that the working class has nothing to lose from income redistribution: they risk losing the social benefits of economic freedom. In particular, redistribution diminishes the profit incentive that encourages entrepreneurs to invest in productive, technological, and scientific innovations. Redistribution forces the working class to forgo the better and cheaper consumer goods and services that entrepreneurial innovations produce over time. Nevertheless, market liberals can reveal through deliberation the highest level of income redistribution they reflectively judge possible to still significantly preserve the profit incentive and consequently preserve the capacity of entrepreneurial innovation to produce better consumer goods for the working class over time.

Deliberation is a dynamic discovery process that persistently allows party politicians to know the complex and changing levels of income redistribution acceptable among progressive liberals and market liberals. Deliberation can empower politicians to persistently infer an acceptable level of income redistribution that progressive liberals reflectively judge will prevent the worst consequences of social inequality and market liberals reflectively judge will preserve the best consequences of economic freedom. More generally, deliberation allows politicians to discover which mildly good compromises to advocate for in order to gain the political alliances and electoral support necessary to resist the bad apples.

Second, vote trading can empower party politicians to avoid policies that a majority of voters mildly support but a minority of voters intensely oppose.<sup>48</sup> Vote trading provides imperfect protection against the tyranny of the majority. Electoral minorities can vote for minority parties or minority members

48 Coleman, "The Possibility of a Social Welfare Function"; and Tullock, "Problems of Majority Voting," 51–53.

within majority parties to advocate for them within or outside of government. Politicians representing electoral minorities can agree to vote for policies that those electoral minorities mildly oppose (or against policies that they mildly support) in return for getting more votes against a policy that those electoral minorities strongly oppose (or for a policy that they strongly support).<sup>49</sup> So vote trading provides imperfect protection against policies that electoral minorities judge are the worst. However, a politician still needs a mechanism that would inform her of the intensity of political sentiments among electoral minorities. As explored above, deliberation is a compromise-discovery process. Deliberation can empower politicians to persistently discover the diverse and dynamic intensities of political sentiments among electoral minorities and to infer which votes to trade; without it, politicians are left mostly in the dark about the complex and constantly changing intensities of political sentiments among electoral minorities. Deliberation can empower politicians to persistently infer which popular policies to oppose in light of mild majority support and intense minority opposition. By persistently discovering what the majority mildly supports and what electoral minorities intensely oppose, deliberation can empower politicians to persistently discover which moral compromises to advocate for to win critical votes during elections and to gain critical alliances when in office in order to resist the bad apples.

#### 4. CONCLUSION

Epistemic democrats typically provide an idealized model of deliberation for direct democracies. However, they provide a potentially unrealistic model of deliberation for representative democracies. Widespread public competence can still yield bad policy outcomes because the choice of good policies is mediated through a choice of bad politicians. Politician incompetence blunts the epistemic benefits of widespread public competence. Epistemic democracy should therefore become much more sensitive to the ordinary incentives of party politicians competing for votes in representative democracies and how they shape the ethical obligations of politicians. A politician has many more ethical obligations than an ethical obligation to promote the truth. In particular, she has a profession-specific obligation to resist the bad apples even if she must compromise on promoting the truth to gain the necessary alliances and votes to do it. A politician should become an interests advocate rather than a values advocate. She should cultivate her democratic integrity as an elected

49 Political philosopher Stuart Hampshire observed that compromise frequently involves both sides of a divide dropping their more minor commitments (*Innocence and Experience*, 154).

representative with a moral responsibility to protect those she represents even if she must compromise her personal integrity as a moral agent with moral convictions in the process. As a consequence, politicians potentially spoil the epistemic benefits of widespread public competence for the rest of us because of the circumstances of politics.

In a different direction, public deliberation may bring other epistemic benefits to representative democracies. Public deliberation provides party politicians with an effective mechanism to know which moral compromises will gain the alliances and votes necessary to resist the bad apples. In light of the constantly changing range and intensity of political sentiments among voters, public deliberation can empower politicians to persistently discover which mildly good compromises a majority mildly supports and most electoral minorities do not intensely oppose. Once party politicians and their obligations are put at the center of political theorizing, epistemic democrats in particular and political philosophers more generally might gain a powerful reason to start modelling public deliberation in representative democracies as a compromise-discovery process that can help conscientious politicians resist the bad apples.<sup>50</sup>

*University of Sheffield*  
r.williams2@sheffield.ac.uk

#### REFERENCES

- Anderson, Elizabeth. "The Epistemology of Democracy." *Episteme* 3, nos. 1–2 (June 2006): 8–22.
- Bellamy, Richard. *Liberalism and Pluralism: Towards a Politics of Compromise*. London: Routledge, 1999.
- Bellamy, Richard, Markus Kornprobst, and Christine Reh. "Introduction: Meeting in the Middle." *Government and Opposition* 47, no. 3 (2012): 275–95.
- Benditt, Theodore. "Compromising Interests and Principles." In *Compromise in Ethics, Law, and Politics*, edited by Roland Pennock and John Chapman, 26–37. New York: New York University Press, 1979.
- Black, Duncan. "On the Rationale of Group Decision Making." *Journal of Political Economy* 56, no. 1 (February 1948): 23–34.
- Brennan, Jason. *Against Democracy*. Princeton: Princeton University Press, 2016.

50 I am very grateful to Dr. Edward Hall, Dr. Matt Sleat, Professor Robert Goodin, Professor Alasdair Cochrane, and two anonymous reviewers for their constructive comments on earlier drafts of this paper. Any errors remain mine.

- . “Murderers at the Ballot Box: When Politicians May Lie to Bad Voters.” In *Ethics in Politics: The Rights and Obligations of Individual Political Agents*, edited by Emily Crookston, David Killoren, and Jonathan Trerise, 11–34. London: Routledge, 2017.
- Caplan, Bryan. *The Myth of the Rational Voter: Why Democracies Choose Bad Policies*. Princeton: Princeton University Press, 2007.
- Cerovac, Ivan. *Epistemic Democracy and Political Legitimacy*. London: Palgrave Macmillan, 2020.
- Cohen, Joshua. “An Epistemic Conception of Democracy.” *Ethics* 97, no. 1 (October 1986): 26–38.
- Coleman, James. “The Possibility of a Social Welfare Function.” *American Economic Review* 56, no. 5 (December 1966): 1105–22.
- Coleman, Jules, and John Ferejohn. “Democracy and Social Choice.” *Ethics* 97, no. 1 (October 1986): 6–25.
- Dovi, Suzanne. *The Good Representative*. Oxford: Wiley-Blackwell, 2007.
- Downs, Anthony. *An Economic Theory of Democracy*. New York: Harper and Row, 1957.
- Elster, Jon. “Arguing and Bargaining in Two Constituent Assemblies.” *University of Pennsylvania Journal of Constitutional Law* 2 (2000): 345–419.
- Estlund, David. *Democratic Authority: A Philosophical Framework*. Princeton: Princeton University Press, 2008.
- Galston, William. “Realism in Political Theory.” *European Journal of Political Theory* 9, no. 4 (October 2010): 385–411.
- Goodin, Robert E., and Kai Spiekermann. *An Epistemic Theory of Democracy*. Oxford: Oxford University Press, 2018.
- Gray, John. *Two Faces of Liberalism*. Cambridge: Polity Press, 2000.
- Grim, Patrick, Aaron Bramso, Daniel J. Singer, William J. Berger, Jiin Jung, and Scott E. Page. “Representation in Models of Epistemic Democracy.” *Episteme* 17, no. 4 (December 2020): 498–518.
- Gutmann, Amy, and Dennis Thompson. *The Spirit of Compromise: Why Governing Demands It and Campaigning Undermines It*. Princeton: Princeton University Press, 2012.
- Hall, Edward. “Integrity in Democratic Politics.” *British Journal of Politics and International Relations* 20, no. 2 (May 2018): 385–408.
- . “Political Compromise and Dirty Hands.” *Review of Politics* 84, no. 2 (Spring 2022): 214–37.
- . *Value, Conflict, and Order: Berlin, Hampshire, Williams, and the Realist Revival in Political Theory*. Chicago: University of Chicago Press, 2020.
- Hamlin, Alan, and Colin Jennings. “Expressive Political Behaviour: Foundations, Scope and Implications.” *British Journal of Political Science* 41, no. 3

- (July 2011): 645–70.
- Hampshire, Stuart. *Innocence and Experience*. Cambridge, MA: Harvard University Press, 1989.
- Horton, John. “Realism, Liberal Moralism and a Political Theory of *Modus Vivendi*.” *European Journal of Political Theory* 9, no. 4 (2010): 431–48.
- Jones, Peter, and Ian O’Flynn. “Can a Compromise Be Fair?” *Politics, Philosophy and Economics* 12, no. 2 (2012): 115–35.
- Landemore, Hélène. *Democratic Reason: Politics, Collective Intelligence, and the Rule of the Many*. Princeton: Princeton University Press, 2012.
- Leopra, Chiara. “On Compromise and Being Compromised.” *Journal of Political Philosophy* 20, no. 1 (March 2012): 1–22.
- Lepora, Chiara, and Robert Goodin. *On Complicity and Compromise*. Oxford: Oxford University Press, 2013.
- Lomasky, Loren, and Geoffrey Brennan. “Is There a Duty to Vote?” *Social Philosophy and Policy* 17, no. 1 (Winter 2000): 62–86.
- Lopez-Guerra, Claudio. *Democracy and Disenfranchisement: The Morality of Electoral Exclusions*. Oxford: Oxford University Press, 2014.
- Machiavelli, Niccolò. *The Prince*. New York: Penguin Books, 1981.
- Mackie, Gerry. “Rational Ignorance and Beyond.” In *Collective Wisdom: Principles and Mechanisms*, edited by Hélène Landemore and Jon Elster, 290–318. Cambridge: Cambridge University Press, 2012.
- May, Simon. “Compromise in Negotiation.” In *Compromise*, edited by Jack Knight, 150–66. New York: New York University Press, 2018.
- McCabe, David. *Modus Vivendi Liberalism: Theory and Practice*. Cambridge: Cambridge University Press, 2010.
- Muirhead, Russell, and Nancy Rosenblum. “The Ethics of Partisanship.” In *Political Ethics: A Handbook*, edited by Edward Hall and Andrew Sabl, 126–47. Princeton: Princeton University Press, 2022.
- Nielson, Kai. “There Is No Dilemma of Dirty Hands.” In *Politics and Morality*, edited by Igor Primoratz, 20–38. Hampshire: Palgrave Macmillan, 2007.
- Philp, Mark. “Public Ethics and Political Judgment.” Report for the Committee on Standards in Public Life (UK), July 2014. [https://www.gov.uk/government/uploads/system/uploads/attachment\\_data/file/336977/2902536\\_CSPL\\_PublicPoliticalEthics\\_acc.pdf](https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/336977/2902536_CSPL_PublicPoliticalEthics_acc.pdf).
- . “What Is to Be Done? Political Theory and Political Realism.” *European Journal of Political Theory* 9, no. 4 (October 2010): 466–84.
- Pincione, Guido, and Fernando R. Tesón. *Rational Choice and Democratic Deliberation: A Theory of Discourse Failure*. Cambridge: Cambridge University Press, 2006.
- Rawls, John. *Political Liberalism*. New York: Columbia University Press, 1993.

- Rousseau, Jean Jacques. *The Social Contract and Other Later Political Writings*. Edited by Victor Gourevitch. Cambridge: Cambridge University Press, 1997.
- Schwartzberg, Melissa. "Epistemic Democracy and Its Challenges." *Annual Review of Political Science* 18, no. 18 (2015): 187–203.
- Somin, Ilya. *Democracy and Political Ignorance: Why Smaller Government Is Smarter*. Redwood City, CA: Stanford University Press, 2013.
- Spang, Friderike. "Compromise in Political Theory." *Political Studies Review* 21, no. 3 (August 2023): 594–607.
- . "Why a Fair Compromise Requires Deliberation." *Journal of Deliberative Democracy* 17, no. 1 (2021): 60–69.
- Tomlin, Patrick. "Should We Be Utopophobes about Democracy in Particular?" *Political Studies Review* 10, no. 1 (January 2012): 36–47.
- Tulloch, Gordon. "Problems of Majority Voting." In *The Selected Works of Gordon Tulloch*. Vol. 1, *Virginia Political Economy*, edited by Charles K. Rowley, 51–61. Indianapolis: Liberty Fund, 2004.
- Urbainati, Nadia. *Representative Democracy: Principles and Genealogy*. Chicago: University of Chicago Press, 2006.
- Usher, Dan. *The Economics of Voting: Studies of Self-Interest, Bargaining, Duty and Rights*. Abingdon: Routledge, 2016.
- Vallier, Kevin. "On Distinguishing Publicly Justified Polities from *Modus Vivendi* Regimes." *Social Theory and Practice* 41, no. 2 (April 2015): 207–29.
- Van Parijs, Philippe. "What Makes a Good Compromise?" *Government and Opposition* 47, no. 3 (2012): 466–80.
- Warren, Mark, and Jane Mansbridge. "Deliberative Negotiation." In *Political Negotiation: A Handbook*, edited by Jane Mansbridge and Cathie Martin, 141–96. Washington, DC: Brookings Institution Press, 2016.
- Weber, Max. "The Profession and Vocation of Politics." Translated by Ronald Speirs. In *Max Weber: Political Writings*, edited by Peter Lassman, 309–69. Cambridge: Cambridge University Press, 1994.
- Weinstock, Daniel. "Compromise, Pluralism, and Deliberation." *Critical Review of International Social and Political Philosophy* 20, no. 5 (2017): 636–55.
- . "On the Possibility of Principled Moral Compromise." *Critical Review of International Social and Political Philosophy* 16, no. 4 (2013): 537–56.
- Wendt, Fabian. *Compromise, Peace and Public Justification: Political Morality beyond Justice*. New York: Palgrave Macmillan, 2016.
- Williams, Bernard. "A Critique of Utilitarianism." In J. J. C. Smart and Bernard Williams, *Utilitarianism: For and Against*, 77–150. Cambridge: Cambridge University Press, 1973.

