

JOURNAL *of* ETHICS
& SOCIAL PHILOSOPHY

VOLUME XXIV · NUMBER 2
March 2023

ARTICLES

- 157 Consumer Complicity and the Problem of
Individual Causal Efficacy
Corey Katz
- 184 Ethical Veganism and Free Riding
Jacob Barrett and Sarah Zoe Raskoff
- 213 Fairness, Costs, and Procreative Justice
Gideon Elford
- 236 In Search of a Stable Consensus: Rawls's Model
of Public Reason and Its Critics
Cyril Hédoïn
- 258 The Purpose and Limits of Electoral
Accountability
Finlay Malcolm
- 291 Bare Statistical Evidence and the Right to
Security
N.P. Adams

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

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

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

Executive Editor

Mark Schroeder

Associate Editors

Saba Bazargan-Forward	Hallie Liberto
Stephanie Collins	Errol Lord
Dale Dorsey	Tristram McPherson
James Dreier	Colleen Murphy
Julia Driver	Hille Paakkunainen
Anca Gheaus	David Plunkett

Discussion Notes Editor

Kimberley Brownlee

Editorial Board

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

Managing Editor

Rachel Keith

Copyeditor

Susan Wampler

Typesetting

Matthew Silverstein

CONSUMER COMPLICITY AND THE PROBLEM OF INDIVIDUAL CAUSAL EFFICACY

Corey Katz

A SUPERMARKET orders meat from a company that runs factory farms that cause the suffering and death of millions of animals. A cellular phone company sources out their manufacturing to factories with egregious labor standards. Is it wrong for an individual consumer to purchase from such companies? We might think it is because making a purchase contributes to the harms or injustices that result from the company's actions. But we also might think that this is mistaken. Is it always the case that an individual's particular purchase contributes to a company's harm or wrongdoing? What about when the company from whom she is purchasing has millions of customers and billions of dollars in revenue? Or, when its decisions about production practices are not being made with an eye to how many units are being consumed at each particular moment? In these cases, claims about the morality of making a particular purchase seem to face the problem of "individual causal inefficacy."¹

It has been difficult to explain why a purchase that makes little to no difference to a producer's wrongdoing is itself morally wrong. Some have recently appealed to the concept of complicity in order to support the idea that consumers have a moral reason to avoid purchasing from companies engaged in wrongdoing. In this paper, I contribute to the development of this direction in consumer ethics. First, I explore how we should define moral complicity. Chiara Lepora and Robert Goodin provide a comprehensive analysis of complicity that captures key aspects of the concept.² In particular, they argue that an act counts as morally complicit if, and only if, it makes a causal contribution to the wrongdoing of another agent. But this definition of complicity seems to send us right back to the problem of individual causal inefficacy. In response, some have claimed that an individual's purchase still makes them complicit in such situations because they are still knowingly joining in with the wrong that the company or other consumers are doing. I argue, however, that it is difficult

1 Hiller, "Climate Change and Individual Responsibility."

2 Lepora and Goodin, *On Complicity and Compromise*.

to see making a purchase as a form of joining in a joint or group action. Instead, I respond to the causal inefficacy problem directly. I argue that even when the outcome is massively overdetermined, an individual purchase still makes a *causal contribution* to the principal harm or wrongdoing, even if it does not *make a difference* to that harm or wrongdoing.

This addresses the concern with applying Lepora and Goodin's conception of moral complicity—a purchaser can become complicit by making a purchase in such cases—but it does not resolve the deeper question of whether and why she should not be complicit. The issue is with how small the causal contribution such an individual purchase makes in cases of massive overdetermination. For the concept of moral complicity to gain traction here, I argue that we may have to appeal to normative concerns besides the outcome of the complicit action. While I argue such concerns are relevant, I conclude that this supplemental normative understanding of moral complicity cannot simply get us around the issue of minimal causal contribution, nor will it support the claim that purchasing from companies is usually all-things-considered morally wrong.

1. COMPLICITY AS CAUSAL CONTRIBUTION AND THE PROBLEM OF INDIVIDUAL CAUSAL EFFICACY

To explore the notion of consumer complicity further, we need a definition of what makes an agent morally complicit in the wrongdoing of another. In the law, complicity is often understood in terms of the crime of aiding and abetting. Someone who aids and abets does not actually commit the main crime, but helps another person or group commit it.³ Lepora and Goodin have developed a definition of moral complicity by extending this line of thinking. They argue that the fundamental element that makes someone complicit with the wrongdoing of a principal agent is that they knowingly act in a way that helps a principal wrongdoing succeed or be worse. They argue for a set of minimal conditions for one to be morally responsible for an act of complicity.

First, the action must make a causal contribution to the principal agent's wrongdoing (it must have "aided" or "helped") and that act must have been "neither involuntary nor accidental."⁴ Second, the agent must have reasonably

3 For example, Andy draws a floor plan of a bank for Bill, despite knowing Bill's intention to rob the bank. After Bill commits the robbery, Alice picks him up from the bank and drives him to a safehouse he has set up. Both Andy and Alice can be charged with aiding and abetting. It is true that they did not actually rob the bank, but because they should have reasonably believed that Bill was planning to rob the bank, and they performed an action that helped him with his crime, they are potentially criminally complicit.

4 Lepora and Goodin, *On Complicity and Compromise*, 83.

believed (or should have reasonably believed) that the principal agent was engaged in wrongdoing or had a wrongful plan and that their action would causally contribute to that wrongdoing.⁵ If these conditions are met, however minimally, Lepora and Goodin argue that the act is morally wrong and the complicit agent has, as they put it, a moral case to answer. They also emphasize that, like all wrongful actions, an act of complicity can be more or less *pro tanto* wrong and may not be all things considered wrong.

For the purposes of exploring the effect that the problem of individual causal inefficacy has on this definition, I would like to bracket the question of whether its epistemic conditions are or could ever be met in the case of complex, mediated, and sometimes global supply chains and production practices.⁶ I would also like to bracket questions about whether the actions of the sort under consideration—running concentrated animal-feeding operations, hiring workers at below living wages, etc.—always count as wrongdoing. Lepora and Goodin’s definition of moral complicity confirms that these questions have relevance for claims about consumer complicity. Still, if a customer’s purchase does not make any causal contribution to the wrongdoing or harm this would, from the start, imply that the customer is not complicit.

To explore the applicability of their definition of complicity further, let us look at an imagined case.

Bob’s Dresses: Aisha sees a cart set up in a mall with a sign that reads “Bob’s Dresses.” Arrayed on Bob’s cart are dresses with the most beautiful and distinctive embroidery Aisha has ever seen. She asks Bob where they are made. Bob explains, “I own a factory in a country where slavery is legal, and I own a number of slaves there, many of them children.” Aisha is in disbelief at the wrongful practices Bob uses to produce the dresses. She thinks it would be wrong to buy a dress now that she knows about those practices because doing so would make her complicit with slavery.

According to Lepora and Goodin’s definition of moral complicity, we must ask whether Aisha’s purchase will causally contribute to the success of Bob’s wrongful plan. Say that Bob’s plan is to make money by selling dresses produced by child slaves and that doing so is morally wrong. If Aisha gives him money, it serves as part of his revenue stream, thereby supporting his continued wrongdoing. Moreover, if Aisha did not buy the dress, it is likely that Bob’s revenue would have been smaller in a way that he might notice and take account of in

5 Lepora and Goodin, *On Complicity and Compromise*, 81–83.

6 Kingston, “Shopping with a Conscience?”

his production plans. By refusing to buy the dress, Aisha's act appears able to have some effect on the success of Bob's ongoing plan.

Now, take a related but different imagined case.

Big Box: Carla is walking through national retailer Big Box's store. She sees a table set out with dresses with the most beautiful and distinctive embroidery she has ever seen. She wonders where they are made and searches for information about them online. She finds out from the Boycott Big Box website that Big Box purchases them from factories that use child slavery. Carla is in disbelief at the wrongful practices used to produce the dresses. She thinks it would be wrong to buy one now that she knows this because it would make her complicit with child slavery.

For the purposes of argument, let us say that Big Box's act of purchasing dresses from producers engaged in wrongful plans is itself wrong.⁷ Many have the intuition that, like Aisha, Carla would be complicit with wrongdoing by purchasing the dress. Yet, when we apply Lepora and Goodin's conception of complicity to Big Box, we may not get that result.

This is because we might think that their conception of complicity loses traction when the consumer-producer relationship is one with the size, scope, or structure of many contemporary retail markets. In such markets, whether or not an individual customer makes a purchase might be expected to have no effect on what the company does. In Big Box, this may be because the company reasonably believes that sales of particular dresses will fluctuate each month, and the question of which dresses to stock will be sensitive to this. But it may also be because they reasonably believe that some dresses will not be sold within a set timeframe and will need to be put on clearance or discarded completely. All of this will likely be built into their plans and pricing from the outset. This means that, from the company's perspective, whether Carla buys the dress or not, the outcome will be identical. They will also have no reason to infer that the explanation for one more dress on the rack is Carla's moral outrage and not normal market dynamics. Large retailers usually expect the vast majority of individual customers to have some personal reason or another not to buy each of the particular products they offer.

For these reasons, it is not clear that any purchase, or refusal to purchase, will influence the success, extent, or severity of Big Box's wrongdoing. As Mark Budolfson puts it,

7 Again, this is an assumption made in order to focus specifically on the issue of causal inefficacy.

if the inefficacy objection is correct that when supply chains are long and complex an individual's consumption cannot be expected to make a difference to *the quantity produced*, then an individual's consumption also cannot be expected to make a difference to *the revenues of producers* for the same reasons. This undermines the more general claim of ethical consumerism that by purchasing morally objectionable products one is *complicit in evil* in an objectionable way because one thereby *supports* objectionable firms *by voting with one's dollars* in a way that benefits those firms.⁸

Yet, some ethical consumers appeal to the concept of complicity in order to describe the morality of situations where they recognize that their purchasing decisions do not make a difference. Someone might say, "Yes, I know my purchase won't make a difference, but I still will not buy from Big Box because I do not wish to be complicit with these injustices." Also, it might seem intuitive that both Aisha and Carla are morally complicit by purchasing a dress. Why should appealing to the structure or size of the market remove complicity? Is not one still knowingly choosing to participate in all of this? Finally, we may even get the worrisome result that Aisha becomes aware of this problem and decides she will only buy slave-made dresses from Big Box so that she can avoid being complicit with slavery.⁹ It is therefore worth exploring further whether the problem of causal inefficacy undermines the applicability of moral complicity to consumer cases like these.

2. CAUSAL CONTRIBUTION AND COMPLICITY

I think the strongest response to the problem is to recognize that complicity does not require that one's action *make a difference* to the success, extent, or severity of the principal agent's wrongdoing, only that it makes a *causal contribution* to that wrongdoing. In this section, I suggest that Carla's purchase makes a causal contribution to Big Box's wrongdoing even when the success, extent, and severity of her wrongdoing is overdetermined. But this causal contribution is likely to be very small. I conclude that the difficulty with applying the concept of moral complicity to cases like this does not lie with the definition of complicity itself, but with the normative reasons we have such that we should try to avoid being complicit with the wrongdoing of others.

As I argued above, the concept of complicity at its core involves helping or aiding another in doing wrong. If Carla's purchase will have no effect on the

8 Budolfson, "Is It Wrong to Eat Meat from Factory Farms?" 92.

9 Driver, "Individual Consumption and Moral Complicity," 67–68.

success, extent, or severity of Big Box's wrongdoing, how can it count as helping or aiding them? Put another way, in such a situation, would not refusing to buy the dress also have no effect on the success, extent, or severity of Big Box's wrongdoing? It therefore seems like we cannot appeal to the notion of helping or aiding to explain why Carla is complicit in these cases. The reason it seems this way is that here the wrongdoing is causally overdetermined. Still, I argue that someone who performs an action that makes a relevant contribution to that overdetermination still counts as a causal contributor to the overdetermined outcome. This is all that is needed for us to see why Carla's purchase still makes her complicit.

Recall that whether Carla purchases or not will have no effect on Big Box's wrongdoing because the way they make decisions about what products to produce/stock are not sensitive to individual purchasing decisions. But such decisions must be responsive to consumer demand at some level and Big Box remaining in business in part depends on this. So, there must be some number of consumers whose decision not to buy a particular product (for whatever reason) would signal to Big Box that there is insufficient demand for the product to make it worth their while to spend money producing or stocking it.¹⁰ The reason Carla's purchase will have no effect on Big Box's wrongdoing is because the number of other customers purchasing slave-made dresses is far above this threshold number, meaning their wrongdoing is overdetermined.

Let us say that the threshold number is ten thousand and let us say that one million slave-made dresses are purchased from Big Box by other consumers each year. The current number of consumers is therefore much higher than the threshold number at which a shift in consumer demand might lead Big Box to stop producing/stocking such dresses. Whether Carla purchases a dress or not, therefore, will have no effect on their wrongdoing because whether one million or one million and one people buy a slave-made dress will not make a difference to Big Box's wrongdoing. But being customer number one million and one still means that Carla makes some sort of causal contribution to Big Box's plans that she would not make if she refused to buy the dress.

While the correct account of causation is a contentious and unsettled question, the sense that Carla still counts as a causal contributor is supported by intuitions in other relevant cases. It is natural to think that *X* causes *Y* if and only if *X* is a necessary condition for *Y*'s occurrence such that, "but for" *X*, *Y* would not have occurred. For example, James throws a rock that hits Chen in the head, killing him. James caused Chen's death because, had James not thrown

10 Kagan, "Do I Make a Difference?"; Norcross, "Puppies, Pigs, and People"; Singer, "Utilitarianism and Vegetarianism."

the rock, Chen would not have died (at that time and in that way). But what if ten thousand uncoordinated strangers each throw a rock at Chen, all the rocks hit him at the same time, and ten rocks hitting Chen would be enough to kill him? Chen is hit by ten thousand rocks, leading to his death. I have the intuition that each rock thrower wrongfully causes Chen's death and that each rock thrower should not have thrown a rock. But if X must have been a necessary condition for Y in order to say that X caused Y , why would we think this? No individual's rock throw is necessary to cause Chen's death, since if one person did not throw a rock, Chen would still have been killed. Yet, I am reluctant to say that no individual of the ten thousand rock throwers caused Chen's death, or that no individual acted wrongly by throwing a rock. Chen's death was causally overdetermined, and each individual made a causal contribution to that overdetermined wrongful harm by throwing a rock.

Similarly, if less than ten thousand customers buying a slave-made dress would signal a shift in demand that would affect Big Box's wrongdoing, and Carla chooses to become purchaser one million and one, she is a causal contributor to Big Box's wrongdoing despite that wrongdoing being overdetermined. Carla might not be identical to a rock thrower who directly causes some share of Chen's wrongful death, but she is still a causal contributor to the wrongdoing that Big Box does. Perhaps her act is not as morally wrong as a rock thrower, but it still makes her complicit with the wrongdoing that Big Box does (and I explore the morality of that below).

These intuitions about overdetermination, causal contribution, and moral responsibility are supported by views of causation that recognize that outcomes can be the result of a set of antecedent conditions or a "joint cause."¹¹ In the rock-throwing case, ten thousand people threw rocks and only ten were needed to kill Chen. We might say that ten of the ten thousand rocks killed him, but which ten? The full set of rocks struck him at the same time and thus the full set of rocks counts as those that caused the threshold to be passed. This idea is supported by an account of causation that holds that X causes Y if X is a necessary element of a sufficient set (NESS) of actual antecedent conditions for Y 's occurrence.¹² This "NESS test" could help explain why an act can make one a causal contributor to an outcome in cases where there are thresholds, even if one's individual act was not necessary for crossing that threshold.

For example, suppose it takes three votes for a measure to pass, but four people (A , B , C , and D) vote. In that situation, each person's vote was neither

11 Harman, "Eating Meat as a Morally Permissible Moral Mistake," 219.

12 Barry and Øverland, "Individual Responsibility for Carbon Emissions"; Hart and Honoré, *Causation in the Law*; Wright, "Causation in Tort Law."

sufficient (one vote is not enough) nor necessary (three votes would have been enough) for the measure to pass. There are four sets of possible actions that are minimally sufficient for the outcome: *ABC*, *ABD*, *ACD*, and *BCD*. Still, what happened was that *ABCD* voted. While *A* could claim that their vote is not necessary for the outcome since *BCD* also voted, *A*'s vote was still a necessary part of the actual set of actions that was sufficient for the outcome (*ABCD*). If *A* had not voted, the actual set of actions that would have been sufficient would have been different: *BCD*. *A*'s vote therefore still counts as "a cause" of the outcome.¹³

Similarly, let us assume that ten thousand purchases are a sufficient condition of the success, extent, or severity of Big Box's plan, but Big Box has ten million actual customers, one of whom is Carla. Carla could claim that her purchase is neither necessary nor sufficient for the success, extent, or severity of Big Box's plan, and that she is therefore not a cause of what they do and is not complicit for that reason. Yet, her action is a necessary part of the actual sufficient set of antecedent conditions for Big Box's plan, just as each voter is in the previous example. If she did not purchase, there would have been a different sufficient set of causes instead. Thus, Carla's purchase can be said to be part of the cause of, or make a causal contribution to, Big Box's wrongdoing.

Another way of understanding why both a rock thrower and Carla count as causal contributors to an overdetermined outcome is with the notion of modal security, which is how Lepora and Goodin understand such cases. They differentiate causal "essentiality" from "potential essentiality." An act is essential when it is a necessary condition for an outcome "in every suitably nearby possible world" or when it is an "individual difference-maker."¹⁴ An act is potentially essential if it could be a necessary condition for an outcome along "some (but not all) possible paths" by which the outcome might occur.¹⁵ In other words, at a time antecedent to the outcome, there "is some suitably nearby possible world in which the act will individually make a difference," even if it turns out not to do so.¹⁶

13 Of course, we might say that the set of *A*, *B*, and *C* voting, plus *E* kicking a rock, would have also been sufficient for the measure passing, and *E* is a necessary element of that set. But we do not think *E* is part of the cause of the measure passing. In response, when we are looking for the set of causes of an outcome, we should focus on the sorts of actions that are of a type that would normally be either necessary or sufficient for that outcome. Kicking a rock is usually itself neither necessary nor sufficient for a measure based on voting to pass. We should therefore look to those who voted as a cause. Still, I am not here defending the NESS account as the correct or best account of causation. Rather, I am arguing that it tracks moral intuitions in both the rock-throwing and *Big Box* cases.

14 Lepora and Goodin, *On Complicity and Compromise*, 61–64.

15 Lepora and Goodin, *On Complicity and Compromise*, 64.

16 Lepora and Goodin, *On Complicity and Compromise*, 64.

Overdetermination causes like that of the rock throwers represent potential essentiality. Each rock throw is potentially essential to the death of Chen because there is a not-too-distant possible world where it is true of any given rock thrower that his throw hits, along with nine others, while all the rest miss, so that his rock becomes the threshold-crossing rock. Prospectively, then, each rock throw provides a sort of modal security to the passing of the relevant threshold. It is therefore fair to say that each rock throw makes a relevant causal contribution to Chen's death. Similarly, there is a not-too-distant possible world where 9,999 people have also, unbeknownst to Carla, decided not to buy the dress, such that her choice not to buy the dress turns out to make a difference to Big Box's wrongdoing.