## THREE KINDS OF PRIORITARIANISM

Carlos Soto

DEREK PARFIT'S 1991 Lindley Lecture "Equality or Priority?" has generated considerable discussion regarding the justification of priority to the worse-off. Parfit argued that prioritarianism provides the most plausible justification of priority to the worse-off. This justification was grounded in a person's lower absolute or nonrelational level of well-being.<sup>1</sup> But there are various ways of understanding prioritarianism.

The most common understanding of prioritarianism present in the literature is axiological. According to axiological prioritarianism, an outcome is better the larger the sum of weighted benefits it contains. The weight or value of a benefit is determined by its size and the absolute level of well-being of potential beneficiaries.<sup>2</sup> Because benefiting people matters more, the worse off these people are, according to Parfit, a smaller benefit for a worse-off person can produce greater moral value and do more to make the outcome better than a larger benefit for a better-off person.<sup>3</sup> This axiology is often coupled with a maximizing version of act consequentialism. According to this combined view, distributive acts are right if and only if they maximize the value of outcomes as described above. I will refer to this combined view as *teleological prioritarianism* or, in Parfit's terms, *telic prioritarianism*.<sup>4</sup>

- 1 I do not offer here an account of well-being or assume any method for its measurement, which may be a limitation of this paper. Furthermore, I will assume—as most other writers on the subject appear to do—that prioritarian judgments are made about a person's overall level of well-being. I will not consider whether prioritarianism should also be applied to particular dimensions of well-being. For the view that prioritarianism should be applied to both, see McKerlie, "Dimensions of Equality."
- 2 Parfit, "Equality or Priority?" 101; McKerlie, "Equality and Priority," 39; and Arneson, "Luck Egalitarianism and Prioritarianism," 343.
- 3 In describing a view as prioritarian, I exclude maximin accounts. Rabinowicz notes that maximin entails giving the *same*—i.e., absolute—priority to the worse-off no matter what their absolute levels or how much they stand to benefit ("Prioritarianism for Prospects," 13). This runs counter to the gradualist conception of prioritarianism in which benefiting people matters more, the worse off these people are.
- 4 Parfit once stated that telic prioritarianism was a view only about the goodness of outcomes ("Equality or Priority?" 101). I do not think that this was Parfit's considered view,

Some philosophers have criticized the fact that telic prioritarianism, as conceived by many of its proponents, applies both interpersonally and intrapersonally. According to Michael Otsuka, Alex Voorhoeve, and Marc Fleurbaey, this unrestricted form of prioritarianism is problematic.<sup>5</sup> Using cases involving risk, they argue that telic prioritarianism is insufficiently sensitive to prudential justifications such as expected utility maximization in one-person cases, thereby violating the unity of the individual, and it is inadequately sensitive to the existence of competing claims, thereby failing to fully respect the separateness of persons.

In response to these criticisms, some authors have developed deontological formulations of prioritarianism. According to deontic prioritarianism, the justification for priority to the worse-off should be grounded in something other than outcome value maximization. For deontic prioritarists, the rightness of distributive acts cannot simply be deduced from axiology. The most prominent versions of deontic prioritarianism all maintain, in one form or another, that people's claims/complaints are what ultimately determines the rightness of acts.

For example, Andrew Williams has developed a form of deontic prioritarianism that does not apply intrapersonally.<sup>6</sup> The impetus for Williams's restriction originates from the contractualist framework within which he operates, a framework that has a long-standing tradition of distinguishing between principles that regulate the distribution of benefits and burdens within lives and across lives. Williams, inspired by Thomas Nagel, argues for a "Nagelian formulation of the Priority View" that grounds priority to the worse-off in a kind of unanimity of outcomes: that is, finding the outcome that is least unacceptable to the person to whom it is most unacceptable, where the acceptability of an outcome is determined in part by how well-off someone is. Grounding deontic prioritarianism in the unanimity of outcomes, Williams claims, provides a means for resolving interpersonal conflicts of competing claims that does not extend to intrapersonal conflict; in intrapersonal cases, one merely has a claim to have one's expected utility maximized, per Williams. By restricting the scope of his version of deontic prioritarianism to interpersonal cases only, Williams avoids the objections levied by Otsuka, Voorhoeve, and Fleurbaey against telic prioritarianism. Matthew Adler and Jacob Nebel likewise espouse restricted

---

since he later described telic prioritarianism as prescribing to agents what they ought to do ("Another Defence of the Priority View," 402).

5 Otsuka and Voorhoeve, "Why It Matters That Some Are Worse Off Than Others" and "Equality versus Priority"; Otsuka, "Prioritarianism and the Separateness of Persons" and "Prioritarianism and the Measure of Utility"; Voorhoeve and Fleurbaey, "Egalitarianism and the Separateness of Persons" and "Equality or Priority for Possible People?"

6 Williams, "The Priority View Bites the Dust?"

versions of prioritarianism based on the idea that moral principles apply only to interpersonal conflict.<sup>7</sup>

Contractualist prioritarianism, however, represents just one version of deontic prioritarianism. The main aim of this paper is to present a case for a noncontractualist version of deontic prioritarianism. It is, according to Parfit, important to understand these distinctions: “Taxonomy, though unexciting, needs to be done. Until we have a clearer view of the alternatives, we cannot hope to decide which view is true, or is the best view.”<sup>8</sup>

The paper is structured as follows. Section 1 assesses several contractualist forms of deontic prioritarianism and argues that they are unsatisfactory. Section 2 argues that telic prioritarian impersonal value is unnecessary and inadequate to fully account for our moral thinking about priority to the worse-off. Section 3 describes one version of noncontractualist deontic prioritarianism and a potential rationale for it. The view is contrasted with contractualist and telic prioritarianism with respect to establishing the moral relevance of absolute levels, the motivation and justification for giving priority to the worse-off, and explaining reactive attitudes. The paper briefly discusses whether the rationale for this view can be developed in ways that also support egalitarianism or hybrid theories. Finally, I give reasons for applying noncontractualist deontic prioritarianism to whole lives as well as parts of lives, and I offer a partial defense against the criticism that this version of prioritarianism appears to be unrestricted. Section 4 concludes.

### 1. CONTRACTUALIST PRIORITARIANISM

Contractualism attempts to justify principles and acts in accordance with some conception of unanimity. Nagel and Williams recognize that when there are conflicting interests, there cannot be complete unanimity regarding outcomes. Nonetheless, they maintain that we should seek to achieve its closest approximation. According to the Nagelian formulation of the Priority View, the relevant unanimity condition consists of finding the outcome that is least unacceptable from an individual point of view. “This means that any other

7 Adler, who also claims to be inspired by Nagel, argues for a version of prioritarianism that is grounded in fairness, and fairness applies only to interpersonal conflict, per Adler, *Well-Being and Fair Distribution*, ch. 5. Adler’s theory, however, can be understood as a version of axiological prioritarianism, which I discuss in section 2. Nebel has proposed a version of deontic prioritarianism to address risky nonidentity cases (“Priority, Not Equality, for Possible People”). Nebel’s account is framed in characteristic contractualist language of minimizing complaints and justifiability to others.

8 Parfit, “Equality or Priority?” 116.

alternative will be more unacceptable to someone than this alternative is to anyone.”<sup>9</sup> But this unanimity condition underdetermines the acceptability of an outcome. The mere goal of achieving unanimity of outcomes does not itself favor prioritarianism over egalitarianism, in which agents are concerned with how individuals fare relative to others.<sup>10</sup> We must assume that absolute levels already matter in order to determine the acceptability of an outcome in Williams’s Nagelian formulation of the Priority View. Yet if we assume that absolute levels already matter, the above unanimity condition presupposes the very feature of deontic prioritarianism in need of justification.

While Nagel’s contractualism emphasizes unanimity of outcomes, Thomas Scanlon’s version of contractualism attempts to find moral principles that no one can reasonably reject, thereby achieving unanimity of moral principles. According to Scanlon’s contractualist formulation of the Priority View, “the worse off people would be if they are not benefited, the stronger their reasons to reject principles that would deprive them of these benefits.”<sup>11</sup> Like the contractualism espoused by Nagel and Williams, Scanlon’s contractualism could be employed to argue for egalitarianism.<sup>12</sup> That either relative or absolute levels of well-being can serve as grounds for reasonable rejection has been suggested by Scanlon himself.<sup>13</sup> But a rationale that explains why absolute levels are morally relevant must be antecedently established in order for there to be a Scanlonian Contractualist Priority View.<sup>14</sup> A major issue here is that there may be ways of establishing this moral relevance that seem to obviate the need for a contractualist framework.<sup>15</sup>

- 9 Nagel, “Equality,” 123.
- 10 Benjamin Lange has also noted that Nagel’s contractualism is compatible with egalitarian and prioritarian readings (“Restricted Prioritarianism or Competing Claims?”).
- 11 Scanlon, “Replies,” 432.
- 12 See, for example, O’Neill, “Constructing a Contractualist Egalitarianism.”
- 13 Scanlon, “Contractualism and Utilitarianism,” 123, and *What We Owe to Each Other*, 226.
- 14 In “Contractualism and Justification,” Scanlon proposes that impersonal values can affect the *reasonableness* of rejection despite not being themselves grounds for rejecting principles, which must be personal. This modification allows consequentialists such as telic prioritarians to claim support from Scanlon’s contractualism.
- 15 Rahul Kumar, a proponent of Scanlonian contractualism, concedes that this theory provides no guidance on determining the relative importance of considerations or how to combine these considerations in order to reach a moral verdict. See “Reasonable Reasons in Contractualist Moral Argument,” 35–36. But contrary to Kumar, Scanlonian contractualism also fails to identify morally relevant considerations. Kumar proposes to use the following purported contractualist commitment as a test of moral relevance: “Can this kind of consideration be considered to be important for being able to live a rationally self-governed, meaningful life?” (“Reasonable Reasons in Contractualist Moral Argument,” 17). If so, then it is relevant for moral argument and cannot be reasonably rejected. However, what is gained

Another serious problem is whether contractualism can successfully support a version of deontic prioritarianism that purports to resolve conflicting claims. This problem is most evident in Parfit's restatement of Scanlon's formulation: "People have stronger moral claims, and stronger grounds to reject some moral principle, the worse off these people are."<sup>16</sup> Parfit did not elaborate on the relation of moral claims to reasonable rejectability. Either moral claims are grounds for the reasonable rejection of principles, or they are products of reasonable rejection.

The notion that moral claims are grounds for reasonable rejection raises several difficulties. First, moral claims provide moral reasons, yet Scanlon's contractualist procedure is supposed to tell us what moral reasons we have rather than presuppose them in the contractualist procedure. Second, moral claims provide agent-neutral reasons, but the grounds for reasonably rejecting principles in Scanlon's system are supposed to be personal or agent-relative. Third, if the worse-off have stronger moral claims prior to the reasonable rejection of principles, then what moral work is done by the notion of reasonable rejectability? The stronger moral claim of the worse-off person appears to be sufficient to settle the matter about what an agent ought to do in conflict cases involving two people. Clearly, strength plays an important and decisive role in contractualism since both Scanlon and Parfit appeal to the strength of reasons in determining what principles can be reasonably rejected. Contractualism, however, does not justify the moral importance of strength, since it depends on this notion to function. So what precludes the strength of claims from being a deciding factor amongst conflicting claims independently of reasonable rejectability?

On the other hand, suppose that moral claims result only from the reasonable rejection of principles. Some of Scanlon's remarks suggest this position, and Frances Kamm has interpreted Scanlon in this way.<sup>17</sup> If so, how can the better-off have competing moral claims? Parfit endorsed the idea that the better-off also have moral claims to a benefit when he introduced Claim Prioritarianism as a version of the Competing Claims View.<sup>18</sup> Suppose that when we ought to

---

by claiming that a consideration cannot be reasonably rejected if it passes this test? It is *already* regarded as morally relevant *in virtue of* its importance for rational self-governance. Kumar's proposal faces a redundancy objection applied to the moral relevance of reasons. For discussion of the redundancy objection as applied to the moral wrongness of acts, see Ridge, "Saving Scanlon" and "Contractualism and the New and Improved Redundancy Objection"; Straton-Lake, "Scanlon's Contractualism and the Redundancy Objection"; and Suikkanen, "Contractualist Replies to the Redundancy Objections."

16 Parfit, *On What Matters*, 2:201.

17 Scanlon, *What We Owe to Each Other*, 169–70; and Kamm, "Owing, Justifying, and Rejecting," 328.

18 Parfit, "Another Defence of the Priority View," 437.

aid the worse-off, the worse-off person, *ex hypothesi*, has a moral claim generated by the reasonable rejection of a principle that would deprive this person of a benefit. The better-off person in this scenario apparently would not have a competing moral claim, since she could not, supposedly, reasonably reject the principle that would deprive her of a benefit by instead directly giving it to the worse-off person. If the worse-off and the better-off both have claims that compete against one another, then these claims seem to be prior to the reasonable rejection of principles. If this is right, then once again we must ask why the strength of claims cannot itself determine what an agent ought to do.<sup>19</sup>

Neither prioritarians nor egalitarians bolster their case by merely appealing to contractualism to defend their views. What is needed to support either of these views appears to lie outside of contractualism.<sup>20</sup>

## 2. TELIC PRIORITARIANISM

Telic prioritarians might think that they can do a better job than contractualists of accounting for the normativity of absolute levels by grounding their importance in the impersonal value of outcomes. I believe this strategy is flawed in several ways.