I think we can therefore conclude that, appearances to the contrary, Lepora and Goodin's definition *does* find Carla to be morally complicit even though purchasing the dress will make no difference to Big Box's wrongdoing. First, Carla's act was voluntary. She consciously chose to buy the dress. Second, Carla meets their basic epistemic condition, because she knew or should have known that Big Box was engaged in wrongdoing. After all, looking up the Boycott Big Box website is not too epistemically taxing. Or, alternatively, if we think Carla's epistemic position is such that she should not have known that Big Box was engaged in wrongdoing, we might conclude that she is morally complicit but not morally responsible for being so.¹⁷ Either way, third, and importantly, Carla's purchase makes a causal contribution to Big Box's wrongdoing.¹⁸ Just

17 For example, suppose Carla buys her dress from the FairClothing website, which sources fair-trade dresses, and in so doing she reasonably believes that she is not contributing to labor exploitation. In fact, FairClothing is a scam and buys dresses from Big Box and sells them for twice the price to suckers like Carla. Intuitively, it does not seem as if Carla is complicit, even though her purchase does make a causal contribution to Big Box's labor exploitation. We can see that in this case it is the epistemic element that seems to make a difference to complicity. If Carla learns about this and continues to buy dresses from FairClothing, only then does she become complicit in what it does, even though her causal contribution is still the same.

18 This may be thought to lead to counterintuitive implications. First, if Carla buys an *ethically sourced* product from Big Box, this may also make a causal contribution to the wrongful parts of its plans and activities, even though her purchase was not wrongfully produced. Yet, we may have the intuition that Carla is not complicit with its wrongdoing because she causally contributed to the ethically appropriate part of its activities. Second, and similarly, to avoid complicity with Big Box, it seems like Carla should boycott the company completely, thereby ensuring that she does not make a causal contribution to the wrongful part of its plans and actions. But these implications of the notion of complicity as causal contribution are not as counterintuitive as they might seem. Take another case where an agent is engaged in both morally blameworthy and praiseworthy actions. Let us say a government is engaged in ethnic cleansing in one part of the country even as they also provide excellent support for the arts. An agency that provides funding for the arts donates

how morally complicit Carla is, and if and why her doing so is wrong, is a topic I turn to later in the paper. But this discussion helps us see how the most basic causal element of Lepora and Goodin's conception of complicity is met even in overdetermination cases.

3. COMPLICITY AS CAUSAL CONTRIBUTION VS. COMPLICITY AS JOINING IN

The reason that Carla's purchase makes her complicit with Big Box's wrongdoing is that Carla knows Big Box is engaged in wrongdoing and her purchase makes a causal contribution to its actions. Other theorists, however, have argued that the reason Carla is complicit even though her purchase does not make a difference is because, by making the purchase, she joins in or participates with the wrongdoing that Big Box, or other customers, are doing.¹⁹ While there is a sense in which I think this is correct, theorists should be careful to remain focused on the element of knowingly making a causal contribution as opposed to any more robust notion of joint or group action. This is because it seems unlikely that most people make a purchase as a way of joining in a group action whose consequences they thereby become responsible for.

One can see the promise of a joint- or group-action approach to understanding Carla's complicity, but it is important to recognize that there is a difference between an aggregate set of uncoordinated individual actions and a joint or group action. For example, someone might want to go to Chicago and decide to buy a train ticket and board and ride the train to get there. But this is different than three people deciding to go to Chicago via train together. Even so, there is debate about how to understand joint or group action. We might think that it requires relatively robust conditions, such as mutual knowledge of each other, mutual responsiveness to each other's actions, and each individual having the intention that the group act together. I do not delve into this debate here but focus on Christopher Kutz's account of joint action to test the claim that Carla joins in with a group action via her purchase. This is because his account identifies minimal conditions for joint action, but also because it is

money to the government specifically earmarked for the arts, even as that agency knows the government is engaged in ethnic cleansing. I think it makes sense to say that the agency is complicit with the government's wrongful action and that support for struggling artists is not a sufficient good to justify its complicity. This could be because donating money to the government for the arts adds a sort of legitimacy to its actions as a whole and this may also be understood in terms of making a causal contribution to the wrongful parts of its actions.

19 Driver, "Individual Consumption and Moral Complicity"; Martin, "Factory Farming and Consumer Complicity"; McPherson, "Why I Am a Vegan (and You Should Be One Too)" and "The Ethical Basis for Veganism"; Schwartz, *Consuming Choices*.

often appealed to in discussions of complicity.²⁰ The element of his account that is initially attractive is his claim that those who participate in a joint or group action can become morally responsible for the consequences of that action even if their joining in makes no difference.

Kutz argues that a set of individuals is engaged in a joint action when each individual (i) has a sufficiently overlapping conception of a shared goal, *G*, and (ii) each performs a voluntary action that aims to contribute a part toward the realization of *G*.²¹ On this view, a group of individuals is taking the train to Chicago together when each of them shares a sufficiently overlapping conception of their goal and each performs an action toward the realization of that goal—say, one looks up the schedule and tells the others, another buys the tickets, and someone else drives them to the station. Or, picking up an example from Kutz, the flight crews who dropped bombs on the city of Dresden in World War II are engaged in a joint action because each has an overlapping conception of *G* (drop bombs on Dresden) and each performs acts toward the realization of *G* (loading bombs, taking off, dropping bombs, etc.).

An individual joins in with the joint action in virtue of their conception of the goal and the aim of their attendant actions. Moreover, each individual who joins in with the group action shares some moral responsibility for the outcomes of that action, even if those outcomes were overdetermined. That is, let us say that whether or not an individual participates in the bombing, the effects would be just as bad. Still, if an individual does join in, he becomes morally responsible for those bad effects, even though he made no difference to them. This is because of how he structures his will; he still shared the goal and joined in with others. Finally, it is important to note that someone can join in with a group action without intending or wanting *G* to be accomplished. For example, someone might be strongly against the bombing but join in because of a threat of dishonorable discharge. All that is needed is that he shares an overlapping conception of *G* with those also engaged in the joint action and voluntarily performs actions that aim to contribute to the achievement of *G*. “Voluntarily” here means meeting basic conditions like consciousness and control of one’s actions.

Returning to the consumer cases at hand, our first possibility is that, by making a purchase, Carla participates in a joint action *with Big Box itself*. It is true that Big Box employees are engaged in a joint action, but becoming a

20 Nefsky, “Fairness, Participation, and the Real Problem of Collective Harm,” 110.

21 Kutz, *Complicity*, 74.

customer of Big Box does not seem to meet the relevant conditions.²² By purchasing, does Carla join in with the group action of Big Box's employees? As David Schwartz explains, to think this we would have to understand Carla's purchase as an "act of intentional participation in a collective action that encompasses the production, marketing, distribution, and consumption of particular consumer products. The collective [would be] . . . thus extended beyond those employed to include those who purchase the output of the manufacturing operations. This intentional participation [would make] . . . the consumer complicit in . . . any blameworthy (or praiseworthy) actions of that collective."²³

In fact, this is just how Schwartz argues that we should think of Carla's purchase. Her act "is essential to the successful collective activity of garment manufacture, marketing, and sales; in fact, consumer participation is the primary motivation—and sole criterion of success—for the entire effort."²⁴ But as a customer it is unlikely that Carla shares the goal or aim that makes it such that the employees of the company are engaged in a joint action. Her aim in buying the dress is to have and enjoy the dress, not to arrange the production or sourcing of various products to sell them to consumers to create a profit. Moreover, she does not buy the dress as a way of contributing to the realization of that goal. Schwartz replies to this issue by suggesting that Carla does have a strong identification with the aim or goal of the company's joint action because she identifies as a consumer.²⁵ Yet, even if she does strongly identify with her role as a consumer, this is not sufficient to provide an overlapping goal with Big Box's employees such that she is engaged in a joint action with Big Box itself, which, as I have suggested, is about selling items to consumers with the goal of making a profit.

A second possibility is that Carla becomes complicit in Big Box's wrongdoing because she joins in a group action *with Big Box's other consumers*. According to Adrienne Martin, that consumer group is complicit in Big Box's wrongdoing

22 First, to count as employees, they need to sufficiently share the overarching goal of arranging the production or sourcing of various products and selling them to consumers to create profit. Second, each employee performs actions that aim to contribute a part toward the realization of some number of sub-goals that in turn aim to contribute a part toward the realization of Big Box's overarching goal. As individuals engaged in joint action, following Kutz's joint-action approach, each employee bears some moral responsibility for the foreseeable harm and wrongdoing that results from the pursuit of their shared goal, including the use of slaves to produce some of the products. Some individuals may bear greater moral responsibility depending on a range of epistemic, causal, and moral factors, but Kutz argues that all the employees will bear some moral responsibility for that outcome.

23 Schwartz, *Consuming Choices*, 74.

24 Schwartz, *Consuming Choices*, 82.

25 Schwartz, *Consuming Choices*, 83.

because it does make a difference to its production decisions. Martin follows this approach in her discussion of purchases of meat from industrialized farms. She acknowledges that no consumer makes a difference to these farms' wrongdoing, and so no individual is complicit as an "accomplice."²⁶ Yet, she argues that, by making a purchase, a customer joins in with the group of consumers who are also making that purchase, a group whose aim or function is to signal demand. That group aids and abets the wrong that industrialized farms do. The individual consumer is complicit with that complicity because she voluntarily joins in and becomes a member of the group that does make a difference.

Martin does not expressly state how they join in simply by making a purchase or what conditions someone must meet in order to count as joining in with what a group does together. But Martin seems to think that someone who joins in shares responsibility for the effects of what the group together causes, whether or not her participation makes any difference to that outcome. She states that "it is not a causal relation doing the moral lifting" in such cases, and that "what matters in such cases is *participation* but describing participation as causal doesn't get at the heart of what matters, morally speaking."²⁷

In the previous approach there was a group or joint action taking place (the actions of Big Box's employees), but here it is not clear that there is any group agency or joint action. Individuals engaged in a joint or group action must at least share an overlapping conception of a shared goal, but it is not clear that a set of consumers of a particular product will do so simply in virtue of the fact that they buy it. In reply, we may be able to say that each has the goal of buying something they need or wants. We might further suggest that this shows that they have an overlapping conception of a shared goal, perhaps something like Martin's suggestion that the group's shared goal is the fulfillment of consumer demand. But I think it is unlikely that Carla has other consumers in mind when she buys the dress, or that, when she buys the dress, she has in mind the goal of fulfilling consumer demand or contributing to the function of the market economy. Moreover, she likely does not buy the dress *as a way* of contributing to any of these putative shared goals; rather, she makes the purchase as a way of fulfilling her own needs or wants. Thus, it is not clear that Carla joined in with what other consumers were doing in any robust way.²⁸

26 Martin, "Factory Farming and Consumer Complicity," 205.

27 Martin, "Factory Farming and Consumer Complicity," 209.

28 Julia Driver also looks to Kutz's idea of intentional participation to address the "no difference" issue. Yet, Driver moves away from a view of complicity that focuses only on an agent's intentions and seems to recognize that there must be some sort of causal element for a purchase to count as participation. She says that an action that "*could never* make a causal contribution" would not count as a participatory act. Moreover, even though

My causal-contribution approach, however, supports the shared sense that by making a purchase Carla is “participating in” what Big Box does, but it does so in a more minimal sense. We can see this by relating my approach to Tristram McPherson’s understanding of complicity. On his view, Big Box is engaged in wrongdoing and has a plan to continue to do so that involves a “pattern of goals.”²⁹ The execution of that plan requires others playing various roles, including the role of customer. By making a purchase, Carla is “knowingly and voluntarily fulfilling a [key] role that needs to be fulfilled” for that wrongful plan “to work.”³⁰ She thereby becomes complicit with the plan. McPherson says it is “plausible” that doing so is wrong even if it does not at all help “the success of the plan.”³¹ McPherson’s account captures the intuitive sense that making a purchase is a form of participation in, or cooperation with, the actions and plans of a company, including its wrongful plans. By performing the action of purchasing, Carla is knowingly playing a role in the wrongdoing, even if she is not literally joining in a group action with the company.

My account supplements this by bringing out the causal element of Carla playing that role. That is, someone might wonder how exactly Carla counts as playing a role in the wrongful plan if whether she does so or not will have no effect on that plan. A joint-action account like Schwartz’s or Martin’s would allow us to point to Carla’s goals or aims in response, but I have argued that is not tenable. What else can we point to in order to explain why Carla counts as “filling a slot” in Big Box’s plan besides the causal relationship between her doing so and the plan itself? Moreover, if it really were true that Carla’s purchase not only has no actual effect on the plan, but *could never* have any effect, it is not clear why her purchase would count as “playing a role” or “participating.” Finally, by assumption, Big Box does not need Carla to play her role since there are already enough others who are doing so. We need the conception of causal contribution I presented above to explain why playing a role in a plan whose success is massively overdetermined still counts as making one morally complicit.

someone’s meat purchase might have no effect on meat production, to say that the purchase makes her a participant in the wrongdoing, the act of purchasing must be the sort of action “that can make causal contributions to similar outcomes.” My discussion of the way a purchase makes a causal contribution to wrongdoing, even if it makes no difference, extends this part of Driver’s discussion of participation complicity. Driver, “Individual Consumption and Moral Complicity” 72–73.

29 McPherson, “Why I Am a Vegan (and You Should Be One Too),” 83.

30 McPherson, “The Ethical Basis for Veganism,” 19.

31 McPherson, “The Ethical Basis for Veganism,” 20.

To conclude, recognizing that Lepora and Goodin's view of complicity as causal contribution maintains traction in overdetermination cases has several important benefits. First, it explains why Carla's purchase makes her complicit without relying on identifying a joint or group action or showing that Carla is joining in that action. In general, it is important to recognize that someone can be complicit with the wrongdoing of others even if she does not actually join in with what they do. Structuring one's will to join in may be morally worse, but one can be complicit simply by making a causal contribution. Second, if Carla's purchase really did make absolutely no causal contribution as some have mistakenly assumed, it would be difficult to see why it would make her complicit. Intuitively, one is complicit with another when one is "part of the story" of what another person does. How can one be part of that story if one makes absolutely no causal contribution to what they do, or if there is absolutely no chance that one's action might aid or help them?³² If the core notion of complicity involves aiding or abetting the wrongdoing of another, it seems like one cannot be complicit with wrongdoing if one's action (or omission) makes absolutely no causal contribution.³³

Finally, insofar as Carla meets the other epistemic conditions on Lepora and Goodin's definition of complicity—she should have known that her act would contribute to wrongdoing—my approach allows us to make sense of how Carla

32 Gardner, "Complicity and Causality."

33 In response, someone might point out the following case that suggests that one can be complicit even if one knows *ex ante* that there is no possibility for one's act to have an effect. Imagine a group of Nazis has a wrongful plan to work together and coordinate their actions in order to promote white supremacy online and get new adherents. One self-identified Nazi is reviled by all of the others, and what he posts is always ignored by all of the others and ineffective at getting new adherents. His posts therefore do not make any difference or make any causal contribution to the wrongdoing that the others do. Still, by making his posts, he seems to be complicit with the wrongdoing of the other Nazis because he is still knowingly trying to play a role or "fill a slot" in the plan. In response, first, I would say that what the reviled Nazi does is more akin to attempted complicity than actual complicity. He intends or hopes to make a causal contribution to the success of what the others do and shares their goal. Yet, his complicity is attempted because he fails to achieve that goal. Attempted complicity may itself be wrong, but it is not wrong because of the causal contribution it actually makes. Second, we should be weary of trying to load too much into the notion of complicity. As Lepora and Goodin point out, there are some ways of acting in relation to another's wrongdoing that are morally worse than mere complicity (e.g., helping to plan or joining in with that wrongdoing) and some that are morally less bad (e.g., simply being in the same room as a known wrongdoer). Lepora and Goodin's conception of complicity as causal contribution gives us an expansive notion of complicity but one that still has some limits. If an action that could *never* have any effect on wrongdoing is to count as making one complicit, this would expand the notion to even further cases and also make it hard to know what counts as complicity and what does not.

might still be thought to voluntarily “participate” in what Big Box does (even if she does not literally join in with them in doing it). Her purchase contributes money that enters Big Box’s revenue stream. That causal contribution distinguishes her even from someone who refuses to purchase but still retrieves a discarded dress from the dumpster behind Big Box. As Driver points out, if it really was the case that Carla’s action made absolutely no causal contribution to what Big Box does, then by purchasing the dress she could, like the dumpster diver, also claim to just be consuming “waste” from the system.³⁴ Intuitively, however, by making a purchase Carla is a “participant in wrong-doing” in a way that the dumpster diver is not.³⁵ The dumpster diver is not a “participant” because their action (most likely) does not make a causal contribution to Big Box’s wrongdoing.

To conclude, my approach applies the most intuitive definition of complicity and explains why Carla’s act makes her complicit. As we will see, however, this does not fully address the problem of individual causal inefficacy for consumer ethics. We will need to better understand the moral reasons there are to avoid complicity and whether they support the intuition that, just as Aisha should not buy a dress from Bob’s Dresses, Carla should not buy one from Big Box.

4. WHY IS IT WRONG TO BE COMPLICIT?

Let us assume that Lepora and Goodin’s epistemic and causal conditions for responsibility for an act of complicity are met in *Big Box*. Acts of complicity are morally wrong and there is a defeasible moral reason that one should not perform such acts (such acts are, in Lepora and Goodin’s words, “*pro tanto* wrong”).³⁶ But how wrong is it to be morally complicit? When is it wrong all things considered to be so? In this section, I describe the problem that the *Big Box* case poses for Lepora and Goodin’s answers to these questions. I then move on to discuss other moral reasons why we should try to avoid complicity in such cases. I argue that there are a number moral reasons to avoid complicity besides the reason to avoid making a causal contribution to the wrongdoing of others.

The issue arises when we recognize that, while Carla’s purchase can be said to make a causal contribution to wrongdoing, that contribution is likely much smaller than it may seem. It is a complex and difficult empirical question whether any market will have the relevant thresholds, and, if it does, whether

34 Driver, “Individual Consumption and Moral Complicity,” 76.

35 Driver, “Individual Consumption and Moral Complicity,” 76.

36 Lepora and Goodin, *On Complicity and Compromise*, 97.

they are not offset by factors that greatly minimize the probability that any particular purchase will have an effect. For example, in some markets there will also be buffers that shield production decisions from sales. Mark Budolfson gives an imagined example of such a buffer. Say a producer has always produced twenty thousand shirts each month and has always sold between fourteen thousand and sixteen thousand shirts each month. If the producer would happen to sell eighteen thousand shirts one month, in response he would produce twenty-five thousand shirts the next month. So, there is a genuine tipping point in the relationship between sales and production; namely, between 17,999 and 18,000 sold. But that threshold has never once been reached. Let us say that both consumers and producers know this. If so, the probability of one more T-shirt purchase having an effect is far less than one in eighteen thousand, since there is an incredibly low expected probability of any more than sixteen thousand shirts being sold that month. As Budolfson says, “we can know enough about supply chains . . . to know that threshold effects are not sufficiently likely and are not of sufficient magnitude to drive the expected effect of consumption anywhere close to the average effect.”³⁷

If it is reasonable to believe that many markets have these sorts of buffers, it is reasonable to believe that the chance that an individual’s purchase will signal a shift in demand or have an effect on a producer’s plans is much less than the average probability and, in large markets, probably vanishingly small.³⁸ Moreover, Lepora and Goodin argue that the amount of causal contribution decreases as the number of people also contributing increases, and as the number of “chance and choice nodes yet to come in the causal chain before the planned wrongdoing” increases.³⁹ Despite these problems, even if the causal contribution Carla’s purchase makes is very small, it still counts as a causal contribution to Big Box’s wrongdoing. She is therefore still complicit. But how morally wrong is it to be only a little bit complicit?