According to Nils Holtug, equal benefits can lead to states of affairs that differ in intrinsic value.<sup>21</sup> The difference in intrinsic value is not a value for anyone, although what is good for a person contributes to these intrinsic values. However, this impersonal value does not play a crucial role in the justification of priority to the worse-off, according to some remarks made by Parfit. Parfit argued that the concept *good* is not fundamental. When some event or act is described as good for someone or impersonally good, these senses of good have no independent

19 Nagel's system suffers from a similar defect: "Each individual with a more urgent claim has priority ... over each individual with a less urgent claim" ("Equality," 118). According to Nagel, some standard of urgency is necessary to order claims or, specifically, order the various needs and interests that ground claims. This standard of urgency will not be determined by Nagel's unanimity condition since the unanimity of outcomes presupposes this standard. It is the standard of urgency that appears to be doing the normative work of mediating conflicting claims.

20 Shlomi Segall has recently criticized Competing Claims Prioritarianism and Competing Claims Egalitarianism ("Equality or Priority about Competing Claims?"). His critiques differ from mine in several ways. His critiques largely center on considerations that are unique to risky nonidentity cases. Additionally, Segall attacks Nebel's version of Competing Claims Prioritarianism by invoking telic prioritarian impersonal outcome value. In contrast, my arguments criticize contractualism as such, both as a mechanism for resolving competing claims and with respect to establishing the normativity of absolute or relative levels of well-being.

21 Holtug, "Prioritarianism," 132.

normative force. These senses of good are merely briefer ways of signaling that there are other facts that give us reason to perform an act or to want an event to occur.<sup>22</sup> Hence, when telic prioritaricians claim that benefits to the worse-off have greater impersonal value, these claims have no independent normative force. The agent's reasons for action in this context are determined by a potential beneficiary's absolute level of well-being and the size of the benefit that can be provided.<sup>23</sup> It is, according to Parfit, simply the strength of reasons that determines what we ought to do.<sup>24</sup> If these reasons are what justifies giving priority to the worse-off, then appeals to impersonal value appear to be superfluous here.<sup>25</sup>

Parfit, moreover, seemed to abandon the idea that there is a "law" of diminishing marginal moral goodness of utility, which one might expect to apply universally.<sup>26</sup> That Parfit did not regard it as a universal law is implicit in his later discussion of population ethics, where he employed different ideas and principles.<sup>27</sup> Several authors have noted that telic prioritarianism implies the Repugnant Conclusion when applied to variable populations.<sup>28</sup> Roughly, there may be greater moral value in the existence of a large population of people with lives that are barely worth living than the existence of a smaller population of different people with high-quality lives. To my knowledge, Parfit never adequately explained why the aforementioned "law" should be barred from population ethics. It is not enough to claim that we need different principles when dealing with variable populations.<sup>29</sup> For there are intrapersonal analogues of the Repugnant Conclusion that pit quality of life against quantity of life.<sup>30</sup> Telic prioritarianism delivers a Repugnant Conclusion in such cases as well.<sup>31</sup>

22 Parfit, *On What Matters*, 1:39–42.

23 Parfit, "Another Defence of the Priority View," 402.

24 Parfit, *On What Matters*, 1:32–33.

25 Relatedly, when Parfit discussed the interpersonal aggregation of benefits, he sometimes appealed to reasons rather than the goodness of outcomes. Parfit claimed that reasons can be combined to produce a stronger set of reasons to act some way, specifically in a way that would benefit people most (*On What Matters*, 1:32).

26 Parfit mentioned this law in "Equality or Priority?" 106.

27 Parfit, "Can We Avoid the Repugnant Conclusion?" and "Future People, the Non-Identity Problem, and Person-Affecting Principles."

28 Holtug, *Persons, Interests, and Justice*, ch. 9; and Tännsjö, "Why Derek Parfit Had Reasons to Accept the Repugnant Conclusion." For a general discussion of prioritarianism and variable populations, see Brown, "Prioritarianism for Variable Populations."

29 Parfit simply asserted this in "Another Defence of the Priority View," 440.

30 Parfit, *Reasons and Persons*, 498; and Rachels, "Repugnance or Intransitivity."

31 Both Parfit and James Griffin have suggested that global preferences may offer a possible solution to the intrapersonal Repugnant Conclusion. According to Griffin, we cannot arrive at the welfare value of a life by simply totting up the goods and evils the life contains. We

Of course, telic prioritarianism is not the only view that implies the Repugnant Conclusion. Thus, the Repugnant Conclusion, it might be claimed, should not be used to discriminate between distributive theories.<sup>32</sup> If we have strong reason to endorse the law of diminishing marginal moral goodness of utility, then it might be argued that this gives telic prioritarianism an advantage over other distributive theories, and the basis for priority to the worse-off is secured.

However, I see neither a strong reason to endorse this “law” nor a reason to regard impersonal value maximization as a convincing basis for priority to the worse-off. The goal of maximizing impersonal value is not necessary to justify priority to the worse-off, nor does it adequately account for all that is significant in our moral deliberations about aiding the worse-off. When we think about what motivates us to give priority to the worse-off—for those so inclined—the morally salient consideration appears to be not the value of a potential state of affairs but simply the condition or plight of the worse-off. As Hilary Greaves has noted, axiology does not capture the greater sense of urgency and empathetic distress that arises when one contemplates priority to the worse-off.<sup>33</sup> Nor, one might add, does axiology adequately account for certain reactive attitudes. The worse-off have reason to feel indignation or resentment when their plight is ignored, but such attitudes seem to be in the first instance directed at what the agent fails to acknowledge about *them* rather than what the agent fails to produce, for a failure to produce does not in itself

---

should instead defer to a person's preference about the kind of life, taken as a whole, he or she wants to live. In conjunction with regarding such global preferences as basic, Griffin proposes a discontinuity in values such that no amount of one kind of value can outweigh any amount of another kind of value (*Well-Being*, 35–36, 86–88). Yet Griffin's analysis of another case undermines the above response. He rejects a holistic evaluative approach as a basis for preferring a life with overall high quality to a longer life with overall lower quality (*Well-Being*, 355). According to holism, the value of the whole is greater than the sum of the values of parts. When considering whether to live a life of seventy years of good quality or a nearly identical life of eighty years with the last ten years of life of poor quality, Griffin maintains that the ten extra years of low-quality life have positive value and count in favor of choosing the longer life even if the life of seventy years, taken as a whole, seems to be a better life. One might say that the local interest in living the ten extra years of bad life that is still worth living competes with and apparently outweighs the global interest in living a shorter life of only good quality, if Griffin is correct. In Griffin's example, if a global preference for the shorter life of higher overall quality is rational, then rational global preferences need not track greater lifetime well-being, and an agent is not rationally required to maximize well-being within his or her life. If an agent rationally ought to have a global preference for the longer life with overall lower quality, then the intrapersonal Repugnant Conclusion looms.

32 Holtug, *Persons, Interests, and Justice*, ch. 9.

33 Greaves, “Antiprioritarianism,” sec. 5.2.

justify these attitudes. Similar claims apply to an agent's reason to feel remorse or regret for failing to aid the worse-off.

It might be argued that a claims-based axiological prioritarianism can answer these criticisms about motivation and reactive attitudes by appealing to the satisfaction or violation of individual claims. According to this view, the better-off and the worse-off have claims to morally valuable outcomes in which they are benefited.<sup>34</sup> This would link an agent's obligation to maximize the moral value of outcomes with the claims of potential beneficiaries in those outcomes.

But are distributive claims necessary for someone's welfare gain to contribute to the moral value of an outcome according to this model? Suppose that *B* is worse off than *C*, and each person stands to benefit equally from our resource. *B*'s claim to our resource is therefore stronger than *C*'s claim, and satisfaction of *B*'s claim would yield the greatest moral value. Now imagine the same case with the addition of *D*: *D* is as well-off as *C*, and *D* would gain a significant amount of pleasure if *C* is benefited. Perhaps *D* is infatuated with *C* but is indifferent to *B*. Our resource would not in any direct way benefit *D*. If *C* is aided, *D*'s welfare gain apparently would contribute to the moral value of this outcome, and this outcome's moral value, we can suppose, would be greater than the moral value of the outcome in which *B*'s claim is satisfied, which is still the strongest individual claim. It does not seem plausible, however, to maintain that *D* has a claim to what happens to *C* or that *D* himself has a claim to being indirectly benefited.<sup>35</sup> That one's well-being gain might contribute to the most morally valuable outcome, according to the weighted sum of benefits, does not itself render one a claimant to the outcome nor entitle one to feel indignation or resentment when an agent decides not to bring about this outcome. If there is to be perfect alignment between respecting the claims of individuals and maximizing morally valuable outcomes, what justifies the view that only benefits to claimants contribute to the moral value of outcomes, thereby excluding *C*'s potential welfare gain? It is unlikely that axiology can provide the necessary rationale. If so, then the strategy of explaining reactive attitudes and the motivation for priority to the worse-off by linking moral value maximization with respecting individual claims seems to fail.

34 Adler, *Well-Being and Fair Distribution*, ch. 5; Holtug, "Prioritarianism: Ex ante, Ex post, or Factualist Criterion of Rightness?"; and Adler and Holtug, "Prioritarianism."

35 For discussion of indirect benefits and the moral significance of directly needing a resource we have to distribute, see Kamm, *Mortality, Mortality*, 1:106–10.

## 3. NONCONTRACTUALIST DEONTIC PRIORITARIANISM

All of the above versions of prioritarianism take the condition of the worse-off to be a necessary ingredient in the justification of priority to the worse-off. Telic and contractualist prioritaricians believe that something more is needed. They appeal to impersonal value maximization or the satisfaction of some unanimity condition in order to supply what they think is missing: a deeper, systematic justification of priority to the worse-off. However, these theorists simply assume that absolute levels are morally important. The consequentialist and contractualist frameworks do not validate this intuition and are themselves problematic. Can noncontractualist deontic prioritaricians provide a deeper rationale beyond simply asserting that the worse-off have stronger claims to a benefit, and there are stronger reasons to aid them?

Deontologists have often invoked the separateness of persons as an important fact about the lives of persons that ought to govern interpersonal relations. But for reasons enumerated by several authors, this approach to justifying priority to the worse-off is beset with difficulties. Dennis McKerlie argued that even if we concede that the separateness of persons supports the objection to balancing benefits and harms across lives as done within a single life and the objection to aggregating benefits and/or harms across lives—i.e., the objections raised by Nagel and Rawls against utilitarianism—there is no clear path from these objections to priority to the worse-off.<sup>36</sup> The objection to aggregation does not imply priority to the worse-off. The objection to balancing, if it is not to exclude priority to the worse-off, must be interpreted in a way that presupposes its legitimacy. David Brink has raised similar worries about the role assigned to the separateness of persons in justifying priority to the worse-off.<sup>37</sup> And Shlomi Segall has argued that the separateness of persons may not be able to help us decide between distributive theories since it can be interpreted in ways that exclude virtually all of them.<sup>38</sup> It does not appear, then, that the separateness of persons offers support for priority to the worse-off in general and prioritarianism in particular. The fact that people live separate lives does not itself determine what aspects of a person's life, taken separately, is of moral importance.<sup>39</sup>

36 McKerlie, "Egalitarianism and the Separateness of Persons"; Nagel, *The Possibility of Altruism*; and Rawls, *A Theory of Justice*.

37 Brink, "The Separateness of Persons, Distributive Norms, and Moral Theory."

38 Segall, "Sufficientarianism and the Separateness of Persons."

39 Nagel subsequently realized that if the objection to balancing were based on the difference between intrapersonal and interpersonal compensation, then this objection would not defeat utilitarianism, for utilitarians need not deny that there is such a difference, nor are they necessarily guilty of extending the principle of individual choice to the social case,

Despite the lack of support from the doctrine of separateness, we need not see persons as mattering less. To the contrary, noncontractualist deontic prioritarianism affirms the importance of persons and the corresponding moral concern they are owed. We do not show proper concern for persons when we ignore an important facet of their condition, that is, when we treat their absolute level of well-being as an insignificant aspect of their lives. Such disregard expresses that how things are with a person count for nothing.<sup>40</sup> This in turn sends the message that *persons matter less*. A person's absolute level of well-being has normative import because responding to this fact is what is required to value persons appropriately.

The central idea behind noncontractualist deontic prioritarianism is that the moral importance of a person's absolute level of well-being is grounded in the value of the person. This is a natural extension of the idea espoused by several philosophers that a person's well-being matters because the person matters.<sup>41</sup> For if a person's well-being matters because the person matters, then it is reasonable to think that an essential aspect of a person's well-being also matters because the person matters. Ignoring absolute levels in our distributive decisions involves, I submit, a failure to respect the value of persons in virtue of failing to count an aspect of their condition made relevant by this value.<sup>42</sup>

This line of thought finds support in some of Harry Frankfurt's remarks on respect: "Failing to respect someone is a matter of ignoring the relevance of some aspect of his nature or of his situation. The lack of respect consists in the circumstance that some important fact about the person is not properly

---

as they rely on a different fundamental principle ("Equality," 120). Otsuka and Voorhoeve seem to miss or ignore this point ("Equality versus Priority").

40 This phrasing mirrors Harry Frankfurt's remarks in another context. See Frankfurt, "Equality and Respect."

41 See, for example, Anderson, *Value in Ethics and Economics*; Velleman, "A Right of Self-Termination?"; and Darwall, *Welfare and Rational Care*.

42 One anonymous reviewer has questioned whether we must appeal to the special value associated with being a person in the above argument. Perhaps the properties/capacities that constitute personhood can themselves ground the normativity of a person's welfare and absolute level of well-being. Some considerations count in favor of appealing directly to the value of persons. We know that persons are valuable. But there is disagreement about the criteria for personhood and which criteria constitute the value of persons. There is also disagreement about whether these properties/capacities matter in themselves. For example, Jeff McMahan has argued that persons may matter in a special way because they have certain capacities, but it does not follow that they matter because their capacities matter (*The Ethics of Killing*, 479). If McMahan were correct, we could not claim that a person's well-being matters because her capacities matter. These disagreements need not first be resolved in order for the value of persons to play a justificatory role.

attended to or is not taken appropriately into account.”<sup>43</sup> Respecting the value of persons requires acknowledging that how things are with a person is morally important.<sup>44</sup> Denial of the relevance of absolute levels for distributive decisions expresses that persons are of lesser importance because how they are doing is deemed to be of no consequence. Valuing a person demands, in part, valuing his or her well-being, and valuing a person’s well-being appropriately, in turn, includes a concern for the person’s well-being level.

Some authors might object that I misunderstand what respect for persons involves. Stephen Darwall, for example, has denied that the normativity of welfare is grounded in an individual’s value as a person.<sup>45</sup> According to Darwall, the attitude that is appropriate to have toward persons as such is respect, an attitude that is responsive to persons being rational agents. Having a value that makes one—and one’s well-being—worthy of care or concern and having a value that makes one worthy of respect are distinct, according to Darwall. Connie Rosati also sharply distinguishes between respect for persons and concern for them and their welfare.<sup>46</sup>

There are two possible replies. First, one might concede a sharp distinction between the attitude of respect on the one hand and the attitude of care or concern for individuals and their well-being on the other. Nonetheless, appreciating the value that makes a person worthy of care or concern involves attending to an aspect of the person’s welfare—namely, her absolute level. The normativity of a person’s absolute level would still depend on a prior value of the person, even if it is not her value as a person as such. Agents fail to show proper concern for persons when they ignore their absolute levels.

43 Frankfurt, “Equality and Respect,” 12. For a critique of Frankfurt’s account of respect for persons, see Raz, “On Frankfurt’s Explanation of Respect for People”. Raz instead proposes that disrespect for persons consists in a denial that persons are of value in themselves. In response, Frankfurt denies that respect for persons is in any important way connected with the value of persons (“Reply to Joseph Raz”). Yet Frankfurt’s account does connect with the value of persons, albeit in a manner not exactly captured by Raz’s account. We might fail to respect persons, for example, not because we deny that they are of value in themselves, but rather because we regard persons as having lesser value or a lower moral status than they in fact have. This illustrates “the circumstance that some important fact about the person is not properly attended to or is not taken appropriately into account.”

44 The notion that respect for persons includes a concern for how lives go is suggested in Scanlon, *What We Owe to Each Other*, 104. Similarly, McMahan has argued that concern for a person’s good is a component of respecting persons (*The Ethics of Killing*, 482–83).

45 Darwall, *Welfare and Rational Care*, 14–15; and Darwall, “Reply to Feldman, Hurka, and Rosati,” 644–45.

46 Rosati, “Darwall on Welfare and Rational Care,” 625–26.

Second, Darwall has overstated the distinction between respect for persons and concern for them and their well-being. Darwall seems to think that because care or concern is an appropriate attitude to have with regard to sentient creatures who are not plausibly regarded as persons, care and concern have nothing to do with the value persons have qua persons.<sup>47</sup> However, it does not follow from this that the value of persons as such plays no role in grounding the normativity of a person's welfare or in shaping the kind of concern owed. In fact, Darwall's conception of recognition respect renders the attitude of respect appropriate in matters concerning a person's welfare. As Darwall has noted, "recognition respect lights up the person's dignity as a person and the constraints on relating to him."<sup>48</sup> Human persons are embodied rational beings whose lives can go better or worse, and surely this fact constrains how we may relate or act toward them. And according to Darwall, "the sort of regard involved in recognition respect is a regard for a fact or feature as having some weight in deliberations about how one is to act."<sup>49</sup> The fact that the lives of persons can go better or worse is a fact meriting weight in our deliberations about how we are to act and thus qualifies attention to absolute levels as a form of recognition respect. Finally, Darwall has maintained that by distinguishing respect for persons from care and concern for them, he follows Kant's conception of respect. But Kant attempted to derive a duty to promote the ends and happiness of others from his second formulation of the categorical imperative—the Formula of Humanity.<sup>50</sup> Kant suggested that furthering the ends of others is a way of contributing to their happiness, which in turn suggests that Kant regarded our rational agency as at least partly constitutive of our well-being.<sup>51</sup> As such, Kant arguably did not see respect for persons as completely divorced from a concern for their well-being.

Insofar as it is plausible to consider well-being in some distributive decisions—a view I find intuitively appealing but do not defend in this paper—my aim here is simply to provide a rationale for why agents are justified in regarding absolute levels of well-being as part of what is morally relevant for such

47 Adler has mysteriously claimed that there are no moral reasons regarding the treatment of nonhuman animals partly because of the claim that norms must be justifiable to a community of persons who can engage in normative reasoning, and nonhuman animals are not members of this community (*Well-Being and Fair Distribution*, 449–50). This is a *non sequitur*. Subjects that govern themselves by norms must have certain capacities, but it does not follow from this that the *objects of normative concern*, which give us reasons for action, must possess the same capacities. The realm of moral reasons is not exhausted by the claims of persons.

48 Darwall, "Reply to Feldman, Hurka, and Rosati," 644.

49 Darwall, "Two Kinds of Respect," 41.

50 Kant, *Groundwork of the Metaphysics of Morals*, 4:430.

51 Similarly, Connie Rosati ("Personal Good") argues that agency is partly constitutive of our personal good.

distributive decisions. This moral relevance is grounded in the more fundamental value of persons and the way it is appropriate to value persons. In contrast, whichever unanimity condition is used in a Contractualist Priority View, it must either presuppose such moral relevance or rely on the kind of rationale I have introduced. But once we have a rationale in place, the contractualist machinery is no longer needed.

Contrary to a goal of impersonal value maximization, the notion of recognition respect provides a credible way of understanding the motivation to give priority to the worse-off and why certain reactive attitudes are warranted. The greater sense of urgency and empathetic distress we feel for the worse-off is engendered by our appreciation of the condition of the worse-off, which we recognize as mattering because persons matter. Accordingly, the worse-off have reasons to feel resentment or indignation, and agents have reason to feel remorse or regret in choosing not to aid them, when and because there is inadequate recognition of an important aspect of the lives of the worse-off.

Of course, there may be constraints on the appropriateness of such reactive attitudes. Someone who is worse off because of fully informed and deliberate choices this person made may not have sufficient reason to feel resentment or indignation when a distributor opts to aid a better-off person to a comparable degree. Clearly, the claim that a person's absolute level matters does not imply that it is the only consideration that matters or that it supersedes all other considerations. Noncontractualist deontic prioritarists can acknowledge that a person's responsibility may have bearing on distributive decisions. I have bracketed questions about responsibility for the sake of simplicity. Presumably, there are cases in which the interests of persons are at stake, yet these people are not responsible for their condition. That the account I have described can explain a range of reactive attitudes in distributive contexts does not entail that it will account for all of them without added complexity.<sup>52</sup>

Relatedly, the greater sense of urgency and empathetic distress we feel for the worse-off appear to have its limits. There may be no sense of urgency or empathetic distress when choosing to confer a benefit on one of two very well-off people who differ in levels of well-being.<sup>53</sup> Should noncontractualist deontic prioritarianism apply only in those circumstances in which a sense

52 A person's absolute level of well-being is often the result of both informed, deliberate choices and factors beyond a person's control, when persons are viewed as temporally extended agents. This view of persons—as opposed to a timeslice view—makes it difficult to ignore the connection between rational agency and well-being. How these components combine to determine a person's responsibility in a given context lies beyond the scope of this paper.

53 For a version of this objection to prioritarianism, see Crisp, "Equality, Priority, and Compassion."

of urgency or empathetic distress is appropriate? If so, this would imply that there is some threshold level of well-being under which deontic prioritarianism is applicable but beyond which it has no jurisdiction. This threshold might correspond with the level of well-being at which a reasonable agent equipped with nonpathological emotional capacities—e.g., they are not sociopaths—no longer feels empathetic distress. However, noncontractualist deontic prioritarians need not be committed to this view. Concern for each person's level of well-being is grounded in a concern for each person, or put another way, a concern for each person dictates a concern for each person's level of well-being. A person's level of well-being does not, it seems to me, become irrelevant simply because this person has a high level of well-being. Thus, while a sense of urgency and empathetic distress on the part of a distributor demarcates an important class of cases, the justification that noncontractualist deontic prioritarians offer in support of priority to the worse-off outstrips the presence of such reactive attitudes. However, this outstripping does not imply that the above rationale plays no role in explaining these reactive attitudes, as these are ultimately responses to persons and their condition. A full complement of emotional responses need not accompany evaluation of the condition of persons at every possible level of well-being in order for all levels to warrant consideration in distributive deliberation.

The reader likely will have observed that I have criticized telic prioritarianism's commitment to impersonal value, yet noncontractualist deontic prioritarians are also committed to a value that is not a value for anyone. But the value at the heart of noncontractualist deontic prioritarianism is not an outcome value that is to be maximized. As Parfit noted, the value of persons is not a kind of goodness.<sup>54</sup> It is a moral status that defines the ways in which we may treat persons.<sup>55</sup> According to noncontractualist deontic prioritarianism, persons assume a central role in our distributive deliberations as opposed to a merely instrumental role in light of what can be produced.<sup>56</sup>

54 Parfit, *On What Matters*, 1:240.

55 Parfit, *On What Matters*, 1:243–44. For a similar view of moral status, see Kamm, *Mortality, Mortality*, vol. 2; and Kamm, *Intricate Ethics*.

56 Because noncontractualist deontic prioritarianism is not concerned with impersonal outcome value maximization, it might avoid the Repugnant Conclusion. The value of persons does not give us reason to produce as much of this value as possible. Nor is there a directive to create new persons *for the sake of* maximizing total well-being or impersonal value, for such a directive inappropriately views persons as mere containers of well-being and subordinates the person to a value that itself depends on the value of the person. Maximizing total (weighted) well-being is not a good that provides us with independent reason to bring it about for its own sake, so it is unclear why an agent should prefer a world containing many lives that are barely worth living to a world containing a smaller number

Now, egalitarians might make similar claims about the ways in which we should view persons in distributive deliberation. And someone might object that the rationale for noncontractualist deontic prioritarianism can be usurped by egalitarians. Egalitarians might insist that relative levels of well-being are not insignificant aspects of human lives, and we must attend to them if we are to show proper concern for persons. Persons, being of equal value and owed equal moral concern, should, other things equal, have lives that go equally well, egalitarians might argue. Extrapolation of the rationale in this way may even help us better understand why some egalitarians maintain that comparative fairness is the basis of relational egalitarianism.<sup>57</sup> If this were right, then my proposal for justifying priority to the worse-off seems to fall into the same camp with contractualism in terms of support for both prioritarianism and egalitarianism.

It is worth noting the distinction between pure prioritarianism and mixed or hybrid prioritarian views. Pure prioritarians subscribe to two theses.<sup>58</sup> According to the positive thesis, absolute levels of well-being are morally important. According to the negative thesis, relative levels of well-being are morally irrelevant. While all prioritarian views must accept the positive thesis, mixed or hybrid views might reject the negative thesis.<sup>59</sup> The rationale for noncontractualist deontic prioritarianism can be viewed as support for the positive thesis. It does not, however, tell us to adopt the negative thesis. In contrast, contractualism does not support the positive thesis. Contractualism does not itself determine whether absolute levels, relative levels, or both are of moral importance.

While the rationale I have provided for priority to the worse-off might be developed in ways that support egalitarianism or a hybrid view, there are some considerations that suggest that the rationale speaks more strongly in favor of a concern for absolute levels of well-being. First, absolute levels are indicative of how things are with someone and reflect how a person is doing, whereas inequality is necessarily parasitic on information about absolute levels in order to be similarly informative. Second, absolute levels represent an essential aspect

---

of different people with lives of high quality. For discussion of how the practical standpoint and the conception of ourselves as moral agents from that standpoint can be used to resist the Repugnant Conclusion, see Mulgan, "Two Parfit Puzzles."

57 For discussion of the relation between fairness and inequality, see Temkin, "Equality, Priority, or What?" The reformulated rationale seems to explain why inequality might be thought to raise concerns about comparative fairness better than Temkin's claims about the impersonal value of outcomes.

58 The framing of pure prioritarianism that follows parallels Paula Casal's conception of sufficientarianism as containing positive and negative theses. See Casal, "Why Sufficiency is Not Enough."

59 For the possibility of mixed views, see Casal, "Why Sufficiency is Not Enough"; O'Neill, "Priority, Preference, and Value"; and Parfit, "Another Defence of the Priority View."

of each person's well-being, an aspect the appreciation of which is called for by our recognition of the value of each person, whereas it is not an essential aspect of a person's well-being that someone else may or may not have the same level of well-being. These considerations suggest that my defense of priority to the worse-off does not equally support prioritarianism and egalitarianism, even if it can be developed in ways to support a concern for both absolute and relative levels.<sup>60</sup> Because a main objective of this paper is to develop a rationale for the normativity of absolute levels, the possibility of a mixed or hybrid view will not be considered here, and I will continue to focus on the simpler formulation of the rationale that is articulated in noncomparative terms.

Noncontractualist deontic prioritarianism, I have claimed, leads us to assign a central role to persons in our distributive deliberations as opposed to a merely instrumental or subsidiary role in light of what can be produced. A person-centered approach to distributive ethics, some have argued, should focus exclusively on complete lives. Nagel, for example, argued that distributive principles apply only to whole lives, a claim he seemed to think follows from the unity of life and the possibility of intrapersonal compensation.<sup>61</sup> Following Nagel, Adler has also wielded the concepts of intrapersonal compensation and the unity of life in support of a lifetime approach to distributive ethics.<sup>62</sup> Williams and Nebel have not indicated any divergence from Nagel on this matter.<sup>63</sup> In contrast, some telic prioritarians such as Parfit and McKerlie have argued that we should also give priority to those who are worse off at particular times.<sup>64</sup> While there seems to be room for reasonable disagreement here, noncontractualist deontic prioritarians might endorse a concern for both whole lives and parts of lives.