For Lepora and Goodin, complicit acts are a moral issue because of their consequences: they causally contribute to the success, extent, or severity of the harm or wrongdoing that others do. When such contributions are made voluntarily and knowingly, agents become morally responsible for their complicit contribution. They assess the wrongness and blameworthiness of an act of complicity in terms of both the moral badness of the principal wrongdoing

37 Budolfson, “The Inefficacy Objection to Consequentialism and the Problem with the Expected Consequences Response,” 1717.

38 Budolfson, “The Inefficacy Objection to Consequentialism and the Problem with the Expected Consequences Response”; and McPherson, “Why I Am a Vegan (and You Should Be One Too)” and “The Ethical Basis for Veganism.”

39 Lepora and Goodin, *On Complicity and Compromise*, 67–68.

and the causal contribution the act makes to that principal wrongdoing. They argue that these two elements interact to shape wrongness and blameworthiness. Specifically, they say that we should multiply them together, allowing the amount of causal contribution to be a “discount factor” on the moral badness.⁴⁰ Holding the moral badness of the principal wrongdoing fixed, they put moral wrongness in a direct, linear relationship with the amount of causal contribution made by the complicit agent.

I also think that the amount of causal contribution an act of complicity makes is relevant to our moral judgments about complicit acts. Making a large causal contribution does seem to make an act of complicity morally worse. Also, moral complicity seems to be a scalar concept since we talk about some people being more complicit than others.⁴¹ For example, let us say that a government permits Big Box to utilize human slaves. We might think what that government does makes them more complicit with Big Box, and is morally worse, than what Carla does. Assuming no complicit agent intends the wrongdoing of the principal agent, this way of talking seems to be tracking how much of a causal contribution someone makes.

Yet, we can see that Lepora and Goodin’s normative analysis of complicity purely in terms of its morally relevant consequences may not help us to explain why Carla’s dress purchase is wrong and something she should not do. As we have seen, her causal contribution is very small. If the only reason complicity is wrong is because of the consequences it has for the harm or wrongdoing done by others, then her purchase of the dress may not ever rise to the level of a moral wrong. The incredibly small causal contribution would serve as a strong discount factor on the moral badness of the principal agent’s act, leading to the act being only a little bit wrong.

On the other hand, the use of slavery in the production of dresses is an egregious moral wrong, so multiplying Lepora and Goodin’s two factors together does increase the *pro tanto* moral reason for her to avoid complicity beyond an infinitesimal weight. That is, it does seem morally worse to make a very small causal contribution to those engaging in human slavery than it is to, say, help someone cheat on a physics test. The issue is that, even in this case, a very small causal contribution to an egregious moral wrong still seems not *that* wrong, as the weight of its moral reason remains very small, at least on Lepora and

40 Lepora and Goodin, *On Complicity and Compromise*, 111–12.

41 This is another reason the conception of complicity as playing a role in another’s wrongful plan needs to be supplemented with an understanding of complicity as causal contribution. We think that playing some roles will make one more complicit than playing other roles. Leaving aside the issue of intending the wrongdoing, what could this be tracking except how much of a causal contribution one makes?

Goodin's understanding. Moreover, if making a very small causal contribution to the death or serious harm to others was very wrong, we may find many everyday activities to be morally wrong, and perhaps impermissible all things considered. For example, when I turn on my gas oven, I create a very small risk of death to others. Although the outcome could be death for others (a very morally bad result), it is not clear that it is wrong to turn on my oven, because the risk I create is very small.

In reply, we might say that even though Carla's contribution is very small, that contribution still makes her morally complicit, is a contribution to an egregious moral wrong, and, thus, should be thought of as being wrong. It is simply, *pro tanto*, not something she should do. The issue with this response is that it is not clear that being a little bit complicit with an egregious moral wrong will often be all-things-considered wrong. The problem is that the moral reason against being very minimally complicit will still be weak and may often be outweighed by any of Carla's morally relevant reasons for making the purchase, even in cases where we have the intuition that she should not buy the dress.

To reflect on this further, let us assume that Carla could go on the FairClothing website and order a dress that was produced by well-treated workers. In a first example, let us also assume the dress was just as nice and the same price. In that case, there do not seem to be any morally relevant reasons for her to buy the slave-made dress from Big Box. Morally, all things considered it would be wrong for her to do so. In a second example, let us assume that the FairClothing dress is three times the price, so things are not equal. Carla is lower income and uses that income to meet her family's needs. She is buying the dress for her daughter's first day at school. Paying three times the price on FairClothing would put a morally relevant burden on her and her family, though in a sense they could afford it. We have the weight of a financial burden to a family on a limited budget weighed against a very small causal contribution to human slavery. In this case, we might think that, all things considered, it is not wrong for Carla to buy the dress. Buying the dress is not something Carla morally should not do.

In a third example, let us say that if Carla buys the dress from FairClothing as opposed to Big Box, the only burden on her is paying three times the price but she could reasonably afford to pay it. Buying the FairClothing dress would leave her less money to spend on other things or lead to a loss of personal enjoyment. Most of us think moral reasons are weightier than personal, nonmoral reasons, even if we do not think that moral reasons have absolute lexical priority over personal, nonmoral ones. We might think that Carla's moral reason not to be infinitesimally complicit outweighs the personal, nonmoral burden of giving up a thing she might like better or paying more than otherwise. In that case, despite Lepora and Goodin's causal discount factor on the moral

wrongness of complicity, it would be wrong all things considered for Carla to purchase the dress if the cost of refusing to do so were only reasonable personal burdens. This would explain why, in this example, Carla morally should not buy the dress because doing so would make her wrongfully complicit despite her minimal causal contribution. This supports intuitive judgments about Carla's moral complicity.

On the other hand, someone might object that even these reasonable personal burdens outweigh the moral wrongness of making a very small causal contribution to an egregious moral wrong. From the perspective of morality, Carla's happiness and personal projects matter. Does morality require her to give up a bit of personal happiness to avoid making a tiny causal contribution to wrongdoing that she does not intend and that others do? While the burden to Carla of buying the FairClothing dress is small, that is a real burden to her and may seem weightier than the burden that buying the Big Box dress imposes on another, which is an infinitesimal causal contribution to wrongdoing or harm to them. If it is morally wrong not to bear even relatively small personal burdens to avoid making infinitesimal causal contributions even to serious harms or wrongdoing, we may find ourselves overburdened with moral wrongs.

I am not sure if these worries are sufficient to undermine the ability of Lepora and Goodin's normative conception of complicity to explain why it would be wrong and morally blameworthy all things considered if Carla were to buy the dress in the third example. We often still appeal to the notion of complicity even when we realize that an act makes very little causal contribution to harm or wrong, and in so doing what we seem to be trying to emphasize is not how much we are contributing to a harm or wrong, but that continuing to make that contribution once we believe or should have believed that harm is taking place says something about ourselves, what we choose to participate in, and our relationship to those harmed or wronged. I therefore think it is valuable to explore the moral reasons that might explain why we should try to avoid intentional acts of moral complicity, even if we accept Lepora and Goodin's claim that the amount of causal contribution serves as a discount factor on moral wrongness. I will suggest that there are two sets of moral reasons an agent might have to avoid knowingly contributing to the wrongs that others do: self-regarding and other-regarding.

Agents have a self-regarding moral reason to avoid being knowingly complicit in the wrongdoing that others do when that wrongdoing runs counter to their moral values and commitments. For example, Kwame Appiah asks why some think that a particular organization morally should divest from companies that are engaged in wrongdoing even if that divestment will have little to no influence. He suggests that the reason for some is the desire to avoid the moral

taint of having dirty hands: "We are associated, through our ownership of the shares, with a wicked system; we play a part in it. Our holding of the shares taints us."⁴² He says that if we are to make any sense of a moral requirement to avoid moral taint, it must be in terms of a tension between one's moral identity and what one chooses to be associated with. This suggests that this taint can undermine moral integrity.⁴³

Similarly, Marina Oshana argues that the reason an agent should avoid being associated or complicit with wrongdoing and wrongdoers is that she should aim to live authentically with respect to her commitments and self-understanding.⁴⁴ If we follow this line of thought, we can say that if Carla has a fundamental belief that slavery is an egregious moral wrong (as she should), choosing to continue to be complicit with Big Box undermines her personal integrity and authenticity. Moreover, failing to live with personal integrity and authenticity may make Carla worse off in the long run. It would be rational for her to give up the enjoyment of buying the dress and pay more for the fair-trade dress in order to promote her own well-being.

Avoiding being complicit with wrongdoers can also help one develop a virtuous moral character. Discussing someone who purchases meat, Julia Driver emphasizes that what someone decides to participate in can display something about their character. First, Driver suggests that from a virtue-consequentialist perspective, if individuals develop the character trait of knowingly making small contributions to others' wrongdoing, this may systematically produce bad overall effects.⁴⁵ Carla may become more indifferent to wrongdoing in general and so may end up being complicit even in situations where she makes a relatively large causal contribution to the wrongdoing of others. Put another way, a world of perfect Carlas who only make small causal contributions to the wrongdoing of others is morally worse than one in which people always refuse to buy slave-made products. Therefore, Carla should not make a habit of knowing complicity with the wrongdoing of others, however small. Second, Driver also suggests that on a virtue-intrinsic approach the character trait of knowing complicity with wrongdoers means that one has an orientation toward good and bad that is itself intrinsically bad. She explains that "virtue involves having the right kind of orientation or attitude toward good and evil," and complicity with wrongdoing may be one such improper orientation.⁴⁶

42 Appiah, "Racism and Moral Pollution," 190.

43 Appiah, "Racism and Moral Pollution," 193.

44 Oshana, "Moral Taint," 368.

45 Driver, "Individual Consumption and Moral Complicity," 76.

46 Driver, "Individual Consumption and Moral Complicity," 76–77.

Yet, there seems to be more involved in avoiding complicity than just a moral concern about integrity, authenticity, and moral character. By contributing to another's wrongdoing one becomes related to the victims of the wrongdoing. While one might care very deeply about their plight, the decision to knowingly contribute can express moral disrespect and disregard. Since all persons are moral equals, moral agents owe each other forms of treatment (actions) and regard (attitudes) that instantiate and express equal respect and mutual recognition.

There are two elements to consider regarding an agent's attitude toward others. The first concerns the attitudes or forms of regard themselves. The second concerns how the agent's actions express those attitudes within the relationship. This is the "meaning" of those actions for others or the significance others have reason to assign them.⁴⁷ Carla should have an attitude of serious moral regard and concern for the plight of the slaves whom Big Box sourced to produce dresses. Choosing to remain complicit with Big Box's wrongdoing even once she reasonably believes, or should reasonably believe, that they are engaged in wrongdoing is disrespectful toward the slaves and expresses a level of disregard for the egregious wrong that is being done to them.

To see how the element of moral disrespect and disregard relates to an act of knowing complicity, we can think about a case where the agent is strongly complicit because the causal contribution is relatively large. For example, Bahani works at the only knife shop in town. Ryder comes in to buy a knife and is acting and speaking in a way such that Bahani reasonably believes that Ryder is going to use the knife to go on a stabbing spree. If Bahani refuses to sell the knife, it will be much more difficult for Ryder to get one. While Bahani does not commit the stabbings, he becomes strongly complicit with them because his action is a large causal contributor (an almost necessary condition) for Ryder's wrongdoing. Given this, part of why Bahani's act is wrong is that it is causally responsible for a large part of the wrongful harm that Ryder does. Not only does it make him strongly complicit with Ryder's actions, but it also seems to "cause harm," even though it does so by helping Ryder cause it. But selling the knife is also wrong because the voluntary decision to still sell the knife despite what Bahani should have believed was disrespectful toward Ryder's potential victims. Imagine one of the victims meeting Bahani after the incident. Bahani may say that he really cared about the plight of Ryder's potential victims and was strongly against Ryder's plan. All of that may be true, yet he still decided to become complicit, which suggests that he did not care as much as he might have.

In this case, Bahani's act makes a relatively substantial causal contribution to Ryder's wrongdoing. His act causes morally relevant harm, makes him

47 Scanlon, *Moral Dimensions*, 54.

complicit with Ryder's wrongdoing, and expresses disrespect and disregard for Ryder's potential victims. Returning to Carla, since her causal contribution is infinitesimal, she is only minimally complicit, and she is only very minimally causally responsible for the harm or wrong done to the slaves. Still, like Bahani, Carla voluntarily and knowingly chose to remain complicit in Big Box's wrongdoing when she could have avoided it at a reasonable cost to herself. While we lose the idea that her complicity causes harm, or makes her strongly complicit, her choice may still express disrespect and disregard for the plight of the slaves.

I do not think that acting in ways that may express disrespect and disregard in such cases rises to the level of a clear moral wrong. Rather, I think it is better to understand it as a form of offense.⁴⁸ The moral reason to avoid even small amounts of complicity with wrongdoers is a "negative" mirror image of "positive" supererogatory moral reasons that has to do with legitimate expectations beyond the realm of rights and duties.⁴⁹ Performing actions that develop and express the attitude of moral respect and regard for the plight of others are just such moral expectations.

There are at least two issues with the claim that Carla's action is disrespectful to those wronged. First, there is the issue of hidden actions. Let us say that Carla buys the dress but does not tell anyone and no one ever finds out, least of all the slaves far away who made it. Why should we think that Carla's action "expresses disrespect" if the action never actually expresses anything to anyone? In reply, it does seem like an act can be disrespectful even if no one is actually disrespected by it. Bahani's action of selling the knife to Ryder given his beliefs about Ryder's plans suggests that Bahani fails to take into sufficient account the plight of potential victims. This is the case even if Ryder decides not to use the knife to harm others.

The second issue is more difficult. Let us say that Carla does care deeply about the plight of the slaves and has joined Students against Sweatshops. Still, she buys the dress because she realizes doing so does not cause harm and makes her only infinitesimally complicit with their situation. But she also does many other things that express moral respect and regard for those enslaved. Why then would the purchase express moral disregard for their plight? To reply to this issue, we only need recognize that having and expressing moral respect and regard for others is a scalar notion. We could often have greater regard for the plight of others, and often perform further actions that express our care

48 Mellema, *Complicity and Moral Accountability*, 88–89.

49 Elizabeth Harman calls these "morally permissible moral mistakes." She argues that buying meat makes one part of the joint cause of the harms and wrongs of meat production. Even if one's causal contribution is infinitesimal, doing so is a morally permissible moral mistake ("Eating Meat as a Morally Permissible Moral Mistake").

and concern. In this situation, Carla has great moral respect and regard for the plight of those wronged by Big Box that she expresses in many of her actions. Nonetheless, she could express a little bit more by avoiding actions that make her even infinitesimally complicit with those engaged in the wrongdoing she so deeply cares about.⁵⁰

5. CONCLUSION

In the first part of this paper, I argued that one can become complicit with the wrongdoing that others do even if one's action does not make a difference to that wrongdoing. All that is needed is that one's act makes a causal contribution to what they do. Moreover, even if that causal contribution is very small, one still counts as complicit. In such situations, I have argued that there is a set of moral reasons to avoid complicity that joins with the moral reason to avoid making a causal contribution to the wrongdoing of others. Some are self-regarding reasons involving having integrity, authenticity, and moral character, while some are other-regarding reasons involving having and expressing attitudes of moral regard and respect for those harmed or wronged. These moral reasons join with the fundamental moral reason to avoid complicity that I argued for above. Carla has a moral reason to avoid complicity at a reasonable cost to herself. Even a very small causal contribution makes her morally complicit, implicating her in the story of the harm and wrongdoing that others do. Voluntarily choosing to continue to be complicit once she reasonably believes, or should reasonably believe, that she is implicated, when she could choose not to be complicit at a reasonable cost to herself, means she plays a role in the wrongdoing that others do. This fact has a moral valence that may outweigh any personal, nonmoral reasons she has to remain complicit.

That said, whether Carla's purchase is all-things-considered wrong because it makes her complicit will continue to depend on her other morally relevant reasons. Let us say that Carla could take the money she saves from purchasing the cheaper, slave-made dress and donate it to effective charities that make a

50 Take a related example. Felix is strongly against and cares deeply about the injustice faced by a particular social group. A political representative who has passed laws that are unjust toward that social group is visiting Felix's town for a parade and festival. Despite his beliefs and attitudes, Felix decides to attend the parade and festival. There are flags, food, and music, and it is a bit of fun. Later, Felix tells his friend who is a member of the relevant social group about his attendance. The friend is hurt by what Felix did and says, "If you cared so much about us, why did you go and support her?" Felix replies, "I am strongly against what she has done, but it was just a bit of harmless fun." Even so, it seems justifiable for the friend to reply, "I know you care about our situation, but you could have shown you care even more by not attending the parade."

direct causal contribution to minimizing suffering and wrongdoing. We might think in that case, even though she has a set of moral reasons to avoid moral complicity, she has a countervailing moral reason to continue to be complicit. Thus, buying the dress may not be all-things-considered wrong to do. Even so, it would certainly be morally better to donate to effective charities even while she minimizes her knowing complicity with the wrongdoing she cares deeply about.

Theorists have found it difficult to explain why consumers morally should not purchase from companies that are engaged in wrongdoing in situations where refusing to purchase would seem to make no difference to the success, extent, or severity of the companies' wrongdoing. I have shown how a complicity-based approach can help. We must recognize that a consumer becomes complicit not because they join in with a group action that is doing wrong, but because a purchase makes a causal contribution to the wrong that the company does, even if it is very small. The choice to knowingly make that contribution implicates the agent in the wrongdoing and that choice has moral relevance in relation to self- and other-regarding moral reasons.⁵¹

Georgian Court University
corey.katz@gmail.com

REFERENCES

- Appiah, Kwame Anthony. "Racism and Moral Pollution." *Philosophical Forum* 18, nos. 2–3 (Winter–Spring 1986–87): 185–202.
- Barry, Christian, and Gerhard Øverland. "Individual Responsibility for Carbon Emissions: Is There Anything Wrong with Overdetermining Harm?" In *Climate Change and Justice*, edited by Jeremy Moss, 165–83. Cambridge, MA: Cambridge University Press, 2015.
- Budolfson, Mark Bryant. "The Inefficacy Objection to Consequentialism and the Problem with the Expected Consequences Response." *Philosophical Studies* 176, no. 7 (July 2019): 1711–24.
- . "Is It Wrong to Eat Meat from Factory Farms? If So, Why?" In *The*

51 I would like to thank Alex Bearden, Bob Fischer, Bob Hartman, Trevor Hedberg, Tristram MacPherson, and Dan Shahar for excellent comments on previous drafts. Audiences at the American Philosophy Association Eastern Division, International Society for Environmental Ethics summer workshop, Society for the Study of Ethics and Animals, and University of Washington Climate Ethics Conference also provided important feedback. Finally, I would like to express deep gratitude for the work of the anonymous peer reviewers for this and other journals whose critical comments helped me to greatly improve this paper.