Persons are temporally extended beings with temporally extended well-being. Each person has connected experiences and psychological states over the

60 I thank an anonymous reviewer for prompting me to clarify my position on the relation between the rationale for noncontractualist deontic prioritarianism and possible support for hybrid views.

61 Nagel, "Equality," 120.

62 Adler, *Well-Being and Fair Distribution*, ch. 6.

63 Williams, "The Priority View Bites the Dust?"; and Nebel, "Priority, Not Equality, for Possible People."

64 Parfit, "Equality or Priority?" and "Another Defense of the Priority View"; McKerlie, "Priority and Time," "Dimensions of Equality," and *Justice Between the Young and the Old*. For critiques of telic prioritarianism and its possible temporal applications, see Tännsjö, "Utilitarianism or Prioritarianism?"; and Andric and Herlitz, "Prioritarianism, Timeslices, and Prudential Value." These critiques target telic prioritarianism's claims about the impersonal value of outcomes—claims that are not made by deontic prioritarians.

course of his or her life.<sup>65</sup> Persons are also agents whose agency unites the various parts of our lives through plans, principles, intentions, commitments, ideals, etc.<sup>66</sup> Our agency plays a role in creating and shaping our diachronic interests or temporally extended personal good.<sup>67</sup> And persons do seem to care about the quality of our lives taken as a whole; each person wants to have lived a good life, however that is defined.<sup>68</sup> These considerations suggest that agents should adopt a global perspective when evaluating a person's life. A person's complete life seems to be especially relevant with respect to certain distributive contexts. One such context involves life-or-death decisions.<sup>69</sup> A prioritarian argument can be offered for giving a stretch of life of comparable quality and quantity to a younger person rather than to an older person when only one person can live. A general preference for the younger in lifesaving contexts seems plausible, but I do not see how it can be justified by views that take moments or parts of life as the sole units of distributive concern.

Notwithstanding, proper recognition respect of persons requires acknowledging that persons have perspectives and occupy points of view, sometimes very different ones throughout the course of a life. Points of view regarding parts of a life have no less normative significance than the point of view that encompasses life as a whole. Connie Rosati's remarks in another context partly explain why: "Because our features affect the quality of our experiences, partly making them what they are for us, and because our features can change, there is no such thing as 'what the experience is like for me.' Rather, there is 'what the experience is like for me, given what I am like at time *T*.'"<sup>70</sup> From within these sub-lifetime points of view, a person's well-being at these times matters—and matters independently of their contribution to the total well-being contained in a life taken as a whole, even if they also can matter in virtue of this contribution.<sup>71</sup> Noncontractualist

65 Parfit once claimed that a psychological reductionist theory of personal identity should lead us to revise our conception of compensation and the scope of distributive principles (*Reasons and Persons*, 335–38). For a critique of Parfit's view, see Adler, *Well-Being and Fair Distribution*, ch. 6. Cf. Jeske, "Persons, Compensation, and Utilitarianism"; and Holtug, *Persons, Interests, and Justice*.

66 For more on the role of agency in determining personal identity and the unity of life, see Korsgaard, "Personal Identity and the Unity of Agency."

67 Velleman, "Well-Being and Time"; and Rosati, "Personal Good."

68 For further discussion of the concern for complete lives, see Griffin, *Well-Being*; Velleman "Well-Being and Time"; Kamm, *Morality, Mortality*, vol. 1; and Temkin, "Aggregation Within Lives."

69 See Kamm, *Morality, Mortality*, vol. 1, for detailed discussion.

70 Rosati, "Persons, Perspectives, and Full Information Accounts of the Good," 317.

71 David Velleman has argued that diachronic interests should not have lexical priority over synchronic interests ("Well-Being and Time").

deontic prioritarians might add that a concern for a person's absolute level during these periods should be included in our appreciation of a person's perspective and concern for the person's well-being during these periods.

The normativity of sub-lifetime levels can be further appreciated from the distributing agent's perspective. The greater sense of urgency we feel for aiding the worse-off is a response to the plight of the worse-off, which we recognize as mattering because they matter. Out of concern for them, we see their plight as something to be alleviated. But this sense of urgency is naturally directed to those who are presently worse off or those who will be worse off, for this is the plight that we can alleviate, i.e., we cannot alleviate a person's past plight. Similar claims apply to the empathetic distress we feel when we contemplate priority to the worse-off. This is not to say that we cannot feel empathetic distress regarding a person's past life. Yet our empathetic distress seems to be heightened when evaluating unfortunate circumstances that we can affect and seems to be dampened when confronted by a person's current good fortune despite his or her bad past.

Taken together, the reasons appreciated from the distributor's perspective and the perspective of a potential beneficiary support a time-specific moral concern. Such concern does not depend on any particular theory of personal identity nor presuppose that parts of lives are metaphysically distinct entities.<sup>72</sup> But if we ascribe distributive importance to sub-lifetime absolute levels, would not this conflict with the possibility of intrapersonal compensation and the unity of life?

Those who wish to apply deontic prioritarianism to parts of lives in addition to whole lives without denying that intrapersonal compensation has normative importance might argue that a person can fail to be adequately compensated in certain contexts. In some cases, there may not be a clear answer regarding a person's lifetime well-being. In other cases, the lack of adequate compensation does not reduce a person's lifetime well-being to an extent that enables lifetime prioritarianism alone to account for priority judgments.

For example, suppose someone has a substantial change in perspective or conception of the good over the course of a single life. Smith may have led the first part of his life as a religiously devout person and then the next part as an atheist. While religious, Smith enjoyed observing the strictures of his religion, but as an atheist, he counts such a life as positively harmful. Smith might reasonably protest that the "benefits" he enjoyed early in life cannot count as adequate compensation for the hardships he endures as an atheist. Yet counting them as harms does not seem quite right either since they were not viewed as such from Smith's earlier religious perspective; and from this early perspective,

72 McKerlie has made similar claims about his telic version of time-specific prioritarianism (*Justice Between the Young and the Old*, ch. 6).

what atheist Smith now regards as benefits might have been regarded as harms. What perspective should be adopted when assessing what counts as benefits or burdens for Smith and computing his lifetime well-being?<sup>73</sup> If we think that only Smith's perspective as an atheist should be used when determining what counts as benefits and burdens to him while an atheist, ignoring his earlier perspective as a religiously devout person, then we are making a time-specific normative judgment. It would be mysterious to then claim that we should ignore this judgment, i.e., demand that we focus only on Smith's complete life despite his change in perspective, whenever atheist Smith is involved in interpersonal conflict.

Another kind of case involves particularly bad periods of life. Robinson endures a period of agony after having lived a very good life. We can give a benefit of a given size to either Robinson or Wilson, who is not in agony but has lived a mediocre life. It seems intuitively compelling that we should aid Robinson. For the example to support a time-specific priority judgment, Robinson's lifetime well-being still must be greater than Wilson's. Yet Robinson's previous good life may not compensate for his agony. The idea here is that adequate compensation may not always be determined simply by the sum total of benefits minus burdens within a life. This idea is supported by common reactions to the intrapersonal Repugnant Conclusion. Most would not accept that the loss of some of the best things in life would be adequately compensated by an indefinitely long life that is barely worth living but promises greater lifetime well-being in aggregate. Perhaps something similar applies to very bad periods of life.

If these two examples are valid, then a general acknowledgement of the possibility of intrapersonal compensation does not preclude time-specific prioritarianism.

This line of argument seems to restrict time-specific prioritarianism to a limited number of cases. Time-specific prioritarians might desire more. Consider the following case:

Person	$T_1$	$T_2$
B	9	3
C	6	6

73 For an argument against using the ideal observer perspective for determining a person's good, see Rosati, "Persons, Perspectives, and Full Information Accounts of the Good." Alex Voorhoeve has argued that intrapersonal intertemporal conflicts would arise even if preferences were idealized, and preference change creates serious difficulties for determining a person's lifetime well-being if idealized preferences function as the measure of well-being ("Preference Change and Interpersonal Comparisons of Welfare").

*B* might be said to be compensated in the sense that he is not worse off overall than if he were at 6 in each temporal period, that is, he is not worse off than if he had not had the advantage in  $T_1$  and the disadvantage in  $T_2$ .<sup>74</sup> If we can distribute 1 unit of well-being to either *B* or *C* at  $T_2$ , *B*'s compensation and lifetime well-being suggests that we should be indifferent.

However, to deliberate *only as if B* were at 6 in each temporal period is in one way to deny the reality of *B*'s situation in  $T_2$ . It fails to regard *B*'s circumstances as what they actually are—namely, that *B* is actually at 3 in  $T_2$ . It is to act as if *B* is currently not experiencing plight, to treat his hardship as if it were absent, which appears to show inadequate recognition respect for *B*. Some of the concerns generated from *B*'s point of view at  $T_2$  are about his life at  $T_2$ , and some of these are taken to matter in their own right, without reference to the whole. Furthermore, *B*'s plight in  $T_2$  engenders greater empathetic distress and a greater sense of urgency to aid *B* than to aid *C* in  $T_2$ . These motivational responses are significant features of moral appraisal and should not be dismissed lightly.

It might be objected that there is another sense of compensation that strongly conflicts with time-specific prioritarianism. Someone is compensated when his or her burden is made up for by a corresponding benefit. When this is the case, *nothing further is owed*. No further action is required to remedy or alleviate the person's burden. If *B*'s advantage in  $T_1$  makes up for *B*'s disadvantage in  $T_2$ , then his burden in  $T_2$  does not itself give us any independent reason for action.

I think we do sometimes understand compensation in this way, for example, when an agent makes informed, deliberate tradeoffs within his or her own life. We are of course assuming that *B* is not responsible for his advantage and disadvantage and are discussing compensation in the context of other-regarding distributive deliberation. Because the above sense of compensation by definition leaves no moral residue, it may be difficult to see how time-specific prioritarianism can gain a foothold unless it is denied that someone is compensated in this way, at least in certain cases. I have described a limited number of cases where this may be so. But should time-specific prioritarians claim more broadly that there is inadequate compensation in more mundane cases like that of *B* and *C*, thereby indicating that something more is owed to *B* than to *C*?

Intuitively, it does appear that there is stronger reason to aid *B* in  $T_2$  despite his past. Taking this intuition seriously would lead to a significant revision of the concept of compensation. For a plurality of cases, the size of benefits/burdens alone would not determine the adequacy of compensation. Revising the concept of compensation in this way seems less plausible for purely self-regarding choices. However, noncontractualist deontic prioritarians need not be

74 See McKerlie, "Dimensions of Equality," for discussion of this sense of compensation.

committed to this sort of revision for self-regarding choices. Unlike telic prioritarrians, noncontractualist deontic prioritarrians do not regard priority weights as intrinsic properties of the outcomes in which individuals are benefited or burdened. Rather, priority weighting arises from the ways in which we ought to value each other.<sup>75</sup> With respect to other-regarding moral concern, if the adequacy of compensation were also determined by people's absolute levels at particular times, then *B* might not be adequately compensated after all. This would allow us to regard *B*'s lower absolute level in  $T_2$  as providing a reason for action.<sup>76</sup> Nevertheless, even on this modified account of compensation, for some sufficiently large advantage in one temporal period, a person's disadvantage in another temporal period could be adequately compensated. In such cases, the benefit makes up for the burden, and nothing further would be owed.

It should be emphasized that it is not the inadequacy of compensation per se that provides the time-specific reason to prefer aiding *B* over *C* in  $T_2$ . *B*'s inadequate compensation is an intertemporal, global feature of *B*'s life, so it can be argued that this aspect of *B*'s *whole life* is what is driving the judgment that *B* should receive priority in  $T_2$ . The point of referring to *B*'s inadequate compensation was to argue that a time-specific concern regarding *B*'s life at  $T_2$  is not morally extinguished or canceled. Consider the following modification:

Person	$T_1$	$T_2$
<i>B</i>	3	9
<i>C</i>	6	6

In this modified example, it remains true that *B* is inadequately compensated on the above proposal. When his life is taken as a whole, this inadequacy might even count in favor of giving priority to *B* in  $T_2$  despite *B* and *C* having

75 If a person's absolute level of well-being matters because the person matters, and time-specific prioritarianism is valid, then does not morality require a concern for absolute levels in both other-regarding *and* self-regarding choices? This depends on whether we can be morally bound to act in certain ways with respect to our own lives, isolated from other-regarding considerations. Some doubt this on the grounds that persons can always release themselves from duties to themselves, which is to say they deny the possibility of self-obligation. The topic of self-regarding duties is an expansive one and not one that I can take up here. If noncontractualist deontic prioritarrians do not reject the possibility of such duties, then they must identify some relevant difference between the self-regarding perspective and the other-regarding perspective from which priority judgments arise, if they intend to preclude prioritarianism from self-regarding choices. This is a topic for future exploration.

76 McKerlie presents a version of this proposal for reconciling a telic version of time-specific prioritarianism with the concept of compensation ("Dimensions of Equality," 279–80).

equivalent lifetime well-being. However,  $C$  is worse off at  $T_2$ , and this itself engenders a time-specific reason for giving priority to  $C$  at  $T_2$ . Consequently, time-specific priority reasons can align with, or oppose, reasons that might arise from someone being inadequately compensated.