- Moral Complexities of Eating Meat*, edited by Bob Fischer and Ben Bramble, 80–98. Oxford: Oxford University Press, 2015.
- Chignell, Andrew, Terence Cuneo, and Matthew C. Halteman, eds. *Philosophy Comes to Dinner: Arguments on the Ethics of Eating*. London: Routledge, 2016.
- Driver, Julia. “Individual Consumption and Moral Complicity.” In *The Moral Complexities of Eating Meat*, edited by Ben Bramble and Bob Fischer, 67–79. Oxford: Oxford University Press, 2015.
- Gardner, John. “Complicity and Causality.” *Criminal Law and Philosophy* 1, no. 2 (May 2007): 127–41.
- Harman, Elizabeth. “Eating Meat as a Morally Permissible Moral Mistake.” In Chignell, Cuneo, and Halteman, *Philosophy Comes to Dinner*, 215–31.
- Hart, H. L. A., and Tony Honoré. *Causation in the Law*. Oxford: Clarendon Press, 1962.
- Hiller, Avram. “Climate Change and Individual Responsibility.” *Monist* 94, no. 3 (July 2011): 349–68.
- Kagan, Shelly. “Do I Make a Difference?” *Philosophy and Public Affairs* 39, no. 2 (March 2011): 105–41.
- Kingston, Ewan. “Shopping with a Conscience? The Epistemic Case for Relinquishment over Conscientious Consumption.” *Business Ethics Quarterly* 31, no. 2 (July 2020): 1–33.
- Kutz, Christopher. *Complicity: Ethics and Law for a Collective Age*. Cambridge: Cambridge University Press, 2000.
- Lepora, Chiara, and Robert E. Goodin. *On Complicity and Compromise*. Reprint ed. Oxford: Oxford University Press, 2015.
- Martin, Adrienne M. “Factory Farming and Consumer Complicity.” In Chignell, Cuneo, and Halteman, *Philosophy Comes to Dinner*, 203–14.
- McPherson, Tristram. “The Ethical Basis for Veganism.” In *The Oxford Handbook of Food Ethics*, edited by Anne Barnhill, Tyler Doggett, and Mark Budolfson, 209–40. Oxford: Oxford University Press, 2018.
- . “Why I Am a Vegan (and You Should Be One Too).” In Chignell, Cuneo, and Halteman, *Philosophy Comes to Dinner*, 73–91.
- Mellema, Gregory. *Complicity and Moral Accountability*. Notre Dame, IN: University of Notre Dame Press, 2016.
- Nefsky, Julia. “Fairness, Participation, and the Real Problem of Collective Harm.” In *Oxford Studies in Normative Ethics*, vol. 5, edited by Mark Timmons, 245–71. Oxford: Oxford University Press, 2015.
- Norcross, Alastair. “Puppies, Pigs, and People: Eating Meat and Marginal Cases.” *Philosophical Perspectives* 18, no. 1 (December 2004): 229–45.
- Oshana, Marina A. L. “Moral Taint.” *Metaphilosophy* 37, nos. 3–4 (July 2006):

353–75.

Scanlon, T. M. *Moral Dimensions: Permissibility, Meaning, Blame*. Cambridge, MA: Belknap Press, 2008.

Schwartz, David T. *Consuming Choices: Ethics in a Global Consumer Age*. Lanham, MD: Rowman and Littlefield Publishers, 2010.

Singer, Peter. "Utilitarianism and Vegetarianism." *Philosophy and Public Affairs* 9, no. 4 (Summer 1980): 325–37.

Wright, Richard W. "Causation in Tort Law." *California Law Review* 73, no. 6 (December 1985): 1735–1828.

ETHICAL VEGANISM AND FREE RIDING

Jacob Barrett and Sarah Zoe Raskoff

THE CREATION of animal products on factory farms causes animals a tremendous amount of pain and suffering. Learning about the severity and extent of this suffering often leads people to change their dietary choices or at least to feel some moral pressure to do so. Many therefore seem implicitly to accept that when deciding what to purchase and eat, animal suffering makes a significant moral difference. Moral philosophers often make this thought explicit, attempting to ground an obligation to go vegan in the horrifying consequences of factory farming.

It turns out, however, to be surprisingly difficult to explain how animal suffering generates any reasons to alter our dietary choices, much less an obligation to go vegan. The standard argument is that we should go vegan to reduce animal suffering.¹ But this argument faces a challenge: thanks to the size and structure of the animal agriculture industry, any individual's consumption decisions are overwhelmingly unlikely to make a difference. *Producing* animal products may be harmful and wrong, but the effects of any individual's *consumption* decisions are insignificant. Going vegan, in other words, is causally inefficacious.² The reduction of animal suffering cannot ground an obligation to go vegan.

The causal inefficacy objection poses a serious challenge to ethical veganism: the view that we have a moral obligation to refrain from purchasing and consuming animal products. Ethical vegans have carefully outlined the conditions of animals on factory farms, the suffering they experience, and the moral significance of reducing this suffering. But the causal inefficacy objection threatens to cut off the rationale for ethical veganism from animal suffering altogether, rendering irrelevant ethical vegans' persuasive arguments on these points. And, in this paper, we argue that common replies to the causal inefficacy objection

1 E.g., Singer, *Animal Liberation* and "Utilitarianism and Vegetarianism."

2 Budolfson, "The Inefficacy Objection to Consequentialism and the Problem with the Expected Consequences Response" and "Is It Wrong to Eat Meat from Factory Farms?"; Fischer, *The Ethics of Eating Animals*, ch. 4; Nefsky, "Consumer Choice and Collective Impact" and "Fairness, Participation, and the Real Problem of Collective Harm"; Shaha, *Why It's OK to Eat Meat*, ch. 4.

are unsatisfactory. Attempts to show that individual vegans are indeed causally efficacious are unsuccessful or at best inconclusive. And arguments that appeal to factors like complicity not only face substantive difficulties but also fail to accommodate the moral significance of *reducing* animal suffering. A stronger argument for ethical veganism would acknowledge that the case for going vegan would be weaker if vegans *as a group* made no difference to reducing animal suffering either. It would tie our obligation to go vegan to the fact that vegans collectively reduce animal suffering, even if no individual vegan makes a difference.

Specifically, we believe that the best response to the causal inefficacy objection relies on the wrongness of free riding. The basic idea is this. As a group, individuals who abstain from animal products create a morally important good: a large reduction in animal suffering. If one recognizes this much yet consumes animal products, one is free riding on vegans. One is recognizing the value of their goal, recognizing that the group makes a significant difference to achieving it, and yet making an exception of oneself by free riding on, rather than participating in, its production. This is wrong because free riding is wrong. And it remains wrong even for those who do not recognize the value of this goal because morality does not let one off the hook so easily: one cannot escape an obligation to go vegan simply by not caring about animal suffering. We are not only obligated to produce morally important goods through our own actions but are also obligated to participate in, rather than free ride on, their collective production. The latter obligation explains why we should go vegan.

We begin by sketching the standard argument for ethical veganism. We then explain the causal inefficacy objection and why we find existing rejoinders inadequate. From here, we develop our anti-free riding argument and consider several objections that lead us to qualify but not abandon our conclusion. The upshot is that even if one settles several controversial issues in ways that make trouble for our argument, there at the very least remain strong reasons for most people to purchase significantly fewer inhumanely raised animal products.

To be clear, our goal is only to examine the connection between animal suffering and our reasons to go vegan or otherwise change our behavior. Parallel considerations apply to other negative consequences of animal agriculture. But nothing we say here bears on the plausibility of grounding reasons to go vegan in something other than the consequences of animal agriculture—for example, in the idea that eating animals is disrespectful.

1. THE STANDARD ARGUMENT

There is overwhelming evidence that factory farming causes animals immense suffering. Animals are kept in horrifying conditions, have their bodies mutilated

without anesthetic, are transported in overcrowded trucks in which they crush one another, and are slaughtered, often painfully and in a state of extreme fear. All this is well documented elsewhere; we spare readers the gruesome details.³ But it is worth noting the scale. In the United States alone, about 9 billion chickens, 120 million pigs, and 30 million cows are slaughtered each year. Of these, 98.2 percent of chickens raised for eggs, 99.9 percent of chickens raised for meat, 98.3 percent of pigs, and 70.4 percent of cows are raised on factory farms.⁴ The severity and scale of suffering on factory farms is staggering.

Pointing to this suffering is the first step in the standard argument for ethical veganism. The second step is to claim that this suffering massively outweighs any compensating benefits of factory farming, such as the pleasure of eating meat or associated cultural experiences.⁵ This step, too, is familiar, but in brief, those who deny it face two challenges. First, it is notoriously difficult to find grounds for believing that animals' interests matter less than humans' interests that do not implausibly commit us to thinking that some humans' interests matter less than others.⁶ Second, even if humans' interests matter more, it is implausible that they matter *so much* more that the value we derive from animal products outweighs the extent of animal suffering. Perhaps if humans suffered as much as animals on factory farms do, this suffering would be even more abhorrent. But that is not the relevant comparison. The extent of animal suffering remains abhorrent, and extremely morally bad, even after factoring in any countervailing benefits.⁷

We believe that the first two steps of the standard argument are successful and assume as much here. Our focus is on the final step, which attempts to derive an obligation to go vegan or otherwise change our dietary behavior from the moral horrors of factory farming. To keep things simple, we focus on veganism for now and take up relevant differences later. According to this final

3 See, e.g., Singer, *Animal Liberation*; Foer, *Eating Animals*.

4 Jacy Reese Anthis, "US Factory Farming Estimates," Sentience Institute, April 11, 2019, <https://www.sentienceinstitute.org/us-factory-farming-estimates>. The statistic for cows refers to those raised in concentrated animal feeding operations as defined by the US Environmental Protection Agency. However, an anonymous referee points out that beef cattle in such operations lead much of their lives in better conditions than our description suggests. For now, we treat all animal products equivalently, but we return to relevant differences later.

5 On such benefits, see Lomasky, "Is It Wrong to Eat Animals?" See also Cohen, "The Case for the Use of Animals in Biomedical Research."

6 Singer, *Animal Liberation* and "Utilitarianism and Vegetarianism"; though see Kagan, *How to Count Animals*; Setiya, "Humanism."

7 DeGrazia, "Moral Vegetarianism from a Very Broad Basis"; Gill, "On Eating Animals"; Kagan, *How to Count Animals*, 5.

step, the connection between veganism and animal suffering is straightforward. Factory farms operate to meet consumer demand for animal products. Going vegan is an effective and relatively low-cost way to reduce demand and so to reduce animal suffering. And since we are obligated to reduce suffering when we can do so at relatively low cost, we are therefore obligated to go vegan.

2. THE CAUSAL INEFFICACY OBJECTION

Unfortunately, the standard argument is too simple. The trouble is that the animal agriculture industry's size and structure appear to render it insensitive to individual consumption decisions. For example, recall that nine billion chickens are slaughtered a year in the United States, working out to over twenty-four million chickens a day. This suggests that if you are at a restaurant deciding between chicken or tofu, your choice will not harm any chickens. The chickens in the restaurant are already dead. And going forward, it is not like ordering tofu sends a signal directly to a factory farmer who says, "Oh! I better produce one less chicken next month!" Instead, the restaurant purchases chickens in bulk, from a distributor who purchases in bulk, from a processor who purchases in bulk, and so on, all the way back to a factory farm. At no point in this supply chain are decisions fine grained enough to reflect individual choices—again, in the United States, *twenty-four million* chickens are slaughtered a day. Rather, reductions in demand only trigger reductions in supply when a series of thresholds is met: enough consumers must refrain from purchasing chickens from enough restaurants and stores that enough restaurants and stores reduce their orders from enough distributors, and so on, that enough processors reduce their orders from factory farms that those farms produce fewer animals. And one choice to order tofu is, unfortunately, not going to trigger all these thresholds. This is the causal inefficacy objection to ethical veganism.

This objection is powerful. The standard argument says that we should go vegan because doing so will reduce animal suffering and the cost is relatively small. But the causal inefficacy objection suggests that going vegan has no such benefit and so is not worth even a small cost. This appears to let omnivores off the hook: they can maintain that factory farming is awful, curse their causal inefficacy, and eat animal products with a clean conscience. Animals should not suffer, but abstaining from animal products does not reduce this suffering. So why should someone have to suffer through tofu when they so much prefer chicken?

Some ethical vegans are unconvinced. They argue that even though no individual decision is likely to make a difference, there must be some number of dietary choices that is large enough to do so—that is, to trigger the aforementioned series of thresholds. For example, even though no choice to abstain

from a whole chicken dinner is likely to save a chicken, perhaps every hundred thousand fewer chickens sold will result in one hundred thousand fewer chickens being produced. In that case, since you have no idea how far you are from triggering the relevant thresholds, you should assign a probability of $1/100,000$ to your abstention from chicken triggering the thresholds and saving one hundred thousand chickens. And, according to expected value theory, saving one hundred thousand chickens with a $1/100,000$ probability is exactly as good as saving one chicken with certainty. Thus, the causal inefficacy objection appears defused. Refraining from chicken is extremely unlikely to save any animals—but when it does, it saves a huge number of them. And this is enough to render one's choice efficacious.⁸

This expected value argument may seem to seal the deal for ethical veganism. But, again, things are not so simple. There are two basic worries—one calling into question the collective impact of large groups of vegans, another the impact of individual vegans.⁹ The first worry is that even if one hundred thousand decisions to refrain from chicken would reduce the number of chickens on factory farms, they are unlikely to reduce the number by one hundred thousand. Instead, they will cause the price of chicken to drop, which will cause some who would not have otherwise bought chicken to do so—at least insofar as farmers are willing to produce chickens at the lower price.

Some proponents of the causal inefficacy objection suggest, on the basis of anecdotal evidence, that such “inelasticities” massively decrease the collective efficacy of vegans.¹⁰ However, the food economists Norwood and Lusk estimate that this effect is slight: in the United States, for a sufficiently large number n , if n choices are made not to buy a chicken, $0.76n$ fewer chickens will be produced.¹¹ If this is roughly correct, then inelasticities cannot plausibly undermine the obligation to go vegan, since the expected benefit of saving roughly 0.76 chickens from suffering is still very significant relative to the associated cost. So the standard argument for ethical veganism withstands the worry that large groups of vegans are causally inefficacious. At least over the long run,

8 Norcross, “Puppies, Pigs, and People,” 232–33; Singer, “Utilitarianism and Vegetarianism,” 335–36; Kagan, “Do I Make a Difference?”

9 Compare Nefsky, “Consumer Choice and Collective Impact,” 273. A third worry is that expected value theory treats small probabilities of large value inappropriately: a $1/100,000$ probability of saving one hundred thousand chickens is not as good as saving one chicken with certainty. We set this aside here.

10 Budolfson, “The Inefficacy Objection to Consequentialism and the Problem with the Expected Consequences Response,” 1718, and “Is It Wrong to Eat Meat from Factory Farms?” 86–89; Nefsky, “Consumer Choice and Collective Impact,” 273.

11 Norwood and Lusk, *Compassion, by the Pound*, 223. Their values for other animal products range from 0.56 to 0.91.

the animal agriculture industry is sensitive to macrolevel market trends—for example, to large numbers of individuals refraining from chicken.¹²

This brings us to the individual impact worry. Even if n decisions to refrain from chicken would trigger a series of thresholds that saves $0.76n$ chickens, it does not follow that a single decision to abstain has a $1/n$ probability of doing so: the *expected impact* of one decision need not equal the *average impact* of n decisions. The animal agriculture industry involves a long supply chain from the farm to the table. At each link in this chain there is considerable *slack*—a margin for error or tolerance for waste. Thanks to slack, one decision to abstain from chicken may have much less than $1/n$ the expected impact of n decisions, and so much less expected value than saving 0.76 chickens.¹³ Consider a stylized case.

Suppose a grocery store decides how many chickens to order each month based on its sales the previous month. It orders chickens in quantities of 500, and typically sells about 9,750 chickens a month. So, for some time, it has been ordering 10,000 chickens a month. The store is willing to tolerate some waste and so will only reduce its order from 10,000 to 9,500 if it sells fewer than 9,500 the previous month. In this case, reducing chickens purchased from the store by 500 is guaranteed to result in the store purchasing 500 fewer chickens the following month. But what is the probability a single decision to refrain from chicken makes this difference? Is it $1/500$, as proponents of the expected value argument assume?

Well, that depends. The probability your decision makes a difference would be $1/500$ if your choice had an equal probability of reducing monthly demand to 9,499 as it did of reducing demand to any other level. But this need not be so. For example, suppose you know trends will hold up: the store will always sell between 9,600 and 9,900 chickens a month. Then, the probability you reduce next month's order by 500 is not $1/500$, but zero. Or suppose you know the store will sell between 9,600 and 9,900 chickens in a month unless a shock occurs, in which case any level is equally probable, and there is a 99 percent probability no shock occurs. Then, you have a $1/500$ probability of causing the store to order 500 fewer chickens *if* a shock occurs, but only a $1/50,000$ probability of doing so overall.

These examples demonstrate that expected impact can come apart from average impact and so defeat any *a priori* argument that the two must converge:

- 12 McMullen and Halteman, "Against Inefficacy Objections"; compare Hedden, "Consequentialism and Collective Action," 536.
- 13 Budolfson, "The Inefficacy Objection to Consequentialism and the Problem with the Expected Consequences Response" and "Is It Wrong to Eat Meat from Factory Farms?" See also Nefsky, "Consumer Choice and Collective Impact"; Fischer, *The Ethics of Eating Animals*, ch. 4.

it is possible that vegans are on average impactful, yet the expected impact of one vegan is very low. The question therefore becomes whether, empirically, we find this divergence. Skeptics argue that we do not because consumers lack the crucial information about thresholds and trends generating the results in the above examples. Absent such information, slack makes no difference to expected impact. Suppose you are in the same store, but you have no idea where thresholds lie or what consumer trends are like. Then, you should estimate that you are just as likely to reduce demand to a threshold value as you are to reduce it to any other level. So, if thresholds occur every n choices, the probability you hit a threshold is $1/n$. And this remains true even if you know there is a lot of slack in the animal agriculture industry. More slack implies that the distance between thresholds is larger, not that your expected impact is lower: it implies that n is larger, not that you have less than a $1/n$ probability of hitting a threshold.¹⁴

But this is too quick. Even absent information about consumer trends or the location of thresholds, consumers might nevertheless know that the probability they reduce demand to precisely some threshold level is lower than $1/n$ because there is some correlation between thresholds and consumer demand.¹⁵ For example, suppose the grocery store decided how many chickens to purchase this month based on its projection of how many it would sell, and it will only change its future orders if this projection proves far off. Then, if the grocery store's projection is fairly reliable—larger errors in its projection are much less probable than smaller ones—the probability this projection is far off will be much lower than the probability it is approximately right. So you are *not* just as likely to reduce demand to the far-off threshold level that makes the store change its future orders as you are to reduce it to any other level. The probability you hit the threshold is much less than $1/n$.

Proponents of the causal inefficacy objection argue that this is relevantly analogous to the position real consumers find themselves in.¹⁶ Grocery stores lose customers who see empty cases, so they have an incentive to overpurchase animal products.¹⁷ This is partly why they produce so much waste—in the United States, for example, a lower-end estimate suggests that grocery stores

14 Hedden, "Consequentialism and Collective Action," 537–39; compare McMullen and Halteman, "Against Inefficacy Objections," 99–100.

15 Hedden notes this possibility but dismisses it as empirically unlikely ("Consequentialism and Collective Action," 539n17).

16 See Fischer, *The Ethics of Eating Animals*, ch. 4; Shahar, *Why It's OK to Eat Meat*, ch. 4. They draw on Budolfson, "The Inefficacy Objection to Consequentialism and the Problem with the Expected Consequences Response" and "Is it Wrong to Eat Meat from Factory Farms?"

17 Fischer, *The Ethics of Eating Animals*, 59; Shahar, *Why It's OK to Eat Meat*, 99.

throw away 4.5 percent of fresh meat.¹⁸ Since demand fluctuates from month to month, grocery stores have a further incentive to tolerate significant variation in sales without adjusting their orders, rather than trying to fine-tune their orders and risk underpurchasing, leaving their customers unhappy.¹⁹ So they have strong incentives to adopt ordering strategies much like in our example. They make projections with some margin of error, such that only large divergences from their projections (signaling a change in market trends), but not minor fluctuations, lead them to change their plans. As our stylized case shows, this can be enough to reduce a consumer's expected impact.

Indeed, the problem may be worse in the real world because similar dynamics arise at each link of the supply chain—the effects of which can compound rapidly to drive down expected impact. Not only do grocery stores and restaurants tolerate waste and variations in demand when deciding how much to purchase from their distributor, so does their distributor when deciding how much to buy from their processor, and so on. Unsurprisingly, then, there is a huge amount of slack in real-world supply chains. For example, in North America and Oceania, a whopping 13 percent of meat initially produced on farms is wasted while working its way down the supply chain to the consumer.²⁰

All this suggests that one's expected impact may be considerably lower than proponents of the expected value argument assume. Of course, just how low is hard to say: the failure of any *a priori* argument that expected impact equals average impact means that one's expected impact depends on controversial claims about messy empirical reality. Here, Norwood and Lusk's figures on elasticity are often cited as showing that individuals have a high expected impact.²¹ But this is a misinterpretation, as these figures only refer to average impact, and Norwood's stated view is that "for all practical purposes, expected effects are impossible to determine."²² Any confident assertion that a typical consumer's expected impact is high enough relative to the cost of going vegan to ground an obligation to do so therefore strikes us as overconfident. This either defeats

18 Buzby et al., "Supermarket Loss Estimates for Fresh Fruit, Vegetables, Meat, Poultry, and Seafood and Their Use in the ERS Loss-Adjusted Food Availability Data." See Fischer, *The Ethics of Eating Animals*, 59, for discussion.

19 Fischer, *The Ethics of Eating Animals*, 60.

20 Food and Agriculture Organization, *Global Food Losses and Food Waste*.

21 E.g., MacAskill, *Doing Good Better*, 88, 228.

22 Fischer, *The Ethics of Eating Animals*, 61 (personal correspondence with Norwood). The misinterpretation is understandable, since Norwood and Lusk do say that buying one unit less of chicken reduces production by 0.76 units (*Compassion, by the Pound*, 223). But (as Norwood confirms in the same personal correspondence with Fischer) their analysis only shows that buying n fewer units reduces production by $0.76n$ units for large n 's.

the standard argument for ethical veganism—which relies on the premise that individuals have a high expected impact—or, depending on where one places the burden of proof, at least leaves the defender and critic of ethical veganism at a stalemate, as far as the standard argument is concerned.²³ But we need not settle for a draw. Controversy about expected impact leaves untouched the (comparatively) uncontroversial claim that vegans have a huge collective impact on reducing animal suffering, both absolutely and relative to the cost of going vegan. Intuitively, this should be enough to ground an obligation to go vegan, regardless of whether *individual* vegans have a significant expected impact. Can some other account explain why?

3. TOWARD A SOLUTION

We began with the idea that the suffering of animals on factory farms somehow grounds an obligation to abstain from animal products. The simplest view is that we should go vegan because doing so reduces suffering, but this runs into the causal inefficacy objection. Many ethical vegans have therefore retreated to the idea that we should go vegan even if doing so fails to reduce animal suffering, because our dietary choices bear some other relation to animal suffering. A typical view appeals to complicity. The animal agriculture industry produces tremendous suffering. Participating in it makes one wrongfully complicit.

Standard objections to complicity views challenge their explanations of either why omnivores are complicit or why complicity is wrong. For example, suppose we understand complicity *causally*, in terms of an individual's expected impact on the maintenance of factory farming, or *expressively*, in terms of what eating animals conveys—say, approval of factory farming or callous disregard for animal suffering.²⁴ The causal inefficacy objection challenges the view that

23 Compare Fischer, *The Ethics of Eating Animals*, 62. For the same reason, attempts to revive the expected value argument by appeal to “indirect effects”—going vegan may cause others to go vegan (Norcross, “Puppies, Pigs, and People,” 233; Almassi, “The Consequences of Individual Consumption,” 404–7)—are unsatisfactory. Such effects are similarly difficult to determine, especially since they are not uniformly positive: vegans may turn others *off* veganism. Although we cannot find a quantitative estimate, our suspicion is that the typical consumer's expected indirect effects are (on balance) positive but small. The most plausible route to high indirect impact is indirect effects compounding, as when each vegan converts two others to veganism, who each converts two others, and so on. But this is not what we find: veganism, unfortunately, is not growing at an exponential rate (compare Nefsky, “Consumer Choice and Collective Impact,” 271).

24 See, respectively, Lepora and Goodin, *On Complicity and Compromise*, ch. 3; and Driver, “Individual Consumption and Moral Complicity.”

omnivores are causally complicit.²⁵ And those convinced by the causal inefficacy objection need not be expressively complicit either: they might abhor factory farming and feel for the suffering of animals, yet consume animal products because they believe doing so makes no difference.²⁶ Or suppose we understand complicity *extractively*, in terms of benefiting from wrongdoing.²⁷ Omnivores are plausibly extractively complicit, but there are many counterexamples to the view that benefiting from wrongdoing is wrong. To take just one: it does not seem wrong to benefit from medical procedures that were developed through wrongful experimentation.²⁸

We are sympathetic to these objections, but rather than fortifying them against potential rejoinders, we focus on a more general problem. The puzzle the causal inefficacy objection raises is that while vegans collectively reduce animal suffering, it is hard to see why individuals are obligated to join in if no individual has a significant impact. An adequate solution to this puzzle should provide some way of connecting individuals' obligations to go vegan with the collective reduction of animal suffering, rather than rendering this reduction irrelevant. Complicity views cannot do this since they sever the connection between the obligation to go vegan and the reduction of animal suffering altogether: we would still be wrongfully complicit, on such views, even if no number of vegans could successfully reduce animal suffering. But vegans do collectively reduce animal suffering, and an argument for ethical veganism can and should take advantage of this fact.

The importance of tying our obligation to go vegan to vegans' collective impact comes out clearly when we consider another view about why eating some animal products is wrong even if individuals are causally inefficacious—the view that it is wrong to consume products whose production *essentially* involves harm, regardless of our causal relation to that production.²⁹ On this view, it is permissible to eat animal products—say, milk—that *can* be produced painlessly, even if their actual production involves suffering. But it is wrong to eat products that are essentially harmful in the sense that they cannot be produced without suffering, even if refraining from them does not reduce this harm.

This account is problematic. Suppose a new technology allows us to produce meat using a device that extracts flesh from animals before immediately

25 Budolfson, "Is It Wrong to Eat Meat from Factory Farms?" 92, and "The Inefficacy Objection to Consequentialism and the Problem with the Expected Consequences Response," 1713.

26 Fischer, *The Ethics of Eating Animals*, 96.

27 McPherson, "Why I Am a Vegan (and You Should Be One Too)."

28 Nefsky, "Consumer Choice and Collective Impact," 277.

29 Budolfson, "Is It Wrong to Eat Meat from Factory Farms?" 94–97.

repairing their bodies without causing any suffering. Since the technology renders harm no longer essential to the production of meat, its development should, on this account, result in the consumption of meat transforming from impermissible to permissible—even if it is never used. This is implausible. Actually using this futuristic technology to produce meat would render eating that meat morally unproblematic (at least on grounds of its connection to animal suffering), but the mere existence of this technology lacks such significance. The lesson is that our moral obligations depend on the actual harm and suffering caused by animal production, rather than this harm's essentiality. More carefully, they depend on the extent to which vegans collectively *reduce* this harm.

To drive this point home, suppose that producing bacon causes the same amount of suffering as producing chicken wings, but (due to differences in their supply chains) no number of individuals refraining from chicken wings would make any difference to this suffering, while relatively small numbers refraining from bacon would make a significant difference. Then, intuitively, even if no individual makes a difference, we have stronger reasons to refrain from bacon than to refrain from wings. And, crucially, this holds regardless of which *essentially* involves more harm or makes us more complicit.

A final view that accommodates this insight holds that individuals have reasons to help bring about valuable outcomes, such as the reduction of animal suffering, even when individual participation makes no difference.³⁰ This view relies on a nonstandard definition of helping, on which an individual helps bring about an outcome when their action plays a *nonsuperfluous* causal role in its production. An action counts as nonsuperfluous when it is possible both for it to be part of the cause of that outcome and for the outcome to fail to come about because not enough people perform actions of that type. So we have reasons to help in such cases, and our reasons to help more impactful collectives are stronger.³¹

Unlike the other views on offer, this view successfully connects reasons to go vegan with the reduction of animal suffering: individuals have reasons to *help* reduce animal suffering. But we worry that it ultimately leaves such reasons unexplained. It is uncontroversial that we have reasons to help bring about outcomes if “helping” is interpreted in its everyday sense, which involves making a causal difference. But the view in question uses “helping” (and associated notions like “nonsuperfluous”) as terms of art. And reasons to help in this special sense do not automatically inherit the credentials of reasons to help in the everyday sense. Absent some further explanation of why we have

30 Nefsky, “How You Can Help, without Making a Difference.”

31 Nefsky, “How You Can Help, without Making a Difference,” 2764.

such reasons, then, this view appears to relocate rather than solve our puzzle: Why do individuals have reasons to “help” if no instance of helping makes a difference? A more satisfying and complete view would provide this further explanation.³² We develop such a view now.

4. THE ANTI-FREE RIDING ARGUMENT

We have seen that an adequate reply to the causal inefficacy objection should hold on to the idea that our obligation to go vegan is somehow related to the reduction of animal suffering. Specifically, it should tie our obligation to go vegan to the collective impact of vegans, even if no individual makes a difference. Thankfully, morality provides us with just the connection we need. In many contexts, it is wrong to free ride on the collective production of important goods. Since the reduction of animal suffering is one such good that vegans collectively produce, those who consume animal products free ride, and this is what makes their consumption decisions wrong. We have moral obligations not only to reduce animal suffering through our own actions but also to participate in, rather than free ride on, collective endeavors that have this impact. The causal inefficacy objection suggests that going vegan may not be an effective way to discharge the former obligation, but it leaves the latter untouched. This explains why we should go vegan.

There is much disagreement about how exactly to formulate the moral requirement not to free ride, but there is also widespread agreement that some such requirement applies in paradigm cases. It is a crucial feature of these cases that free riding is wrong even if individual instances of free riding make no difference. For example, many explain why we should pay to ride public transportation—rather than literally riding for free—by appeal to the wrongness of free riding. Although our own measly fare is unlikely to affect the public transit system, the system would cease to function if nobody paid to ride. Since anybody could equally help themselves to the justification that *their* participation makes no difference, allowing that this provides a genuine exemption from paying would preclude us from the good of public transportation. Morality helps us overcome this problem by denying that this is a genuine exemption and by demanding that we pay our fare even if doing so makes no difference.

32 A related problem of incompleteness arises for McPherson, who suggests that there may be reasons to participate in collectively beneficial social patterns and to abstain from collectively detrimental ones (“The Puzzle of Inefficacy,” 240–42). Such reasons might solve our puzzle, but we need a further explanation of why they exist. (To be fair to McPherson, he does not purport to defend this view, but merely gestures to it as having the right general shape to solve our puzzle.)

This, if nothing else, is what seems right about Kantian approaches to morality: we ought not to make an exception of ourselves, at least in cases where anyone else could make the same exception, and where granting this exception would preclude the production of an important good.³³

It is true that paradigmatic cases of wrongful free riding involve free riding on the collective production of what economists call a *public* good: a (nonexcludable, nonrivalrous) good that benefits the free rider rather than some third party. It is because *I* benefit from riding public transit that I should contribute my fare; those who do not ride public transit are not guilty of free riding. This might seem to render the case of reducing animal suffering disanalogous from standard cases of wrongful free riding: it is animals rather than vegans who benefit from the collective action of vegans. We address this worry in two steps.

The first step is to note that those who acknowledge the great importance of reducing animal suffering do derive value from the actions of vegans. To put things, again, as economists might: those who value the reduction of animal suffering would be willing to pay some monetary cost to reduce it, and so benefit from its reduction, at least in the broad sense of “benefit” relevant here. Indeed, there are many familiar examples of public goods whose production you should not free ride on even though they benefit you, not in the narrow sense that they further your self-interest, but in the broad sense that they further what you care about or prefer to achieve. Common examples include poverty reduction and herd immunity against illnesses that do not threaten you (assuming you care about others’ welfare and health). In fact, certain governments, such as the United Kingdom’s, explicitly identify animal welfare as a public good given widespread preferences for improved animal welfare.³⁴ It is similarly common to treat animal welfare as a public good in economic analyses.³⁵

For those who recognize the great value of reducing animal suffering on factory farms, then, the anti-free riding argument succeeds. Reducing animal suffering is an important public good, and it is a familiar feature of such goods that we should not free ride on their collective production even if our own contribution makes no difference.

This conclusion is already significant. Dialectically, the causal inefficacy objection is mounted by those who acknowledge the great importance of reducing animal suffering and so would acknowledge an obligation to go vegan if individuals were causally efficacious. Our argument implies that anyone in

33 This preliminary gloss on the wrongness of free riding owes much to Cullity, “Moral Free Riding” and “Public Goods and Fairness.”

34 Coe and Finlay, *The Agriculture Act 2020*.

35 E.g., Norwood and Lusk, *Compassion, by the Pound*, ch. 10.

this camp should go vegan, so it responds to the causal inefficacy objection in the dialectical context in which it typically appears.

But this may seem unsatisfying. The standard argument for ethical veganism defends the stronger conclusion that one has an obligation to go vegan regardless of whether one cares about animal suffering. And it seems odd to accept our weaker conclusion that those who care about animal suffering have this obligation, without also accepting the stronger one. After all, to claim otherwise is to allow that individuals can duck their moral obligations—in this case, not to purchase or consume animal products—simply by failing to care (or not caring enough) about what they morally ought to care about—in this case, animal suffering. Morality, however, should not let such callousness get us off the hook. If those who care about animal suffering have an obligation to go vegan, then so too should those who do not.

The second step of our argument bridges the gap between the weaker and the stronger conclusions. The most straightforward route appeals to the idea that callousness cannot extinguish obligations. The argument is simple. The first step of our anti-free riding argument establishes that those who care about animal suffering should go vegan. But, we now add, callous indifference to animal suffering cannot absolve one of such an obligation. So one must have an obligation to go vegan regardless of whether one cares about animal suffering. More generally, if callousness does not exempt one from moral obligations, then the existence of an obligation not to free ride on the collective production of goods one cares about implies an associated obligation not to free ride on the collective production of goods one morally ought to care about, in the sense that not caring about them would exhibit the moral failing of (perhaps among other things) callousness. In other words, it implies a general obligation not to free ride on the collective production of goods that we either care about or morally ought to care about. And this general obligation entails the stronger conclusion that one should go vegan regardless of whether one cares about animal suffering.

This anti-callousness argument is difficult to resist. Given the weaker conclusion that those who care about animal suffering have an obligation not to free ride on collective endeavors that reduce it, one can block it only by insisting that callousness can absolve one of this obligation. But this is implausible. In the individual case, it would be a nonstarter for someone to claim that they are not obligated to produce a morally important good simply because they do not care about it. And it seems equally bizarre to allow that simply not caring about a morally important good can absolve one of an obligation to participate in a collective endeavor to produce it. If callousness does not exempt us of obligations in individual cases, then it should not do so in collective cases either. “I do not care about the suffering of others” is no better a justification for failing

to participate in collective endeavors that reduce suffering than it is for failing to reduce suffering through one's own actions.

Our approach so far has been to argue from the widely acknowledged obligation not to free ride on the collective production of public goods to a similar obligation not to free ride on the collective production of morally important goods. We now strengthen our case for the existence of this latter obligation—first, by drawing out a counterintuitive implication of rejecting it and, second, by noting that most moral theories converge on it.

First, in the absence of this obligation, morality would include a counterintuitive loophole absolving individuals of their obligations to participate in collective action in cases where no individual makes a difference, even though (i) the group is successfully producing a morally important good, (ii) the group is only able to produce this good because its members participate despite their individual inefficacy, and (iii) individuals would be obligated to participate if the group produced the same benefit but through a different causal mechanism that rendered each individual causally efficacious. But we submit that morality does not have loopholes. Much as in the case of riding public transit, morality does not grant us permission to free ride when that permission would be equally available to everyone, and where granting it would preclude the creation of an important good.

To make the counterintuitiveness of this loophole vivid, note that if it existed, morally motivated agents could be manipulated by bad actors in a peculiar way. Suppose that advances in technology reduce slack to the point that every dietary choice makes a significant causal difference. In this world, morally motivated individuals (who otherwise prefer omnivorism) go vegan to reduce animal suffering by their individual actions, driving factory farms to the brink of viability. Realizing their error, factory farmers hatch a devious plot: they will exploit the loophole by reintroducing slack into the system, rendering individual dietary choices again inefficacious. This has no effect on the collective impact agents have on reducing animal suffering, but assuming the loophole in question exists, it removes any obligation to go vegan. The factory farmers' plot succeeds. Morally motivated agents resume consuming animal products and so cease reducing animal suffering.³⁶

We think it clear that something has gone wrong in this story. Morality should not include a loophole allowing bad actors to manipulate morally motivated agents in this way: changes to individual efficacy should not lead morally motivated agents to stop participating in morally important collective action

36 We adapt this thought experiment from Budolfson, "The Inefficacy Objection to Consequentialism and the Problem with the Expected Consequences Response," 1713n1.

that is efficacious at the group level. An obligation not to free ride is exactly the right shape to plug this loophole. So morality must include one.³⁷

Second, although our argument has proceeded at the intuitive level, we may strengthen it by noting that most moral theories converge in endorsing an obligation not to free ride on the collective production of morally important goods. Most obviously, pluralistic deontologists may accept this obligation as a foundational duty, or as one grounded in a requirement of fairness. Indeed, the most influential and well-developed account of free riding holds that free riding on the collective production of both public and morally important goods is unfair, and hence wrong, for the same reason: roughly, because both involve exempting oneself from an obligation in a context where the production of a good requires individuals not to exempt themselves in this way.³⁸

Other moral theories accommodate the same obligation.³⁹ For example, rule-consequentialists will ground it in its good consequences, because individuals adhering to it produces better outcomes. Contractualists will agree, holding that no one can reasonably reject this obligation since its rejection would place larger burdens on those who would otherwise benefit from the production of such goods.⁴⁰ Virtue theorists will ground an obligation not to free ride in the virtue, say, of being *cooperative*, where cooperative people are willing to join in morally important collective endeavors.⁴¹ It is only orthodox act-consequentialists who cannot accommodate this obligation, but that is unsurprising.