This approach to reconciling intrapersonal compensation over time with time-specific prioritarianism leaves a difficult question unanswered. How do time-specific prioritarians individuate temporal periods of a person's life? This question matters because how these temporal periods are individuated will determine what distributive decisions time-specific prioritarians will make.<sup>77</sup> If the individuation is arbitrary, then that seems to make time-specific prioritarianism arbitrary. I do not have a solution to the problem of individuating temporal periods. Nonetheless, this does not render the distinction between a concern for whole lives and a concern for parts of lives untenable. In other contexts, we acknowledge the normative importance of certain distinctions even if there is some degree of arbitrariness in how these distinctions are drawn, e.g., the age of consent, the poverty line, the speed limit. Furthermore, there is some degree of arbitrariness within pure lifetime prioritarian and egalitarian views. Each view must determine how much weight it ascribes to various lifetime aspects of well-being when assessing a person's overall condition.<sup>78</sup> Additionally, if proponents of pure lifetime views also think that structural goods are morally relevant, e.g., a life that starts off badly but ends well is better than a life that starts off well but ends badly, while the contents of each life are identical, then some decision must be made about how to weight and aggregate these different kinds of good to reach an overall verdict. And all prioritarian and egalitarian views—excluding maximin—must determine how much weight to give to aiding the worse-off. If arbitrariness does not invalidate these views, then opponents of time-specific prioritarianism should offer some account of when arbitrariness is or is not invalidating. There is enough support, I believe, to render time-specific prioritarianism worthy of further exploration by non-contractualist deontic prioritarians as a component of a view that combines a concern for whole lives and parts of lives.

Let us return to the problem with which I began this paper. Recall that some proponents of deontic prioritarianism have underscored the fact that it avoids the objections levied by Otsuka, Voorhoeve, and Fleurbaey against telic prioritarianism, objections that purport to show that unrestricted prioritarianism

77 McKerlie denies that this is a problem for time-specific prioritarianism, but he seems to be mistaken here (*Justice Between the Young and the Old*, 105–9). See Bykvist, review of *Justice between the Young and Old*, for the relevant counterargument.

78 Nagel discusses ordering various lifetime needs and interests, e.g., health, education, work, freedom, self-respect, and pleasure (“Equality,” 117).

does not adequately respect the unity of the individual or the separateness of persons. Given that the rationale for noncontractualist deontic prioritarianism makes no reference to interpersonal conflict, the question arises whether the view extends beyond interpersonal conflict. As it stands, the rationale seems to license a concern for levels of well-being in one-person cases. That non-contractualist deontic prioritarianism shares this feature with standard telic prioritarianism may not be all that surprising. After all, according to both theories, agents are guided by a value that is present in one-person cases. But if noncontractualist deontic prioritarianism is an unrestricted view, then it may be vulnerable to the objections raised by Otsuka, Voorhoeve, and Fleurbaey.

Their objections have generated a sizeable literature. Although some of the replies on behalf of unrestricted telic prioritarianism appear to show that their objections are not decisive, a thorough discussion of these objections and replies lies beyond the scope of this paper.<sup>79</sup> Yet in closing, I wish to express my skepticism about restricted views. The scope of *other-regarding* morality extends beyond interpersonal conflict.<sup>80</sup>

If absolute levels were irrelevant and morality had no place in one-person cases, then it should make no difference to the strength of our reason to aid when Sam is badly off and when Sam is well-off, supposing we can benefit Sam to the same degree. But this seems false. The concept of urgency is not limited only to cases in which we must decide to distribute aid among contestants.<sup>81</sup> The same goes for reactive attitudes. There is clearly greater urgency to aid Sam when he is badly off than when he is well-off. And greater moral approbation

79 See O'Neill, "Priority, Preference, and Value"; Parfit, "Another Defence of the Priority View"; Porter, "In Defence of the Priority View"; Segall, "In Defense of Priority (and Equality)"; Holtug, "Prioritarianism: Ex Ante, Ex Post, or Factualist Criterion of Rightness?"; and Agmon and Hitchens, "Prioritarianism."

80 Williams, Voorhoeve, and Fleurbaey appear to view morality exclusively as a way of mediating between conflicting claims. Yet these authors suggest that in one-person cases, individuals have *claims* to have their expected utility maximized. See Williams, "Priority View Bites the Dust?" 323; and Voorhoeve and Fleurbaey, "Egalitarianism and the Separateness of Persons," 398. What is the nature of this claim against a morally motivated stranger? If it is a moral claim, then morality is *not* solely concerned with mediating between competing claims. If it is not a moral claim, then what obligates strangers to maximize another person's expected utility in one-person cases, and who can be wronged when strangers decide against maximizing another's expected utility in such cases? It is unclear what these competing claims theorists have to offer here.

81 Nagel, recall, suggests that there is a standard of urgency that orders various needs and interests ("Equality," 117). The standard could apply to one-person cases even if the unanimity condition that presupposes it is designed for resolving conflicts between contestants for aid. That is, if a person can have more urgent and less urgent claims to aid, this involves a noncomparative judgment made with reference to the standard of urgency.

of Bob is warranted if he fails to aid Sam for no good reason when Sam is badly off than if Bob fails to aid Sam when Sam is well-off.<sup>82</sup>

What could be the basis for a conception of morality that is solely concerned with interpersonal conflict? We have already seen one prominent answer: Nagel claimed that the possibility of intrapersonal compensation invalidates the application of distributive principles within a life.<sup>83</sup> I argued above that intrapersonal compensation over time involving actual gains and actual losses can be reconceived to accommodate the application of distributive principles to parts of a life. I believe that Nagel's argument has even less force when applied to possible gains and possible losses within a life. We can appreciate this latter point without needing to revise the concept of compensation as was done in the discussion of temporal scope.

Otsuka, Voorhoeve, and Fleurbaey, inspired by Nagel, argue that a prospective greater gain compensates for a prospective lesser loss in intrapersonal gambles, and an agent ought to accede to this view when deciding whether to expose another person to a gamble, lest the agent be accused of failing to take seriously the unity of life. However, this sense of compensation may not have the same normative significance as compensation involving actual gains and actual losses. Nor does a prospective gain compensate for an actual loss.<sup>84</sup> One is left to wonder why a conception of compensation in prospects should always take precedence over the fact that the person we expose to a gamble will actually go uncompensated if she loses a gamble, falling to a lower absolute level. Invoking the unity of life only seems to beg the question about the importance of other senses of compensation. Presumably, followers of Nagel who endorse a purely whole-life approach to distributive ethics would ascribe special importance to intrapersonal compensation of an actual loss.

Consider the following example from Parfit.<sup>85</sup> Suppose that it is equally likely that either Tom is very well-off or Tom is very badly off. If we do *X*, Tom will receive a benefit if he is very badly off. If we do *Y*, Tom will receive a slightly greater benefit if he is very well-off. Doing *Y* obviously maximizes expected utility. But if compensation of actual losses matters, then we have reason to do

82 Assume that there is no cost to Bob in aiding Sam in either scenario, i.e., there is no interpersonal conflict. There appears to be a duty to aid in both cases, if the aid cannot otherwise be administered. Now suppose there were some cost to Bob. If there is a magnitude of cost such that there is no duty to aid Sam when he is well-off, yet there remains a duty to aid when Sam is badly off, then the *stringency* and *defeasibility* of a duty to aid is determined in part by a person's absolute level.

83 Nagel, "Equality," 120.

84 O'Neill made this point in "Priority, Preference, and Value," 346.

85 Parfit, "Another Defence of the Priority View," 408.

*X*, since *X* arranges for Tom to receive some measure of compensation when he would need compensation, i.e., if Tom turns out to be very badly off. Tom would not need compensation if he turned out to be very well-off. And doing *Y* would have no bearing on compensation of an actual loss. The notion of compensation, then, does *not* itself preclude prioritarian concern for levels of well-being in one-person cases and may actually align with such concern. Compensating someone becomes more urgent the worse off this person would be without the compensation.

This defense of prioritarianism in risky one-person cases differs from what is proposed by Parfit and Luc Bovens.<sup>86</sup> Both authors claim that we should be risk averse when making decisions on another person's behalf.<sup>87</sup> In contrast, my argument does not rely on claims about risk aversion or the impersonal value of outcomes. Rather, it claims that if we care about compensation, as Otsuka, Voorhoeve, and Fleurbaey profess they do, then certain decisions may be more consistent with this concern despite contravening expected utility maximization.

Noncontractualist deontic prioritarians do not view persons as mere sites or loci for value production, including expected utility maximization. Persons are not mere instruments for maximization of the good, not even their own good or expected good. Fixating solely on expected utility maximization in one-person cases such as Tom's loses sight of the person, the *end in itself* for whom we ultimately act. A concern for a person's level of well-being is rooted in a concern for the person, and this concern for levels is not rendered irrelevant simply because a person does not compete with others. Since noncontractualist deontic prioritarians are ultimately responding to the value of each person, it is difficult to see how extending prioritarianism to one-person cases inappropriately values persons.

#### 4. CONCLUSION

A noncontractualist version of deontic prioritarianism is a viable contender within the spectrum of prioritarian views. This view is more plausible than its telic or contractualist counterparts in explaining our moral thinking about priority to the worse-off. In counting how things are with a person, i.e., counting her absolute level of well-being, we acknowledge and express that the person matters. We see that a person's absolute level matters because the

86 Parfit, "Another Defence of the Priority View," 423; and Bovens, "Concerns for the Poorly Off in Ordering Risky Prospects," 404.

87 For a critique of this approach and the notion that there is a divergence in the goodness of outcomes and what is expectably best for someone in risky one-person cases, see McCarthy, "The Priority View."

person matters, and we respect their value by attending to their condition. This rationale grounds the normativity of absolute levels, elucidates the motivation to give priority to the worse-off, and explains why various reactive attitudes are warranted. The rationale might be developed to support hybrid or mixed views; but its support for a concern for relative levels of well-being appears to be weaker. I have contended that noncontractualist deontic prioritarianism can be applied to both whole lives and parts of lives, which is compatible with a person-centered distributive ethic. And although noncontractualist deontic prioritarianism is not restricted to interpersonal conflict, this is not an embarrassment for the view.<sup>88</sup>

*California Hospital Medical Center*  
*cmsoto@proton.me*

#### REFERENCES

- Adler, Matthew. *Well-Being and Fair Distribution: Beyond Cost-Benefit Analysis*. Oxford: Oxford University Press, 2012.
- Adler, Matthew, and Nils Holtug. "Prioritarianism: A Response to Critics." *Politics, Philosophy, and Economics* 18, no. 2 (February 2019): 101–44.
- Agmon, Shai, and Matt Hitchens. "Prioritarianism: A (Pluralist) Defense." *Journal of Ethics and Social Philosophy* 15, no. 1 (January 2019): 19–42.
- Anderson, Elizabeth. *Value in Ethics and Economics*. Cambridge, MA: Harvard University Press, 1993.
- Andric, Vuko, and Anders Herlitz. "Prioritarianism, Timeslices, and Prudential Value." *Australasian Journal of Philosophy* 100, no. 3 (May 2021): 595–604.
- Arneson, Richard. "Luck Egalitarianism and Prioritarianism." *Ethics* 110, no. 2 (January 2000): 339–49.
- Bovens, Luc. "Concerns for the Poorly Off in Ordering Risky Prospects." *Economics and Philosophy* 31, no. 3 (November 2015): 397–429.
- Brink, David. "The Separateness of Persons, Distributive Norms, and Moral Theory." In *Value, Welfare, and Morality*, edited by R. G. Frey and Christopher W. Morris, 252–89. Cambridge: Cambridge University Press, 1993.
- Brown, Campbell. "Prioritarianism for Variable Populations." *Philosophical Studies* 134, no. 3 (June 2007): 325–61.
- Bykvist, Krister. Review of *Justice between the Young and Old*, by Dennis McKerlie. *Ethics* 125, no. 3 (April 2015): 895–900.

88 I thank the anonymous reviewers for their helpful comments on earlier versions of this paper.

- Casal, Paula. "Why Sufficiency Is Not Enough." *Ethics* 117, no. 2 (January 2007): 296–326.
- Crisp, Roger. "Equality, Priority, and Compassion." *Ethics* 113, no. 4 (July 2003): 745–63.
- Darwall, Stephen. "Reply to Feldman, Hurka, and Rosati." *Philosophical Studies* 130, no. 3 (September 2006): 637–58.
- . "Two Kinds of Respect." *Ethics* 88, no. 1 (October 1977): 36–49.
- . *Welfare and Rational Care*. Princeton: Princeton University Press, 2002.
- Frankfurt, Harry. "Equality and Respect." *Social Research* 64, no. 1 (Spring 1997): 3–15.
- . "Reply to Joseph Raz." In *The Contours of Agency: Essays on Themes from Harry Frankfurt*, edited by Sarah Buss and Lee Overton, 316–20. Cambridge, MA: MIT Press, 2002.
- Greaves, Hilary. "Antiprioritarianism." *Utilitas* 27, no. 1 (March 2015): 1–42.
- Griffin, James. *Well-Being: Its Meaning, Measurement, and Moral Importance*. Oxford: Oxford University Press, 1986.
- Holtug, Nils. *Persons, Interests, and Justice*. Oxford: Oxford University Press, 2010.
- . "Prioritarianism." In *Egalitarianism: New Essays on the Value of Equality*, edited by Nils Holtug and Kasper Lippert Rasmussen, 125–56. Oxford: Oxford University Press, 2007.
- . "Prioritarianism: Ex Ante, Ex Post, or Factualist Criterion of Rightness?" *Journal of Political Philosophy* 27, no. 2 (June 2019): 207–28.
- Jeske, Diane. "Persons, Compensation, and Utilitarianism." *Philosophical Review* 102, no. 4 (October 1993): 541–75.
- Kamm, Frances. *Intricate Ethics: Rights, Responsibilities, and Permissible Harm*. Oxford: Oxford University Press, 2007.
- . *Morality, Mortality*. Vol. 1, *Death and Whom to Save from It*. Oxford: Oxford University Press, 1993.
- . *Morality, Mortality*. Vol. 2, *Rights, Duties, and Status*. Oxford: Oxford University Press, 1996.
- . "Owing, Justifying, and Rejecting." *Mind* 111, no. 442 (April 2002): 323–54.
- Kant, Immanuel. *Groundwork of the Metaphysics of Morals*. Translated and edited by Mary Gregor. Cambridge: Cambridge University Press, 1998.
- Korsgaard, Christine. "Personal Identity and the Unity of Agency: A Kantian Response to Parfit." *Philosophy and Public Affairs* 18, no. 2 (Spring 1989): 101–32.
- Kumar, Rahul. "Reasonable Reasons in Contractualist Moral Argument." *Ethics* 114, no. 1 (October 2003): 6–37.