An obligation not to free ride on the collective production of morally important goods generally, and so on the collective reduction of animal suffering specifically, therefore stands on firm ground. Shortly, we will consider some challenges to our claim that this obligation implies that we must go vegan,

37 Our claim here is not that bad actors can *never* manipulate morally motivated agents but merely that the possibility of the particular sort of manipulation in the above example—involving a reduction of individual efficacy but no change to collective efficacy or to anything else—is especially implausible. Thanks to an anonymous referee for raising this issue.

38 Cullity, “Public Goods and Fairness” and *The Moral Demands of Affluence*, 62–65. More carefully, Cullity holds that I unfairly free ride when I rely on others to do their parts in a collective imperative to bring about some good without doing my own part. The grounds of the collective imperative differ depending on whether the good is public or morally important, but free riding on each is unfair in the same way. (Cullity reserves the term “free riding” for cases of public goods, but this difference is merely terminological.)

39 Compare Brennan, “Polluting the Polls,” 540.

40 On some versions of contractualism, burdens on animals do not count, rendering such an obligation inapplicable to the case at hand. We take such theories to be implausible, at least insofar as they are meant to provide a complete theory of morality.

41 Cullity, *Concern, Respect, and Cooperation*, ch. 3.

which will require us to qualify our view. But first we clarify our position by addressing three unsuccessful objections.

5. CLARIFYING THE POSITION

The first objection comes from those sympathetic to our conclusion yet skeptical that our argument provides the right explanation of why it is wrong to purchase and consume animal products. The anti-free riding argument suggests that we are obligated to go vegan because it is wrong to free ride on *vegans*. Animals, here, seem oddly missing from the story.

It is true that our explanation of why one should go vegan is that, otherwise, one is free riding on vegans. But the anti-free riding argument still accommodates the intuitive sense in which a concern for animals ultimately grounds this obligation, because there remains the further question of why it is wrong to free ride on vegans. And our answer is that this is wrong because vegans are collectively producing a morally important good: the reduction of animal suffering. If animal suffering were morally unimportant, or if vegans were failing collectively to reduce it, the anti-free riding argument would not succeed. Our explanation of why it is wrong to purchase and consume animal products therefore makes essential reference to the moral importance of reducing animal suffering. Although it would be theoretically neater to claim that individuals should go vegan to reduce animal suffering by their own actions, the causal inefficacy objection calls this into doubt. Our account therefore provides exactly what we set out to provide: a plausible way of grounding an obligation to go vegan in the moral importance of reducing animal suffering. Animals are not missing from the story; they take center stage.

A second objection concerns cost. We only have obligations not to free ride, the thought goes, when we can do so at relatively low cost. We are happy to grant this. But this is unsurprising; proponents of ethical veganism nearly universally acknowledge that the obligation to go vegan does not apply to those for whom veganism would be especially costly. With the exception of those for whom veganism poses a serious economic burden or those who have a relevant medical condition, however, this limitation does little work. Indeed, despite common misconceptions, a vegan diet is cheaper for most people in high- and upper-middle-income countries.⁴² It is healthier than typical omnivorous diets.⁴³ And vegans do not lead worse lives in general, at least as far as

42 E.g., Springmann et al., "The Global and Regional Costs of Healthy and Sustainable Dietary Patterns."

43 See Garrett, "Utilitarianism, Vegetarianism, and Human Health," for an argument that the expected value of going vegetarian is therefore significantly positive (with a survey of the

their (self-reported) subjective well-being is concerned.⁴⁴ Real hardship may justify omnivorism, just as it may justify those in severe poverty sneaking onto public transit and riding for free. But most people reading this paper are lucky enough not to have this justification available.

We return to cost in the next section, but for now consider a third objection: that the strength of our reason not to free ride depends on how many others are participating. The worry is that if almost no one is participating in the production of some good, then one's reason to participate is weaker and may not generate an obligation. And since vegans are a small minority, this might seem to undermine the anti-free riding argument.

There is something to this thought. In many cases, the fact that very few are participating in collective action extinguishes our obligation not to free ride. Consider an analogous case. Early in the COVID-19 pandemic, social distancing was encouraged as a method for slowing the spread of the virus. Certain models, however, suggested that social distancing is only beneficial when nearly universally practiced.⁴⁵ If this is correct, then under conditions where social distancing was nowhere near, and was never likely to reach, this level, individuals did not seem to have an obligation to social distance (grounded in the wrong of free riding). Since individuals engaged in social distancing were failing to reduce virus transmission, others were not wrongfully free riding by failing to do so. The lesson is that one cannot free ride on the production of a good if too few are participating to produce that good, or perhaps if one does not anticipate that enough will eventually participate.

The same example can, however, be repurposed to more constructive ends. Other models of social distancing suggest that even though no individual makes a significant difference, the more who engage in social distancing the better: 50 percent of people social distancing reduces virus transmission more than 40 percent, which reduces it more than 30 percent, and so on. On these models, social distancing produces not a binary (or single-step) good but a multistep good. The good is not all or nothing, only kicking in at, say, 70 percent of people social distancing. Rather, small numbers of people social distancing produce the good of reduced virus transmission to some degree, and increases in the number of people social distancing result in this good being produced to a greater degree if those increases are sufficiently large.

empirical evidence at notes 24 and 25).

44 See Pfeiler and Elgoff, "Do Vegetarians Feel Bad?" for an attempt to measure this while controlling for various factors, and Iguacel et al., "Vegetarianism and Veganism Compared with Mental Health and Cognitive Outcomes," for a meta-analysis.

45 Chang et al., "Modelling Transmission and Control of the COVID-19 Pandemic in Australia."

If these latter models are correct, then, intuitively, we did have an obligation to social distance grounded in the wrong of free riding, even when, say, only 25 percent of people were doing so (at least assuming the collective benefits were worth the cost). What matters is not the total number or proportion of people who are participating in the production of a good, but whether enough are participating to collectively make a significant difference—and perhaps whether further increases in participation will, at some point, pass some further threshold or step that results in even greater production of the good.

As we have seen, the reduction of animal suffering is a multistep good in the same way. Even if no individual vegan makes a difference, large enough increases in decisions to abstain from animal products do. So the fact that there are few vegans does not imply that we lack an obligation not to free ride on them, so long as they are genuinely reducing animal suffering. Our obligation not to free ride on the collective production of multistep goods is not extinguished merely because a small minority are participating. It is only extinguished when too few are participating to make any difference at all.⁴⁶

According to a 2018 Gallup poll, approximately 3 percent of individuals in the United States (where veganism is not especially popular) self-identify as vegans.⁴⁷ This strongly suggests that vegans are indeed making a collective difference: as we have argued, although the supply chain for animal products may be insensitive to individual decisions, it is highly sensitive to macrolevel market trends. As further evidence of this, consider how many more vegan products are available at restaurants and grocery stores than there used to be. It follows by the anti-free riding argument that it is wrong to free ride on vegans' reduction of animal suffering, even if 97 percent of people in the United States are guilty of doing so.

46 Contra Nefsky ("Fairness, Participation, and the Real Problem of Collective Harm," 255), we see it as a benefit, not a cost, that the anti-free riding argument implies no obligation to go vegan when not enough are doing it to make a difference. But we stress that this does not further imply that we are never obligated to initiate novel forms of collective action. We may very well have such an obligation for some other reason—say, due to the expected impact (via the unusually large "indirect effects") of starting a new movement. But, crucially, the actions required to initiate collective action often differ from those required not to free ride on ongoing collective endeavors—for example, going vegan may, but need not, be an effective way to start a movement to reduce animal suffering (compare Cullity, *Concern, Respect, and Cooperation*, ch. 3). So it is an advantage of our account that it does not treat the two cases identically. (And, if one disagrees, one may modify our account to say that we also have an obligation not to free ride on *potential* morally important goods.)

47 R.J. Reinhart, "Snapshot: Few Americans Vegetarian or Vegan," Gallup, August 1, 2018, <https://news.gallup.com/poll/238328/snapshot-few-americans-vegetarian-vegan.aspx>.

6. WIGGLING OUT OF VEGANISM

This brings us to an important issue we have bracketed so far. We have framed our discussion as an argument that we have an obligation to go vegan, rather than to modify our choices in related ways. But have we really earned this conclusion? There are four worries to consider.

The first worry is that our argument relies on the moral significance of animal suffering and therefore cannot explain why purchasing and consuming animal products from humane, suffering-free farms is objectionable. As we have noted, however, the overwhelming majority of animal products in the United States and much of the world comes from factory farms. And there are reasons to doubt whether allegedly humane farms are genuinely humane: a 2012 report from the Animal Welfare Institute, for example, suggested that production methods on US chicken farms certified by the US Department of Agriculture as “superior” for animal welfare “are not materially different from conventional production methods” and that standards are very laxly enforced.⁴⁸ Many ethical vegans therefore argue that, in practice, there is hardly any difference between buying whatever animal products one pleases and only buying animal products that are produced in allegedly humane environments. But there may be exceptions, and some certifications do seem more reliable than others. Indeed, the Animal Welfare Institute has more recently published a useful guide of which animal welfare food labels are trustworthy and which are misleading.⁴⁹

Here, there remain further questions about the morality of raising and slaughtering animals, even painlessly, and so about the morality of buying animal products from genuinely humane farms. On views on which painlessly killing animals is bad (say, because doing so violates rights), we can construct a structurally parallel anti-free riding argument against purchasing and consuming animal products from humane farms. This argument, however, will be weaker in force since those who purchase meat from humane farms are free riding only on the reduction of animal deaths, rather than on the reduction of both deaths and suffering. But whether it is bad for animals to be painlessly slaughtered is much less obvious than whether it is bad for animals to suffer, and we will not investigate the question further here.

A second worry is that our argument runs together purchasing and consuming animal products, but it might seem only to establish an obligation not to purchase them. It might leave dumpster diving or eating dumpster-bound

48 Mathews, “Humanewashed,” 1.

49 Animal Welfare Institute, “A Consumer’s Guide to Food Labels and Animal Welfare.”

leftovers permissible, since even widespread engagement in such activities does not increase demand for animal products.

We are unsure about this. Perhaps individuals who publicly abstain from eating free animal products significantly contribute, as a group, to the reduction of animal suffering through their influence on others' dietary choices or by helping to dismantle a broader ideology of "carnism" that upholds factory farming.⁵⁰ Or perhaps individuals who abstain from consuming animal products in private are less likely to purchase animal products in the future and so collectively have a larger impact.⁵¹ Again, we are unsure. These are the sorts of empirical questions the anti-free riding argument requires us to attend to. Like the standard argument, it makes our reasons to refrain from animal products contingent on facts about the empirical consequences of our dietary choices—only it focuses us on the effects of groups rather than individuals and requires us not to free ride on efficacious groups.

This brings us to the third way our argument may fall short. Vegans are not the only ones who collectively reduce animal suffering. So too, say, do vegetarians. At first glance, our argument suggests no principled case for going vegetarian rather than vegan since dairy cows and chickens raised for eggs suffer on factory farms. Nevertheless, we may consider vegetarianism one strategy of reducing one's use of animal products, akin to other strategies like adhering to Meatless Mondays, Veganuary, or just trying to eat less meat. And it is not obvious that our argument establishes an obligation to go vegan rather than merely to reduce.

Now, the anti-free riding argument does suggest that we have stronger reasons to go vegan than merely to reduce: reductarians, after all, free ride on vegans, whereas the reverse is not true.⁵² So in the absence of weighty counter-

50 See Joy, *Why We Love Dogs, Eat Pigs, and Wear Cows*, ch. 2. Compare John and Sebo ("Consequentialism and Nonhuman Animals," 575–76), who argue against eating humanely raised meat on similar grounds.

51 Compare Almassi, "The Consequences of Individual Consumption," 407–8; John and Sebo, "Consequentialism and Nonhuman Animals," 572–74.

52 We here assume that reductarians form a larger collective of which vegans are a subset. This raises thorny questions about collective membership conditions, since, thanks to the causal inefficacy objection, we cannot simply count someone as a member when they make a difference to the outcome a collective produces (Nefsky, "Fairness, Participation, and the Real Problem of Collective Harm"). In general, the key to solving this issue is to identify members of a collective by their contribution to the same "underlying dimension" in which each *does* make a difference (say, the number of animal products purchased), where large enough changes in this underlying dimension cause changes to the morally important dimension (say, the amount of animal suffering) (Wieland and Oeveren, "Participation and Superfluity"). However, we do not defend any particular account of the relevant dimension here and instead rely on intuitive judgments about collective membership.

vailing reasons to reduce rather than to go vegan (deriving from the greater cost of veganism), the anti-free riding argument suggests that we must go vegan. And it might seem like such reasons are unavailable, since, as we have argued, veganism is not excessively costly, at least for most healthy and affluent people.

Yet this is too quick. The claim we defended above is that the difference in cost between going vegan and failing to participate in the collective reduction of animal suffering is insufficient to outweigh our reasons to participate, which derive their strength from the large collective impact of vegans on reducing animal suffering. But this leaves open the possibility that veganism is excessively costly when compared to some forms of reducing, even for healthy and affluent people, since veganism might be significantly more costly than reducing yet have little additional collective impact.

For example, suppose that someone finds it somewhat costly to go vegan but almost costless to eat largely vegan while allowing themselves some wiggle room, say, in certain social contexts or when traveling. If the difference in the collective impact of strict veganism and wiggly veganism is low but the difference in cost between the two is relatively significant, then perhaps the person is justified in being wiggly vegan. Similarly, if the difference in collective impact between veganism and vegetarianism is low and one finds it much less costly to eliminate meat but not eggs and cheese from one's diet, then perhaps merely going vegetarian is justified.

We find it plausible that some are justified in going less than strictly vegan for such reasons. But this concession is not as significant as it might seem. If one can eliminate the bulk of the cost of going vegan by building in some wiggle room, then doing so may be justified, but this provides no license for failing to go vegan outside narrowly tailored cases. For example, suppose one finds veganism costly because one's social life is organized around meals with nonvegan family and friends. Then, even granting that the significance of such social costs permits one wiggle room in social situations, this provides no justification for eating animal products when alone or in social contexts where vegan options are available.

So while it is hard to say precisely how wiggly any individual's veganism may be, we have not yet found a challenge to our claim that most of us are obligated to significantly reduce our consumption of animal products. Our discussion has assumed, however, that strict vegans are collectively more impactful than reducetarians, such that we have stronger reasons to be strict. But the opposite may hold if reducetarians more often inspire others to join the cause, say, because strict veganism tends to turn others off.⁵³ If so, reducetarianism may be morally preferable.

53 Thanks to an anonymous referee for this point.

Fourth and finally, consider the possibility that one might discharge one's obligation not to free ride without altering one's dietary choices at all. Suppose someone generously donates their money or time to animal welfare charities or activism. Does the anti-free riding argument imply that they must *also* go vegan (or reductarian)? Or have they already done everything required of them to discharge their obligation not to free ride?

In fact, the worry is more general. We have defended an obligation to alter one's dietary choices as an implication of a general obligation not to free ride on the collective production of morally important goods. But an obligation not to free ride on *all* relevant collective endeavors might seem overly demanding, even if participating in any particular endeavor is not. This suggests that we might have latitude about not only how to participate in the collective reduction of animal suffering but also whether to participate in this or some other collective endeavor when discharging our general obligation not to free ride.⁵⁴ Put another way, it might seem implausible that each group can obligate *everyone* to devote themselves to its cause, simply by initiating effective collective action in its pursuit.⁵⁵

This raises deep questions about morality's demandingness that we cannot fully address here. But there are, broadly speaking, three relevant sorts of views. The first is an extremely demanding view on which moral demands *iterate*. If particular demands are not too costly, then we must meet each of them, even if the demands add up to something highly costly in aggregate. On this view, we are obligated to go vegan (or reductarian) since doing so, considered on its own, is not too costly. So too are we obligated to participate in every other way and in every other cause, so long as each additional form of participation, considered on its own, is not too costly.

But less demanding conceptions of morality can also ground an obligation to go vegan or reductarian. Consider a moderately demanding conception that factors in the *aggregate* cost of various demands, but with further provisos. For example, perhaps morality cannot demand so much of us, in aggregate, that we are left unable to lead a flourishing and autonomous life, with various personal projects and commitments. But it demands that when choosing between projects, we opt for morally preferable ones when the resulting package of projects does not leave us significantly worse off or conflict with core commitments (including to our loved ones). This is rough, and the devil is in the details.⁵⁶ But

54 Shahaar, *Why It's OK to Eat Meat*, ch. 5. Compare Fischer, *The Ethics of Eating Animals*, ch. 9.

55 Thanks to an anonymous referee for articulating the point this way.

56 For one attempt to spell out such a moderately demanding view, see Cullity (*The Moral Demands of Affluence*, ch. 9), from whom we have also adapted the *iterative* versus *aggregate* distinction.

views of this type will generally imply that we must participate in the collective reduction of animal suffering. Since one must eat regardless of how else one spends one's time, and since we have seen that veganism is cheaper and so leaves one with more resources than omnivorism without undermining one's (subjective) well-being, adding veganism (or reductarianism) to one's portfolio of projects is clearly compatible with leading a good life full of other projects and commitments.⁵⁷ This represents a disanalogy with many other forms of activism, the addition of which would threaten to swallow up one's time and resources and so one's other projects and commitments.

Finally, consider mildly demanding conceptions of morality on which one must pay some aggregate cost, and that is all. On such views, we are not obligated to go vegan or reductarian so long as we meet the relevant cost threshold in other domains of life. However, even here it is implausible that we have *absolute* latitude or equally strong reasons to participate in any way and in any cause. For example, we intuitively have weaker reasons to donate to less important causes and in less (collectively) effective ways. And our reasons to participate in the collective reduction of animal suffering are plausibly very strong, even if not uniquely so. In the first place, veganism—or at least reductarianism—is unusually collectively cost effective, in the sense that it is low cost at the individual level but produces a massive benefit at the collective level. There are few other cases, if any, where individuals can sacrifice so little to collectively do so much.⁵⁸ Second, when it comes to reducing animal suffering, there is, so to speak, no neutral option. It is not as if the only alternative to joining in a collective endeavor, say, to reduce domestic violence is to (individually or collectively) produce domestic violence. Most of us do neither. But in the case of reducing animal suffering, neutrality is not an option. Some collectively produce animal suffering; others collectively reduce it. One must pick a side.

57 Shahar raises similar considerations but dismisses them on the grounds that they are only relevant on an extremely demanding conception of morality (*Why It's OK to Eat Meat*, 109–13). In so doing, he appears to overlook the possibility of a moderately demanding view.

58 Shahar is skeptical that veganism produces large collective benefits and denies that it is low cost because it causes social friction, involves “treating every meal as a weighty ethical decision,” deprives one of gustatory pleasure, and involves transition costs, like relearning how to cook (*Why It's OK to Eat Meat*, 116–19). However, we take our earlier discussion of elasticity and average impact to support high collective impact: the average decision to abstain from chicken, recall, saves 0.76 chickens, at least according to Norwood and Lusk (*Compassion, by the Pound*, 223). (Shahar approvingly cites this point [*Why It's OK to Eat Meat*, 107n14].) Further, as we have noted, one can, if necessary, reduce social friction by allowing certain exceptions, and as we will discuss shortly, one can reduce the cost of moral deliberation by adopting simple rules. This leaves transition costs and pleasure. But such costs, though they may loom large in the minds of omnivores, are small in comparison to the long-term (collective) benefits of maintaining vegan or reductarian diets.

Both considerations suggest that our reasons to participate in the reduction of animal suffering are uncommonly strong.

So, on extremely and moderately demanding conceptions of morality, we are obligated to go vegan or reducetarian; on mildly demanding conceptions, we have strong reasons to do so. Or, at the very least, we have strong reasons to avoid purchasing factory-farmed products. Whether we should also avoid *consuming* animal products or purchasing *suffering-free* animal products—and how strict about this we should be—depends on how we answer the difficult empirical and normative questions we have raised in this section but cannot hope to resolve here. Instead, we end with three pragmatic considerations.

The first is that we generally have stronger reasons to avoid animal products for which similar reductions of demand would trigger greater reductions in suffering, and this can yield unintuitive results. For example, avoiding eggs plausibly has a larger collective impact than avoiding beef or pork since chickens raised for eggs are more abused and less efficient than beef cattle or pigs (as measured in animals per calorie produced), and since the supply chain for chicken and eggs is more elastic.⁵⁹ The second is that it is typically easier to consistently follow simple, rigid rules than vague or highly complex ones.⁶⁰ This does not bear on how strictly one should abstain from animal products, but it does suggest that a policy of, say, never purchasing animal products at grocery stores, going vegetarian, or eliminating chicken and eggs may be better than a policy of “eating less meat.”⁶¹ Adopting rigid rules that do not require deliberation on a case-by-case basis may also eliminate certain costs of reducing consumption, since many find it costly to treat what they previously saw as the “morally free” zone of dietary choices as a domain that now requires moral deliberation.⁶² The third is that given our tendency to manifest a self-serving bias when engaging in moral reasoning, it may be that we should err on the side of adopting stricter policies. Those who enjoy eating animal products can easily convince themselves that there is nothing wrong with doing so, and

59 See MacAskill (“Effective Reductarianism,” 70), who argues that the three “most effective way[s] to reduce animal suffering [are] to stop eating chicken, then eggs, then pork.” The issue is complicated by other negative consequences of animal agriculture, such as greenhouse gas emissions, which cattle disproportionately produce. But for an argument that animal agriculture has much larger welfare effects on animals than its emissions have on humans (drawing on climate economic models of the social cost of carbon), see Kuruc and McFadden, “Monetizing the Externalities of Animal Agriculture.”

60 E.g., John and Sebo, “Consequentialism and Nonhuman Animals,” 574–75.

61 Rothgerber finds that compared to vegetarians, “conscientious omnivores” report both greater difficulty following their diet and more frequent violations of it (“Can You Have Your Meat and Eat It Too?”).

62 On this cost, see again Shahar, *Why It’s OK to Eat Meat*, 116.

those of us trained in analytic philosophy are especially good at coming up with rationalizations of this choice.⁶³

University of Oxford
 jacob.barrett@philosophy.ox.ac.uk
 sarah.raskoff@philosophy.ox.ac.uk

REFERENCES

- Almassi, Ben. "The Consequences of Individual Consumption: A Defence of Threshold Arguments for Vegetarianism and Consumer Ethics." *Journal of Applied Philosophy* 28, no. 4 (November 2011): 396–411.
- Animal Welfare Institute. "A Consumer's Guide to Food Labels and Animal Welfare." Washington, DC: Animal Welfare Institute, August 2021. <https://awionline.org/sites/default/files/uploads/documents/21FoodLabel-GuideFull.pdf>.
- Brennan, Jason. "Polluting the Polls: When Citizens Should Not Vote." *Australian Journal of Philosophy* 87, no. 4 (December 2009): 535–49.
- Budolfson, Mark Bryant. "The Inefficacy Objection to Consequentialism and the Problem with the Expected Consequences Response." *Philosophical Studies* 176, no. 7 (July 2019): 1711–24.
- . "Is It Wrong to Eat Meat from Factory Farms? If So, Why?" In *The Moral Complexities of Eating Meat*, edited by Ben Bramble and Bob Fischer, 80–98. Oxford: Oxford University Press, 2015.
- Buzby, Jean C., Hodan Farah Wells, Bruce Axtman, and Jana Mickey. "Supermarket Loss Estimates for Fresh Fruit, Vegetables, Meat, Poultry, and Seafood and Their Use in the ERS Loss-Adjusted Food Availability Data." *EIB-44*. US Department of Agriculture, Economic Research Service. March 2009.

63 The authors would like to thank Rhys Borchert, Mark Budolfson, David DeGrazia, Jamie Draper, Christopher Howard, Mario Juarez-Garcia, Joseph Millum, Andreas T. Schmidt, Danny Shahar, Hannah Tierney, Alice van't Hoff, David Wendler, and two anonymous referees for helpful comments and discussion. Thanks also to members of the Philosophy Work in Progress Group at the University of Arizona, the Davis Area Group in Ethics and Related Subjects at the University of California, Davis, and the Centre for the Study of Social Justice Seminar Series at the University of Oxford. This research was supported in part by the National Institutes of Health Clinical Center and the Intramural Research Program. The views expressed in this article are the authors' own. They do not represent the position or policy of the National Institutes of Health, the Department of Health and Human Services, or the United States government.

- Chang, Sheryl L., Nathan Harding, Cameron Zacreson, Oliver M. Cliff, and Mikhail Prokopenko. "Modelling Transmission and Control of the COVID-19 Pandemic in Australia." *Nature Communications* 11 (November 2020): 19393–96.
- Coe, Sarah, and Jonathan Finlay. *The Agriculture Act 2020*. Briefing Paper no. CBP 8702. London: House of Commons Library, 2020. <http://researchbriefings.files.parliament.uk/documents/CBP-8702/CBP-8702.pdf>.
- Cohen, Carl. "The Case for the Use of Animals in Biomedical Research." *New England Journal of Medicine* 315, no. 14 (October 1986): 865–69.
- Cullity, Garrett. *Concern, Respect, and Cooperation*. Oxford: Oxford University Press, 2018.
- . *The Moral Demands of Affluence*. Oxford: Oxford University Press, 2004.
- . "Moral Free Riding." *Philosophy and Public Affairs* 24, no. 1 (January 1995): 3–34.
- . "Public Goods and Fairness." *Australasian Journal of Philosophy* 86, no. 1 (March 2008): 1–21.
- DeGrazia, David. "Moral Vegetarianism from a Very Broad Basis." *Journal of Moral Philosophy* 6, no. 2 (March 2009): 143–65.
- Driver, Julia. "Individual Consumption and Moral Complicity." In *The Moral Complexities of Eating Meat*, edited by Ben Bramble and Bob Fischer, 67–79. Oxford: Oxford University Press, 2015.
- Fischer, Bob. *The Ethics of Eating Animals: Usually Bad, Sometimes Wrong, Often Permissible*. New York: Routledge, 2020.
- Foer, Jonathan Safran. *Eating Animals*. New York: Little, Brown and Company, 2009.
- Food and Agriculture Organization. *Global Food Losses and Food Waste: Extent, Causes and Prevention*. Rome: Food and Agriculture Organization of the United Nations, 2011. <http://www.fao.org/3/I2697E/i2697e.pdf>.
- Garrett, Jeremy R. "Utilitarianism, Vegetarianism, and Human Health: A Response to the Causal Impotence Objection." *Journal of Applied Philosophy* 24, no. 3 (August 2007): 223–37.
- Gill, Michael B. "On Eating Animals." *Social Philosophy and Policy* 30, no. 1–2 (January 2013): 201–7.
- Hedden, Brian. "Consequentialism and Collective Action." *Ethics* 130, no. 4 (July 2020): 530–54.
- Iguacel, Isabel, Inge Huybrechts, Luis A. Moreno, and Nathalie Michels. "Vegetarianism and Veganism Compared with Mental Health and Cognitive Outcomes: A Systematic Review and Meta-analysis." *Nutrition Reviews* 79, no. 4 (April 2021): 361–81.

- John, Tyler M., and Jeff Sebo. "Consequentialism and Nonhuman Animals." In *The Oxford Handbook of Consequentialism*, edited by Douglas W. Portmore, 564–91. Oxford: Oxford University Press, 2020.
- Joy, Melanie. *Why We Love Dogs, Eat Pigs, and Wear Cows*. San Francisco: Conari Press, 2009.
- Kagan, Shelly. "Do I Make a Difference?" *Philosophy and Public Affairs* 39, no. 2 (Spring 2011): 105–41.
- . *How to Count Animals, More or Less*. Oxford: Oxford University Press, 2019.
- Kuruc, Kevin, and Jonathan McFadden. "Monetizing the Externalities of Animal Agriculture: Insights from an Inclusive Welfare Function." *Social Choice and Welfare* (forthcoming).
- Lepora, Chiara, and Robert Goodin. *On Complicity and Compromise*. Oxford: Oxford University Press, 2013.
- Lomasky, Loren. "Is It Wrong to Eat Animals?" *Social Philosophy and Policy* 30, no. 1–2 (January 2013): 177–200.
- MacAskill, William. *Doing Good Better: How Effective Altruism Can Help You Help Others, Do Work That Matters, and Make Smarter Choices about Giving Back*. New York: Avery, 2015.
- . "Effective Reductarianism." In *The Reductarian Solution: How the Surprisingly Simple Act of Reducing the Amount of Meat in Your Diet Can Transform Your Health and the Planet*, edited by Brian Kateman, 69–71. New York: TarcherPerigee, 2017.
- Mathews, Rachel. "Humanewashed: USDA Process Verified Program Misleads Consumers about Animal Welfare Marketing Claims." Washington, DC: Animal Welfare Institute, March 2012. https://awionline.org/sites/default/files/publication/digital_download/fa-humanewashedreportonusdapvp_o.pdf.
- McMullen, Steven, and Matthew C. Halteman. "Against Inefficacy Objections: The Real Economic Impact of Individual Consumer Choices on Animal Agriculture." *Food Ethics* 2, no. 2–3 (May 2019): 93–110.
- McPherson, Tristram. "The Puzzle of Inefficacy." In *Philosophy and Climate Change*, edited by Mark Budolfson, Tristram McPherson, and David Plunkett, 222–51. Oxford: Oxford University Press, 2021.
- . "Why I Am a Vegan (and You Should Be One Too)." In *Philosophy Comes to Dinner: Arguments about the Ethics of Eating*, edited by Andrew Chignell, Terence Cuneo, and Matthew C. Halteman, 73–91. New York: Routledge, 2016.
- Nefsky, Julia. "Consumer Choice and Collective Impact." In *The Oxford Handbook of Food Ethics*, edited by Anne Barnhill, Mark Budolfson, and Tyler

- Doggett, 267–86. Oxford: Oxford University Press, 2018.
- . “Fairness, Participation, and the Real Problem of Collective Harm.” In *Oxford Studies in Normative Ethics*, vol. 5, edited by Mark Timmons, 245–71. Oxford: Oxford University Press, 2015.
- . “How You Can Help, without Making a Difference.” *Philosophical Studies* 174, no. 11 (November 2017): 2743–67.
- Norcross, Alastair. “Puppies, Pigs, and People: Eating Meat and Marginal Cases.” *Philosophical Perspectives* 18, no. 1 (December 2004): 229–45.
- Norwood, F. Bailey, and Jayson L. Lusk. *Compassion, by the Pound: The Economics of Farm Animal Welfare*. Oxford: Oxford University Press, 2011.
- Pfeiler, Tamara M., and Boris Egloff. “Do Vegetarians Feel Bad? Examining the Association between Eating Vegetarian and Subjective Well-Being in Two Representative Samples.” *Food Quality and Preference* 86 (December 2020): 104018.
- Rothgerber, Hank. “Can You Have Your Meat and Eat It Too? Conscientious Omnivores, Vegetarians, and Adherence to Diet.” *Appetite* 84 (January 2015): 196–203.
- Setiya, Kieran. “Humanism.” *Journal of the American Philosophical Association* 4, no. 4 (Winter 2018): 452–70.
- Shahar, Dan C. *Why It’s OK to Eat Meat*. New York: Routledge, 2022.
- Singer, Peter. *Animal Liberation: A New Ethics for Our Treatment of Animals*. New York: Harper Collins, 1975.
- . “Utilitarianism and Vegetarianism.” *Philosophy and Public Affairs* 9, no. 4 (Summer 2020): 325–37.
- Springmann, Marco, Michael A. Clark, Mike Rayner, Peter Scarborough, and Patrick Webb. “The Global and Regional Costs of Healthy and Sustainable Dietary Patterns: A Modelling Study.” *Lancet Planet Health* 5, no. 11 (November 2021): e797–e807.
- Wieland, Jan Willem, and Rutger van Oeveren. “Participation and Superfluity.” *Journal of Moral Philosophy* 17, no. 2 (April 2020): 163–87.

FAIRNESS, COSTS, AND PROCREATIVE JUSTICE

Gideon Elford

IT IS A FAMILIAR AND INTUITIVE view across a wide range of different positions in moral and political thought that it can be right to hold persons responsible for the foreseeable consequences of their choices. What, more precisely, this general thought entails varies considerably depending on the broader normative perspective of which it is a part. A commitment to holding persons responsible has in recent years come to find a central expression in certain popular liberal egalitarian views. The views in question maintain that the value of equality is quite consistent with, indeed may perhaps require, holding persons responsible for the consequences of their choices in the form of liabilities or claims that in turn result in inequalities between persons. Otherwise unfair inequality is thereby rendered fair or just in virtue of a sensitivity to responsibility.

Despite the influence and profile of this responsibility-sensitive egalitarian approach, I contend that the basis for the commitment to responsibility at its heart remains imprecisely understood. More specifically, there is a failure to distinguish between two distinct grounds for holding persons responsible for the consequences of their choices. On the one hand, there is a *fairness*-type basis for responsibility—that an agent’s being better or worse off than others in virtue of their own choice renders that inequality just.¹ On the other hand, there is an importantly distinct *cost-internalization*-type basis for responsibility—that agents ought to bear the consequences of their choices because others have claims against certain costs being imposed upon them.

These respectively different bases for the commitment to responsibility in responsibility-sensitive egalitarian views are importantly related to one another, as I will go on to explain. However, they are also crucially distinct and, as I will show, failing to distinguish between them embodies the danger that we erroneously suppose that egalitarian fairness itself demands a more extensive commitment to responsibility than it necessarily does. I uncover the distinctness of the respective fairness and cost-internalization bases for responsibility

1 Or, similarly, that the choice deprives an agent or others of grounds for complaint with respect to that inequality; see Cohen, “Fairness and Legitimacy in Justice.”

in responsibility-sensitive egalitarian views by way of the examination of an area of debate in which issues of egalitarian responsibility figure prominently—namely the debate concerning procreative distributive justice. In the context of that debate, attending to the distinction in question shows that a commitment to a responsibility-sensitive understanding of equality does not necessarily commend holding parents liable for the costs of children.

1. RESPONSIBILITY-SENSITIVE EGALITARIANISM AND PROCREATIVE JUSTICE

The egalitarian move to responsibility sensitivity involves distinguishing between advantages that are appropriately related to an individual's own responsible choices and those that are not. These responsibility-sensitive egalitarian views—of which the famous so-called luck egalitarian view is arguably the flagship—therefore situate the idea of responsibility centrally when it comes to thinking about the value of equality.² Such views maintain that it is only with respect to responsible choices that inequalities are fair. As Cohen puts it, “luck egalitarianism accounts it an unfairness when some are better off than others through no fault or choice of their own.”³ In shorthand, *unchosen inequality is unfair*. Moreover, this responsibility-sensitive egalitarian approach typically maintains that (relevantly) *chosen inequality is not unfair*.

These responsibility-sensitive egalitarian commitments have played a prominent role in emergent literature concerning procreative justice.⁴ Until recently, comparatively little work among political philosophers has been devoted to addressing the question of who should pay for the costs involved in having and raising children.⁵ Quite clearly, caring for and raising children is costly, and in that way burdensome, not only in the sense that time and resources are

2 See Arneson, “Equality and Equal Opportunity for Welfare”; Cohen, “On the Currency of Egalitarian Justice”; Dworkin, “What Is Equality? Part 2”; Temkin, *Inequality*.

3 Cohen, “Luck and Equality,” 442. See also Temkin, who puts this thought in terms of the worse off when he says “it is bad (unfair and unjust) for some to be worse off than others through no fault (or choice) of their own” (*Inequality*, 13).

4 Important contributions include: Bou-Habib and Olsaretti, “Equality, Autonomy, and the Price of Parenting”; Casal, “Environmentalism, Procreation, and the Principle of Fairness”; Casal and Williams, “Rights, Equality, and Procreation” and “Equality of Resources and Procreative Justice”; Clayton, *Justice and Legitimacy in Upbringing*; Olsaretti, “Children as Public Goods” and “Children as Negative Externalities?”; Rakowski, *Equal Justice*; Tomlin, “Should Kids Pay Their Own Way?”; Steiner and Vallentyne, “Libertarian Theories of Intergenerational Justice”; Vallentyne, “Equality and the Duties of Procreators.”

5 Casal and Williams (“Equality of Resources and Procreative Justice,” 150, 158) mention of some notable exceptions to this relative neglect in the form of Malthus, *An Essay on the Principle of Population*; and George, “Who Should Bear the Cost of Children?”

needed for their development into capable and independent adult members of the community, but also in the sense that as emergent adult members they will have justice claims on others, the satisfaction of which will in turn incur costs.⁶ So children not only need to be cared for, they also, as adults, will have justice entitlements that others may be required to satisfy, the content of which will depend on one's particular conception of justice.⁷

Existing debates surrounding the issue of procreative justice and how the costs of children are covered have tended to fall into two broad camps: the "pro-sharing" camp that favors sharing the costs of children across society, and the "anti-sharing" camp that favors holding particular persons or groups, typically parents, liable for the costs of children.⁸ For reasons that will become clear in what follows, thinking about the question in such dichotomous terms is misleading. It is more appropriate to ask *how much* of the costs of children it is reasonable to expect parents to internalize. As is often emphasized in the discussion of this issue, it is essential to be clear that the question of who should pay is being posed against the background of an otherwise just distribution. Under nonideal, less than fully just conditions, there might be any number of grounds for society at large to bear some of the costs involved in raising children, such as some parents, perhaps especially single parents and women, already being unjustly disadvantaged and being less able (even unable) to provide for their children and/or being more vulnerable in virtue of doing so. The relevant question in the discussion of procreative distributive justice at hand is rather, who should ideally be held liable for these costs, abstracting away from of other such complicating justice considerations.