- Lange, Benjamin. "Restricted Prioritarianism or Competing Claims?" *Utilitas* 29, no. 2 (June 2017): 137–52.
- McCarthy, David. "The Priority View." *Economics and Philosophy* 33, no. 2 (July 2017): 215–57.
- McKerlie, Dennis. "Dimensions of Equality." *Utilitas* 13, no. 3 (November 2001): 263–88.
- . "Egalitarianism and the Separateness of Persons." *Canadian Journal of Philosophy* 18, no. 2 (June 1988): 205–25.
- . "Equality and Priority." *Utilitas* 6, no. 1 (May 1994): 25–42.
- . *Justice between the Young and the Old*. Oxford: Oxford University Press, 2012.
- . "Priority and Time." *Canadian Journal of Philosophy* 27, no. 3 (September 1997): 287–309.
- McMahan, Jeff. *The Ethics of Killing: Problems at the Margins of Life*. Oxford: Oxford University Press, 2002.
- Mulgan, Tim. "Two Parfit Puzzles." In *The Repugnant Conclusion: Essays on Population Ethics*, edited by Jesper Ryberg and Torbjörn Tännsjö, 23–44. Boston: Kluwer Academic Publishing, 2004.
- Nagel, Thomas. "Equality." In *Mortal Questions*, 106–27. Cambridge: Cambridge University Press, 1979.
- . *The Possibility of Altruism*. Oxford: Oxford University Press, 1970.
- Nebel, Jacob. "Priority, Not Equality, for Possible People." *Ethics* 127, no. 4 (July 2017): 896–911.
- O'Neill, Martin. "Constructing a Contractualist Egalitarianism: Equality after Scanlon." *Journal of Moral Philosophy* 10, no. 4 (January 2013): 429–61.
- . "Priority, Preference, and Value." *Utilitas* 24, no. 3 (September, 2012): 332–48.
- Otsuka, Michael. "Prioritarianism and the Measure of Utility." *Journal of Political Philosophy* 23, no. 1 (March 2015): 1–22.
- . "Prioritarianism and the Separateness of Persons." *Utilitas* 24, no. 3 (September 2012): 365–80.
- Otsuka, Michael, and Alex Voorhoeve. "Equality versus Priority." In *Oxford Handbook of Distributive Justice*, edited by Serena Olsaretti, 65–85. Oxford: Oxford University Press, 2018.
- . "Why It Matters That Some Are Worse Off Than Others: An Argument against the Priority View." *Philosophy and Public Affairs* 37, no. 2 (Spring 2009): 171–99.
- Parfit, Derek. "Another Defence of the Priority View." *Utilitas* 24, no. 3 (September 2012): 399–440.
- . "Can We Avoid the Repugnant Conclusion?" *Theoria* 82, no. 2 (May

- 2016): 110–27.
- . “Equality or Priority?” In *The Ideal of Equality*, edited by Matthew Clayton and Andrew Williams, 81–125. New York: Palgrave MacMillan, 2002.
- . “Future People, the Non-Identity Problem, and Person-Affecting Principles.” *Philosophy and Public Affairs* 45, no. 2 (Spring 2017): 118–57.
- . *On What Matters*. 2 vols. Oxford: Oxford University Press, 2011.
- . *Reasons and Persons*. Oxford: Clarendon Press, 1984.
- Porter, Thomas. “In Defence of the Priority View.” *Utilitas* 24, no. 3 (September 2012): 349–64.
- Rabinowicz, Wlodek. “Prioritarianism for Prospects.” *Utilitas* 14, no. 1 (March 2002): 2–21.
- Rachels, Stuart. “Repugnance or Intransitivity: A Repugnant but Forced Choice.” In *The Repugnant Conclusion: Essays on Population Ethics*, edited by Jesper Ryberg and Torbjörn Tännsjö, 163–86. Boston: Kluwer Academic Publishing, 2004.
- Rawls, John. *A Theory of Justice*. Rev. ed. Cambridge, MA: Harvard University Press, 1999.
- Raz, Joseph. “On Frankfurt’s Explanation of Respect for People.” In *The Contours of Agency: Essays on Themes from Harry Frankfurt*, edited by Sarah Buss and Lee Overton, 299–315. Cambridge, MA: MIT Press, 2002.
- Ridge, Michael. “Contractualism and the New and Improved Redundancy Objection.” *Analysis* 63, no. 4 (October 2003): 337–42.
- . “Saving Scanlon: Contractualism and Agent-Relativity.” *Journal of Political Philosophy* 9, no. 4 (December 2001): 472–81.
- Rosati, Connie. “Darwall on Welfare and Rational Care.” *Philosophical Studies* 130, no. 3 (September 2006): 619–35.
- . “Personal Good.” In *Metaphysics after Moore*, edited by Terry Horgan and Mark Timmons, 107–32. Oxford: Oxford University Press, 2006.
- . “Persons, Perspectives, and Full Information Accounts of the Good.” *Ethics* 105, no. 2 (January 1995): 296–325.
- Scanlon, Thomas. “Contractualism and Justification.” In *Reason, Justification, and Contractualism: Themes from Scanlon*, edited by Markus Stepanians and Michael Frauchiger, 17–44. Boston: De Gruyter, 2021.
- . “Contractualism and Utilitarianism.” In *Utilitarianism and Beyond*, edited by Amartya Sen and Bernard Williams, 103–28. Cambridge: Cambridge University Press, 1982.
- . “Replies.” *Ratio* 16, no. 4 (December 2003): 424–39.
- . *What We Owe to Each Other*. Cambridge, MA: Harvard University Press, 1998.

- Segall, Shlomi. "Equality or Priority about Competing Claims?" *Economics and Philosophy* 38, no. 2 (July 2022): 242–65.
- . "In Defense of Priority (and Equality)." *Politics, Philosophy and Economics* 14, no. 4 (November 2015): 343–64.
- . "Sufficientarianism and the Separateness of Persons." *Philosophical Quarterly* 69, no. 274 (January 2019): 142–55.
- Stratton-Lake, Philip. "Scanlon's Contractualism and the Redundancy Objection." *Analysis* 63, no. 1 (January 2003): 70–76.
- Suikkanen, Jussi. "Contractualist Replies to the Redundancy Objections." *Theoria* 71, no. 1 (March 2005): 38–58.
- Tännsjö, Torbjörn. "Utilitarianism or Prioritarianism?" *Utilitas* 27, no. 2 (June 2015): 240–50.
- . "Why Derek Parfit Had Reasons to Accept the Repugnant Conclusion." *Utilitas* 32, no. 4 (December 2020): 387–97.
- Temkin, Larry. "Aggregation within Lives." *Social Philosophy and Policy* 26, no. 1 (January 2009): 1–29.
- . "Equality, Priority, or What?" *Economics and Philosophy* 19, no. 1 (April 2003): 61–87.
- Velleman, David. "A Right of Self-Termination?" *Ethics* 109, no. 3 (April 1999): 606–28.
- . "Well-Being and Time." *Pacific Philosophical Quarterly* 72, no. 1 (March 1991): 48–77.
- Voorhoeve, Alex. "Preference Change and Interpersonal Comparisons of Welfare." In *Preferences and Well-Being*, edited by Serena Olsaretti, 265–79. Cambridge: Cambridge University Press, 2006.
- Voorhoeve, Alex, and Marc Fleurbaey. "Egalitarianism and the Separateness of Persons." *Utilitas* 24, no. 3 (September 2012): 381–98.
- . "Priority or Equality for Possible People?" *Ethics* 126, no. 4 (July 2016): 929–54.
- Williams, Andrew. "The Priority View Bites the Dust?" *Utilitas* 24, no. 3 (September 2012): 315–31.

# NOT LIVING MY BEST LIFE

## A REPLY TO MASNY

*Guy Fletcher*

**D**ESPITE THE EXTENT of their other disagreements, philosophers of well-being have tended to assume that the prudential goodness of a life is determined by what *actually* happens or its *actual* features. Following Michal Masny, I will call this assumption the “orthodox view” of the prudential goodness of a life.

In an interesting recent paper, Masny has put forward a novel theory of the goodness of a life that explicitly *rejects* the orthodox view: the *Dual Theory*.<sup>1</sup> The Dual Theory, if true, would have significant implications for various issues in normative ethics, such as duties of beneficence. It is thus worthy of serious attention. In this paper, I first explain the Dual Theory and the motivation that Masny provides for it. I then put forward three general problems for the Dual Theory and Masny’s case for it.

### 1. THE DUAL THEORY

Masny introduces the Dual Theory thus:

How good a life is for someone is determined jointly by their level of well-being and the degree to which they realize their potential. (7)

This contrasts with what Masny calls the “orthodox view”:

How good a life is for its subject depends exclusively on the things that actually happened within it. (6)

The crucial part of the orthodox view for Masny and for my discussion is the claim that the goodness of a life is determined by *actual* events or features of the life.

To give a sense of just *how* orthodox this assumption is, notice first that it is neutral on the division between all the main theories of well-being: hedonism, desire theories, perfectionism, objective list theories, hybrid theories, value-fulfilment

1 Masny, “Wasted Potential.” All parenthetical page references are to Masny’s paper.

theories, etc. Furthermore, though it might superficially seem otherwise, the orthodox view is also neutral on various debates *within* such camps. For example, desire fulfilment theorists face questions about the relation between a subject's desire, the timing of the desired state of affairs, and the time of the prudential goodness generated.<sup>2</sup> The orthodox view is compatible with every view on this question. The orthodox view is also fully compatible with highly *idealized* forms of desire views, so long as such views maintain, as they do, that it is only the *actual* events or features of our lives that contribute to their goodness (by satisfying the desires of some hypothetical counterpart).<sup>3</sup> Moreover, the orthodox view is compatible also with the possibility of posthumous harms and benefits, as long as only actual events determine someone's level of well-being posthumously. (It is also compatible with rejecting such posthumous effects.)<sup>4</sup>

Thus, to repeat, on the *orthodox* view, the goodness of a life is determined by one thing: how much well-being it *actually* instantiates. By contrast, on Masny's Dual Theory, the goodness of a life is determined by two things: how much well-being it actually instantiates *and* how much of its potential is realized.

I will examine the Dual Theory in more depth in the next section. It is helpful first to see the motivation for it. Masny provides two main cases to support the Dual Theory and to undermine the orthodox view. The first case involves Sophie Germain, a nineteenth-century mathematics prodigy:

She was born to a wealthy Parisian family and enjoyed a life rich in meaningful relationships, sophisticated pleasures, and important achievements. However, much of her exceptional academic talent was wasted because of the obstacles she faced as a woman. Early on, her parents tried to hinder her youthful fascination with mathematics. Later, she was barred from attending the Ecole Polytechnique and the meetings of the Paris Academy of Sciences, and both her manuscripts and published work were regularly ignored by her contemporaries. (6)

Masny argues that the case of Sophie evokes a kind of "evaluative ambivalence":

On the one hand, her story is uplifting: she experienced a lot of what makes life valuable. On the other, there is a sense of tragedy that we just

2 On these issues see, for example, Bradley, *Well-Being and Death*, sec. 1.3; Baber, "Ex Ante Desire and Post Hoc Satisfaction"; Dorsey, "Desire-Satisfaction and Welfare as Temporal" and Dorsey, "Prudence and Past Selves"; Bruckner, "Present Desire Satisfaction and Past Well-Being"; and Lin, "Asymmetrism about Desire Satisfactionism and Time."

3 For defense and discussion of such views, see Railton, "Facts and Values"; and Rosati, "Internalism and the Good for a Person."

4 On posthumous harms and benefits, see Pitcher, "The Misfortunes of the Dead"; and Boonin, *Dead Wrong*.

cannot shrug off: she could have achieved much more, but didn't, and could have been more appreciated and spared many frustrations, but wasn't.<sup>5</sup> (6)

This ambivalence seems genuine to me. Masny argues that the Dual Theory makes sense of this ambivalence, among its other virtues (and that the orthodox view cannot make good sense of such ambivalence). Let us now see the Dual Theory in detail.

## 2. THE DUAL THEORY EXPLORED

The best way to appreciate the details of the Dual Theory of the prudential goodness of a life is by tracing Masny's route through various choices for this view. First, is someone's potential determined by facts about them as an individual or by some reference class to which the individual belongs? Masny holds that it is the *individual's* potential. Second, what is the relevant degree of potential? Is it someone's level of well-being in a close possible world or something more demanding? Masny suggests that their potential is their "*maximal* possible well-being" (16). Third, is it the well-being that was actually possible or that which the individual *believed* to be possible? Masny opts for the *objective* version.

I have no reservations about these choices (particularly the first and third).<sup>6</sup> A trickier one, as Masny acknowledges, is the fourth choice: What is the relevant kind of possibility? Here there is a challenge. Construing possibility extremely broadly, such as logical possibility, would end up giving everyone the same, extremely high level of potential well-being. Thus the Dual Theory needs something more restrictive if it is not to end up dialectically uninteresting. Masny characterizes a more restrictive account of possibility in terms of "intrinsic potential":

We may refer to the distinction between *intrinsic* and *extrinsic* potential. The former is the potential whose attainment does not require any direct or significant alteration of the individual's constitution, whereas the latter notion is more encompassing. . . . Of these two notions, intrinsic potential appears to me to be closer to what matters. (18)

Masny illustrates this with his second case, the example of a boy named Billy, introduced earlier in the paper as having a "serious cognitive impairment" (12). Masny writes:

- 5 Nothing in Masny's discussion or this reply rules out *other* kinds of ambivalence—for instance, judging that a life was prudentially good but morally disappointing.
- 6 Masny's choice on the second question is one with which people who are attracted to the Dual Theory might want to disagree. Constraints of space preclude discussion here.

Billy’s intrinsic potential is not much higher than his actual well-being, whereas his extrinsic potential is considerably higher. Of these two notions, intrinsic potential appears to me to be closer to what matters. After all, there is something heartening about Billy’s life. I will assume this view for the remainder of the discussion. (18)

Thus, in sum, the full version of the Dual Theory holds:

The goodness of a life is determined by two factors: (a) the level of well-being actually realized and (b) the degree to which the individual attains the maximal level of well-being that it is possible for that individual to realize without direct or significant alteration of their constitution.

When it comes to the interaction between these two factors, Masny outlines two models:

According to the *Addition Model*, to determine the overall value of a life, we need to add the value of realized potential to the contribution made by well-being, where the former is represented by a non-negative number. . . . By contrast, the *Subtraction Model* determines the overall value of a life by subtracting the disvalue of unrealized potential from the value of well-being. . . . (19)

Here Masny opts for the Subtraction Model.

Having seen the Dual Theory in full detail, we can now apply it to the cases of Sophie and Billy, deploying some purely illustrative numbers (table 1). The Dual Theory can note that Sophie’s life contained a lot of well-being but fared poorly relative to the well-being she could have attained without direct or significant alteration of her constitution. (The relevant change needed was in her society.) Conversely, the theory can hold that while Billy fared less well in absolute terms, he fared *well* relative to the level of well-being he could have attained without direct or significant alteration of his constitution.

Table 1

	Well-Being Actually Instantiated	Maximal Possible Well-Being without Direct or Significant Alteration of Constitution	Unrealized Well-Being	Dual Theory (Subtraction Model)
Sophie	10	25	15	$(10 - 15) = -5$
Billy	2	3	1	$(2 - 1) = 1$

Note that the numbers here are purely illustrative. Understandably, Masny does not take a stand on the overall noncomparative goodness or badness of either Billy’s life or Sophie’s life, nor on what difference, if any, there is between them.

I use these numbers only to explain the *structure* of the view, not to make claims about its specific implications in the cases Masny mentions.

### 3. THE DUAL THEORY VERSUS THE ORTHODOX VIEW

#### 3.1. Preserving Ambivalence

As noted above, Masny seems right about the case of Sophie. We do feel ambivalent. Her life was good and also tragic. Yet I think that an orthodox view, with proper supplementation, offers a satisfactory explanation of this.

The supplementation that the orthodox view needs is the addition of an alternative explanation (one that Masny rejects in the paper): treating our evaluative ambivalence as stemming from the fact that we make two different types of judgments about Sophie's life. We make a noncomparative judgment of how good it *was* (her life went well) and a comparative judgment of how good it *could have been* but for the prejudice that blighted it (her life could have gone much better). I will call this supplement the "two judgments strategy."

One advantage for the two judgments strategy over the Dual Theory is that it preserves the *ambivalence* that Masny uses to motivate the Dual Theory. Taking the two judgments strategy, we can think that Sophie's life went well and that it could have gone much better. We can be simultaneously pleased for her and frustrated or disappointed.

The Dual Theory view, by contrast, removes the ambiguity. This is because the two factors—well-being actually realized and attainment of maximal possible well-being—serve to generate a single verdict about the goodness of her life (through either the Addition or Subtraction Models). This makes it surprising that Masny articulates evaluative ambivalence thus:

A full and accurate description of the quality of her life seems to preclude merely summing the two evaluations, positive and negative. Instead, it requires that we maintain both judgments at once. (9)

The Dual Theory of course gives the overall goodness of life two subparts. But it precisely does *not* require that we maintain both judgments at once. Instead, it holds that there is one overall fact about the goodness of the life—one determined by the interplay of two factors. We see this in the way that the Dual Theory would give us *one* verdict about each case. This might be, for instance, that Sophie's life went badly and that Billy's life was minimally good. By contrast, the orthodox view, with the two judgments strategy, can preserve ambivalence. We judge that Sophie's life went well, which pleases us. And we judge separately that it could have gone much better, a source of regret. Thus one

strike against the Dual Theory—and one source of support for the orthodox view—is that the Dual Theory removes evaluative ambivalence.

One might reply that the Dual Theory can say that the different *subsidiary* evaluations of the actual well-being and the realisation of potential well-being, even if they feed into a single judgment, are still sufficient to preserve ambivalence. But this explanation of ambivalence is no better than the explanation that the orthodox view offers. The orthodox view says that we are ambivalent because we make a positive (noncomparative) judgment of the goodness of the life and a negative (comparative) judgment of how good it was relative to how good it could have been. The Dual Theory holds that the inputs into the single judgment of the goodness of the life that it provides are a positive judgment and a negative judgment. If the subsidiary judgments were enough to preserve ambivalence, it is unclear why the two judgments that can be offered by the orthodox view are any worse. Thus, to put the objection of this section more concessively: the Dual Theory either eliminates ambivalence or has no advantage over the orthodox view in explaining it.<sup>7</sup>

### 3.2. Irrelevant Potentials and Differences

The orthodox view can allow that unrealized potential goodness contributes *instrumentally* to the goodness or badness of a life (via unhappiness, etc.). What is at issue between the Dual Theory view and the orthodox view is whether unrealized potential well-being affects the goodness of a life *intrinsically* and *separately* from actualized well-being.<sup>8</sup> When we note this point, we see one way in which the case of Sophie is a less than pure test case. After all, Sophie was aware and presumably anguished by the ways her circumstances prevented her life from going better than it did. Let us consider another case—about someone named Megan—that amends this feature, through which we see another objection to the Dual Theory view:

*Megan:* Megan “enjoyed a life rich in meaningful relationships, sophisticated pleasures, and important achievements.” Megan’s developed talents, unlike Sophie’s, lay in the kinds of activities that were unrestricted to women at that time. However, Megan, like Sophie, lived before the

7 Masny appears to concede that the orthodox view is able to say satisfactory things about some of the cases. He provides other cases in the paper that he claims speak against the orthodox view (sec. 6). I regret that I lack the space to address Masny’s interesting cases there.

8 The “separately” is necessary to distinguish the Dual Theory from a theory of well-being with “unrealized potential” on its list of actual well-being determinants. The difference between such a view and the Dual Theory is an interesting question I lack the space to pursue.

development of soccer as a sport. Unbeknownst to anyone, Megan had the potential to be the greatest soccer player of all time, and this would have enhanced her well-being greatly.

Note that Megan’s realizing her potential would not have involved a direct or significant alteration of her constitution. Rather, as for Sophie, it would simply have required the wider world to have been different (by containing soccer).<sup>9</sup> Structurally speaking, as shown by table 2, the Megan case thus looks like the Sophie case, according to the Dual Theory:

Table 2

	Well-Being Actually Instantiated	Maximal Possible Well-Being without Direct or Significant Alteration of Constitution	Unrealized Well-Being	Dual Theory (Subtraction Model)
Megan	10	25	15	$(10 - 15) = -5$

Yet the fact that Megan had such unrealized and entirely unknown potential does not itself seem to make her life less good or regrettable, considered non-comparatively (even if we can regret that it prevented her life from being *even better*). Nor does it seem like it would have made her life better to have *lacked* this unrealized, unknown potential—an implication of the Dual Theory that Masny is commendably explicit about (14). To see this, consider Megan’s twin sister Twegan, who differed only in some minimal way that would have prevented her from being a successful football player.<sup>10</sup> Comparing the two sisters, we get table 3:

Table 3

	Well-Being Actually Instantiated	Maximal Possible Well-Being without Direct or Significant Alteration of Constitution	Unrealized Well-Being	Dual Theory (Subtraction Model)
Megan	10	25	15	$(10 - 15) = -5$
Twegan	10	12	2	$(10 - 2) = 8$

The Dual Theory will contend that Twegan’s life went better than Megan’s because Megan *could* have had a higher level of well-being (in a world containing soccer). This is despite the facts that neither Megan nor Twegan had

9 We could substitute some other talent or attribute here. All that must be true is that it be something that someone could have without being aware of it (or without being aware of how it could unlock substantially more well-being for them).

10 I here assume that such a case could be compatible with the “significant alteration of constitution” clause of the Dual Theory.

any inkling of this; they both lived in a world without soccer; and their lives contained the same amount of actual well-being.

To reiterate a point from above, it is consistent with the orthodox view to lament that Megan lived in the world without soccer and so missed out on the better life that she could have had. But the fittingness of this is distinct from taking this to detract from the actual goodness of her life and thinking that Megan’s life was a worse life than Twegan’s.

To take the point further, suppose that Megan had a friend, Fregan, whose life instantiated more well-being (+12) than Megan’s but had a lower maximum possible well-being than Megan’s. Consider table 4:

Table 4

	Well-Being Actually Instantiated	Maximal Possible Well-Being without Direct or Significant Alteration of Constitution	Unrealized Well-Being	Dual Theory (Subtraction Model)
Megan	10	25	15	$(10 - 15) = -5$
Twegan	10	12	2	$(10 - 2) = 8$
Fregan	12	20	8	$(12 - 8) = 4$

The Dual Theory would rank these lives as follows: Twegan > Fregan > Megan. Yet it seems reasonable to be indifferent between the lives of Megan and Twegan and to rank Fregan’s life over the other two, given that it contained more well-being. This, I suggest, is because we can—and do—separate our judgments of how good a life could have been from how good it actually was, with the latter tracking the well-being actually instantiated. This is easier to see once we think about *unknown* potential.

### 3.3. Extreme Skepticism

The previous problem for the Dual Theory stemmed from the verdicts that it would reach about cases and how it would reach them. A further worry is the degree of *skepticism* about the goodness of lives that the Dual Theory thereby produces. Masny introduces and then dismisses the worry:

One might also worry that as we expand our understanding of the goodness of a life, it becomes less transparent what is in our interest at any given time and how our lives are going as a whole. This is because now we need to know not just facts about the actual world, but also about various possible worlds. And this, in turn, might be taken to have a paralyzing effect on the ability of our theory of prudence to inform our lives.... This concern is overstated. Even the orthodox view on which the goodness of a life is determined solely by the level of well-being

gives us at most rough guidance regarding matters of self-interest. For example, we would all be hard-pressed to report our lifetime hedonic scores as of this morning. Likewise, the primary ambition of the Dual Theory is not to guide us through every single decision, but rather to help us better understand what makes life good. I believe it is successful in this regard. (32)

Masny is correct that the orthodox view generates difficult questions. But that does not undermine the comparative point. The Dual Theory makes it hugely more difficult to know whether a life went well because we must know both how well it went and how maximally well it *could* have gone. But this latter test seems so difficult to pass as to generate an extreme kind of skepticism, one where we could never, or almost never, know that someone's life had gone well. But this seems too stringent. It seems more plausible that our judgments about whether a life went well are not so modally sensitive, because they track something noncomparative: the actual amount of well-being attained. This is why we are able to know that some lives are good despite our not knowing how good they could have been.

#### 4. CONCLUSION

I have provided three main objections to Masny's Dual Theory. As compared with the orthodox view of the prudential goodness of a life, the Dual Theory eliminates ambivalence or is no better at explaining it than the orthodox view. It makes the goodness of lives sensitive to irrelevant factors and differences, and it generates an implausible degree of skepticism about our ability to know when a life goes well.<sup>11</sup>

University of Edinburgh  
guy.fletcher@ed.ac.uk

#### REFERENCES

Baber, Harriett. "Ex Ante Desire and Post Hoc Satisfaction." In *Time and Identity*, edited by Joseph Keim Campbell, Michael O'Rourke, and Harry Silverstein, 249–67. Cambridge, MA: MIT Press, 2010.

11 I hereby thank Daniel Groll, James Brown, Declan O'Gara, Debbie Roberts, Bryan Pickel, and my Philosophy of Well-Being cohort in 2022–23 for helpful discussion and also thank an anonymous reviewer from this journal for very helpful comments.

- Boonin, David. *Dead Wrong: The Ethics of Posthumous Harm*. New York: Oxford University Press, 2019.
- Bradley, Ben. *Well-Being and Death*. Oxford: Clarendon Press, 2009.
- Bruckner, Donald. "Present Desire Satisfaction and Past Well-Being." *Australasian Journal of Philosophy* 91, no. 1 (2013): 15–29.
- Dorsey, Dale. "Desire-Satisfaction and Welfare as Temporal." *Ethical Theory and Moral Practice* 16, no. 1 (February 2013): 151–71.
- . "Prudence and Past Selves." *Philosophical Studies* 175, no. 8 (August 2018): 1901–25.
- Lin, Eden. "Asymmetrism about Desire Satisfactionism and Time." In *Oxford Studies in Normative Ethics*, vol. 7, edited by Mark Timmons, 161–83. Oxford: Oxford University Press, 2017.
- Masny, Michal. "Wasted Potential: The Value of a Life and the Significance of What Could Have Been." *Philosophy and Public Affairs* 51, no. 1 (Winter 2023): 6–32.
- Pitcher, George. "The Misfortunes of the Dead." *American Philosophical Quarterly* 21, no. 2 (April 1984): 183–88.
- Railton, Peter. "Facts and Values." *Philosophical Topics* 14, no. 2 (Fall 1986): 5–31.
- Rosati, Connie S. "Internalism and the Good for a Person." *Ethics* 106, no. 2 (January 1996): 297–326.



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

The *Journal of Ethics and Social Philosophy* (JESP) is a peer-reviewed online journal in moral, social, political, and legal philosophy. The journal is founded on the principle of publisher-funded open access. There are no publication fees for authors, and public access to articles is free of charge. Articles are typically published under the CREATIVE COMMONS ATTRIBUTION-NONCOMMERCIAL-NODERIVATIVES 4.0 license, though authors can request a different Creative Commons license if one is required for funding purposes.



Funding for the journal has been made possible through the generous commitment of the Division of Arts and Humanities at New York University Abu Dhabi.

جامعة نيويورك أبوظبي



NYU ABU DHABI