A cursory consideration of the issue might suggest that responsibility-sensitive egalitarians should favor an anti-sharing stance. When it comes to unchosen advantages or disadvantages, responsibility-sensitive egalitarianism favors pro-sharing—indeed, equal (overall) sharing. But when it comes to the consequences of responsible choices, responsibility-sensitive egalitarianism seems to be against sharing costs and benefits. After all, a child's being born is characteristically the procreative choice of someone—usually those who

6 Such costs may include environmental ones; see Casal, "Environmentalism, Procreation, and the Principle of Fairness."

7 Olsaretti, "Choice, Circumstances, and the Costs of Children."

8 I use the term "parents" to refer to both those who procreate children and those who parent children, in the sense of offering care and nurturing. Though the differences between those choices are important, they do not affect the substance of the argument here. For the view that it should be children themselves who pay such costs (in later life) see Tomlin, "Should Kids Pay Their Own Way?" It is also worth noting that those who affirm an anti-sharing position in several cases take their view to be a *pro tanto* rather than all-things-considered position.

also serve as their parents. Indeed, it is this view, with some caveats, that has proved attractive to some responsibility-sensitive egalitarians. Eric Rakowski elaborates in a passage that has become something of a touchstone for the anti-sharing view in this context.

But babies are not brought by storks whose whims are beyond our control. Specific individuals are responsible for their existing. It is therefore unjust to declare . . . that because two people decide to have a child, or through carelessness find themselves with one, *everyone* is required to share their resources with the new arrival, and to the same extent as its parents. With what right can two people force all the rest, through deliberate behavior rather than bad brute luck, to settle for less than their fair shares after resources have been divided justly?⁹

The anti-sharing stance is also defended in various guises by Casal and Williams, Clayton, and Steiner and Vallentyne.¹⁰ The position those authors affirm is that those who have made responsible choices to procreate and parent should, under relevantly ideal conditions, be held liable for the costs associated with the creation of and care for those children. For instance, Casal and Williams state their position as follows:

Though potential parents should be allowed to decide whether or not to increase family size, some injustice exists if resources are redistributed from others to their offspring as a result of their reproductive decisions. Transfers to the latter . . . should ideally take place at the expense of only their parents' share of resources and should not impinge upon others.¹¹

In a similar vein, Vallentyne says: "The question is whether [the costs of satisfying a child's equality rights] must be borne by the procreators. A plausible conception of the rights and duties of equality will, I claim, answer affirmatively."¹²

Now, while it is true that children are costly, they also offer benefits. Of specific relevance in this context are the benefits that children offer as economic

9 Rakowski, *Equal Justice*, 153.

10 Casal and Williams, "Rights, Equality, and Procreation" and "Equality of Resources and Procreative Justice"; Clayton, *Justice and Legitimacy in Upbringing*; Steiner and Vallentyne, "Libertarian Theories of Intergenerational Justice."

11 Casal and Williams, "Equality of Resources and Procreative Justice," 161–62. It is worth also noting that Casal and Williams allow that where there are positive externalities resulting from the reproductive decisions then those decisions do not necessarily reflect the impingement of costs on others; see "Rights, Equality, and Procreation," 101–3.

12 Vallentyne, "Equality and the Duties of Procreators," 206–7.

assets in the form of future labor and general social contributors. A good portion of the existing debate engages the issue of whether those positive externalities associated with bearing children can ground an enforceable duty on nonparents to contribute toward the costs of raising and providing for them.¹³ I am not going to address that question directly here, though I will return to it in section 5. Instead, I want to cast doubt on the idea that a commitment to a responsibility-sensitive understanding of equality necessarily commends holding parents liable for the costs of children in the way that these authors argue. In particular, I am going to suggest that the commitment to parental liability is apt to run together two quite distinct grounds for holding persons responsible for the costs of their choices. On the one hand, a fairness-based ground for responsibility—that an agent's being better or worse off as a consequence of their own choices renders the inequality a just one. On the other hand, a cost-internalization basis for responsibility—that others have claims against certain costs being imposed upon them.

I illuminate the tendency to conflate these different grounds for responsibility by examining a prominent objection to a pro-sharing stance from certain responsibility-sensitive egalitarians—namely, that if the costs of caring for children are shared, then nonparents have costs unfairly imposed upon them. That is to say that, independent of a choice they have made themselves, and as a result of someone else's choice, nonparents are, *ceteris paribus*, made worse off. If, as responsibility-sensitive egalitarianism suggests, it is unfair for persons to be worse off than others through no fault or choice of their own, then it seems that there is an egalitarian fairness-based objection to sharing the costs of children with nonparents.

2. OLSARETTI AGAINST ANTI-SHARING

As a backdrop to explaining the anti-sharing argument, it is worth briefly exploring an argument from Serena Olsaretti in response to certain responsibility-sensitive egalitarians who defend the anti-sharing stance. Olsaretti notes that responsibility-sensitive egalitarian accounts differ with respect to how far parents ought to be held liable for the costs of children. She rightly points out that neither a belief in the unfairness of unchosen inequality nor a commitment to the claim that some choice-derived inequalities are fair entail any particular account of *which* inequalities are fair ones.¹⁴ In this way, one can

13 Casal and Williams, "Equality of Resources and Procreative Justice"; Olsaretti, "Children as Public Goods?"; Tomlin, "Should Kids Pay Their Own Way?"

14 Olsaretti, "Responsibility and the Consequences of Choice" and "Children as Public Goods?"

affirm a responsibility-sensitive egalitarian view without being committed to the fairness of any particular inequalities. Therefore, nothing about a commitment to a responsibility-sensitive understanding of the value of equality as such entails any particular account as to which inequalities are considered fair. What is required, in addition, is what Olsaretti calls a *principle of stakes*—the costs and benefits that persons ought to be held liable for or entitled to as a consequence of their choices. Olsaretti illustrates the claim in the context of a general discussion of responsibility-sensitive egalitarianism, not specifically related to issues of proactive justice. She draws on the case of a motorcyclist who is injured as a result of riding recklessly, often discussed in the context of responsibility-sensitive egalitarianism, and asks rhetorically:

Just what should the consequences of the motorcyclist's actions be? Do they include being left to the side of the road, even if this means that she might die there. . . . Or are they that she should be taken to a hospital and pay for the treatment of all her injuries. . . . And at what price should the treatment be charged. . . .? Are the consequences of her action also that passers-by may also appropriate her motorbike from the side of the road? That she may lose her job if . . . her limpness makes her a less attractive employee?¹⁵

It is clear, then, that responsibility-sensitive egalitarianism needs an account that answers such questions, and, furthermore, that such an account is not deducible merely from a commitment to responsibility-sensitive equality itself. In other words, the principle that only choices render inequality just does not entail a particular account as to which choices justify which inequalities.

As Olsaretti explains, responsibility-sensitive egalitarian views have often been presented with an implicit principle of stakes in mind—often, roughly, that people ought to be held liable for the “natural” consequences of their choice, or that they ought to be held liable for the actual consequences of their choice, where those actual consequences are determined by the uncoerced, equality-respecting choices of others.¹⁶ What Olsaretti and others show is that any such claims about the consequences persons ought to be liable for as a consequence of their choices do not follow from a commitment to fairness as such—namely the claim that persons are only justly worse off than others as a consequence of their own choices.¹⁷ We still require a further normative account concerning the

15 Olsaretti, “Responsibility and the Consequences of Choice,” 172.

16 Olsaretti, “Responsibility and the Consequences of Choice,” 174–81.

17 See also Dekker, “Choices, Consequences and Desert”; Fleurbaey, *Fairness, Responsibility and Welfare*; Stemplowska, “Making Justice Sensitive to Responsibility”; Vallentyne, “Brute Luck, Option Luck.”

respects in which individuals ought (or ought not) to be worse off than others, given their choices. The import of this for the procreative justice debate is that it seems, therefore, that an anti-sharing view does not follow from commitment to fairness thus understood. An alternative principle of stakes might, in fact, favor pro-sharing policies instead. An opposition to unfair inequality need not therefore commit us to parents' liability for the costs of children.

Although I think Olsaretti's position here is ultimately correct and indicates that a commitment to fairness does not entail a particular view about which costs, resulting from their choices, persons ought to internalize, it appears vulnerable to a compelling rejoinder. Understanding the nature of the rejoinder, and why it fails, helps to reveal the nature and importance of the distinction that I will defend in what follows.

3. BRUTE LUCK AND OTHER-AFFECTING CHOICE

The rejoinder runs along the following lines. It might be conceded that when we attend only to the *self-affecting* consequences of a person's choice, it is an open question which inequalities are fair and, as such, we require a principle of stakes to determine what it would be fair to hold persons substantively responsible for. We need, that is, an account that explains why certain choices ought to receive certain consequences. That much from Olsaretti's argument can be accepted. However, the rejoinder runs, when we attend to the *other-affecting* consequences of a person's choice, it is clear that the principle of stakes we affirm is itself constrained by considerations of fair equality. Because our choices *have an impact on other people*, what we hold people responsible for must take account of this. Specifically, the argument goes, negative consequences resulting from a person's choices that would otherwise render others *worse off* must be the chooser's liability, lest others be made worse off through no fault or choice of their own. If Olsaretti's motorcyclist crashes into another person's vehicle, causing costly damages, then the owner of the vehicle is worse off despite not being responsible themselves for being worse off. And yet responsibility-sensitive egalitarianism objects, on grounds of fairness, to persons being worse off than others in cases where that is not the result of their own responsible choices. As such, there seems to be an built-in, responsibility-sensitive egalitarian objection to any principle of stakes that fails to maintain that the costs of the motorcyclist's reckless riding should be internalized by her. This is because any such costs that are not borne by the motorcyclist herself look like they must be borne by others. Even if the motorcyclist does not crash into another's car but merely finds herself in need of hospital treatment, this treatment must be funded—if not by the motorcyclist then by someone else. Similarly, procreation imposes costs on others. The

introduction of new members of moral concern into the domain creates burdens of provision on others. Not only is the care of children costly, as we have said, but the adults they eventually become are members of the community that has justice claims on others that also involve costs. If parents are not held liable for those costs then someone else is required to cover them, and so the procreative choices of some would in that case make others worse off. A commitment to responsibility-sensitive equality would therefore seem to militate in favor of procreative liability. That is, cost internalization is implied by a responsibility-sensitive egalitarian principle of fairness precisely because other-affecting (non-internalized) costs constitute unchosen disadvantage for others.

Indeed, something along these lines seems to undergird the position of some of the foremost advocates of the anti-sharing stance. Casal and Williams, for instance, explicitly appeal to the equality of resources view offered by Ronald Dworkin.¹⁸ Briefly, Dworkin's view holds that the value of equality is best understood in terms of a measure of equality of resources, under which, very roughly speaking, persons are equal where they do not prefer one another's bundle of resources, even if, say, their bundle generates for them less welfare than others gain from their respective bundles. In Dworkin's terms, no one "envies" another's bundle.¹⁹ Dworkin articulates and defends his conception of equality of resources via a well-known desert island example in which shipwreck survivors run an auction using clamshells as currency in order to divide up the island's resources and arrive at a distribution that is free of "envy" as Dworkin defines it.²⁰ Crucially, this is a way of offering a measure of the value of resources that takes everyone's preferences into account. The value of any given resource is, through the auction, a function of what others would give up to procure it for themselves.²¹ In this way, the distribution is a fair one precisely because the bundle of resources devoted to each person's life is a function of the cost to others of this person having it. A defining merit of the Dworkinian view, then, is that the fairness of a distribution is defined in terms of the costs one's own life, and the resources devoted to it, impose on others. It is for these reasons that Casal and Williams conclude that there is an egalitarian injustice if the costs of children are shared among procreators and non-procreators in that it involves forcing the latter to pay for the costly choices of the former.²²

18 Casal and Williams, "Equality of Resources and Procreative Justice."

19 Dworkin, "What Is Equality? Part 2," 285.

20 Dworkin, "What Is Equality? Part 2," 285–90.

21 Dworkin, "What Is Equality? Part 2," 294.

22 Casal and Williams, "Equality, of Resources," 161–62.

I suggest that such a view obscures the two different bases for responsibility that I contend it is important to disaggregate. To understand why, it is important to distinguish between two different ways in which the choices of some can make others worse off. The first is the *absolute* sense—where someone is worse off than they were prior to a choice, irrespective of any comparison with how well off anyone else is. If one member of a village drains a well on Tuesday, all of the village members are without water and thereby absolutely worse off than they were prior to the well being drained on Monday, given that they are left without water, even though none is worse off than any other. The second is the *person-comparative* sense—where someone is worse off in comparison with others. If the draining of the well only deprives some village members of water and not others, those without water are person-comparatively worse off than those with a water supply.

Applying this to the issue of procreative justice, let us assume that while the procreative choices of parents might render everyone absolutely worse off, the costs are shared in such a way that no one is made person-comparatively worse off—both parents and nonparents are made worse off to the same extent.²³ In this instance it might seem that responsibility-sensitive egalitarianism will *not* have a fairness-based objection to sharing the costs (even if particular principles of stakes object to it). It will not have grounds to object, that is, on the basis that persons are person-comparatively worse off through no choice of their own. After all, no one is person-comparatively worse off under the cost-sharing scheme. It seems, then, that a cost-sharing principle of stakes of this kind will not run afoul of fairness considerations written into to responsibility-sensitive egalitarianism. However, although in one respect nonparents are not made person-comparatively worse off than parents, they *are* person-comparatively worse off in another respect. To appreciate this, it is worth attending to the distinction at the center of responsibility-sensitive egalitarian views, between disadvantages that result from an individual's own choices and those that do

23 It is worth noting that Casal and Williams suggest that even in cases where procreation does not leave nonparents worse off than parents, there would still be reasons to require parents to internalize the costs of their procreative choices. Even here, though, their argument seems to depend on an implicit appeal to the comparative advantage parents enjoy. They compare the case of parents demanding others share in the costs of procreation to the case of Louis, offered by Dworkin, who calls for additional resources to satisfy his voluntarily acquired expensive tastes. Casal and Williams cite approvingly Dworkin's conclusion that "It is quite *unfair* that [Louis] should . . . be able, at the expense of others, to lead a life that is *more expensive* than theirs." Casal and Williams, "Rights, Equality, and Procreation," 110, emphasis added. See also their "Equality of Resources and Procreative Justice," 161. It is the contention of the paper that the Dworkin conclusion in fact *fuses* fairness and cost internalization in a way that would be helpfully distinguished.

not—their *brute luck*. Responsibility-sensitive egalitarianism, as I understand it, is committed to the fairness of (at least some) inequalities reflecting the former, and the unfairness of inequalities reflecting brute luck. So, in respect of those gains and losses that *are not the result of (reasonably anticipated, avoidable) choices*, persons ought to be equal.

Under this responsibility-sensitive egalitarian approach, it *does seem* that there is a meaningful sense in which nonparents are person-comparative worse off under a cost-sharing scheme. Let us say that the procreative choices of parents halve everyone's level of absolute advantage (in terms of whatever one's favored metric of advantage is). Nonparents are absolutely half as well off as they were as a result of bad brute luck—namely not as a result of their *own* choices but the procreative choices of parents. In contrast, while parents are also absolutely half as well off as they were, this is a direct result of *their own* choices. As such, they are *not brute-luck worse off at all* but as a result of the exercise of their own responsibility. So it seems that there is an objectionable inequality between parents and nonparents in the respect that the latter experience a form of bad brute luck that the former do not. Unless parents pay for the costs of their own procreative choices, those costs are borne by nonparents in a way that makes them *person-comparatively worse off in brute-luck terms*. Anything less than full cost internalization on the part of parents seems to entail a kind of inequality (of brute luck) that responsibility-sensitive egalitarians explicitly oppose on grounds of fairness. It seems, then, that responsibility-sensitive egalitarianism, if it is committed to brute-luck equality, must force parents to internalize the costs of their choices and must do so on grounds of fairness (that no person should be person-comparatively worse off, other than as a result of their own choice).

Despite the surface appeal of this argument, in fact it reflects what I argue is the conflation between cost internalization and fairness when it comes to the place of responsibility in egalitarian thinking. This is because it trades on a mistaken supposition that inequality in brute luck and unfair inequality are coterminous.²⁴ Once we appreciate that inequality in brute luck and *unfair* inequality are two quite different things, we can better understand that there are two importantly separable ideas playing a role in much egalitarian thinking. In turn we can see that the anti-sharing view defended by Casal and Williams and others cannot trade on fairness alone to contend that the costs of children ought not to be shared, but must be supplemented by a further specific account of why the costs of children ought to be internalized by procreators. This is significant because it demonstrates that fairness *as such*, despite having powerful appeal, does not militate in favor of an anti-sharing stance.

24 See Elford, "Equality and Other-Affecting Choice."

Consider the following simple example to illustrate why brute-luck inequality is not unfair. Michael and Martellus both have ten units of brute-luck advantage. Michael gives Martellus five units of his bundle of advantage units, resulting in the distribution—Michael five, Martellus fifteen. Responsibility-sensitive egalitarians for the most part have no objection to the resulting unequal distribution. Because the inequality between Michael and Martellus is a result of Michael's choice, the inequality is a fair one. This is the responsibility sensitivity of such egalitarian views in action. Notice, though, that not only is there an inequality in advantage (of ten units) but there is also an *inequality in brute-luck* advantage—Martellus's five-unit gain is, from his perspective, a matter of brute luck and not of a choice he made (but a choice of Michael's).²⁵ So the five added units represents a brute-luck gain that Martellus enjoys that Michael does not. And yet this does not seem at all unfair. It does not seem unfair because even though Michael is worse off in brute-luck terms, it is the result of a choice he made (he chose in a way that made Martellus brute-luck better off).²⁶ He could have avoided the brute-luck inequality if he had refrained from giving to Martellus, or, indeed, if he had simply destroyed five units of his bundle. So, I suggest, provided Michael had the (equal) opportunity with Martellus to be as well off, there is no egalitarian unfairness in his being worse off.²⁷ Michael could have remained equal with Martellus at ten units each had he refrained from giving five units.²⁸

Let me offer a further example to more clearly demonstrate the import of the discussion for debates over procreative justice. Suppose that Melody and Matilda live next door to one another and have contrasting tastes in music—Melody likes jazz and Matilda likes country. While both enjoy their respective favorite forms of music, they also appreciate peace and quiet and prefer it at all

- 25 It might be said that Martellus's gain is *not brute luck*, given that his receipt of the five-unit gain will be a consequence of a choice he makes—namely to accept the gain. I do not think this is quite right. First, it is not generally true that in order to enjoy a benefit one must "accept" it; some benefits are unavoidable. Second, and still more pertinent, I would maintain that the mere "opportunity to accept" the five-unit gain is itself an advantage in the relevant sense. To put it another way, a person who enjoyed a far greater range of significant opportunities to gain than others would be advantaged compared with those others even before they chose whether to take advantage of those opportunities.
- 26 One might also maintain that in order for the inequality to be fair it must be reasonable to expect Michael to anticipate his choice resulting in this comparative disadvantage.
- 27 Subject to one's favored account of stakes. That is, one's favored forms of responsibility-sensitive egalitarianism might deny that a choice should result in a given liability. In which case, the fairness of the inequality is not vindicated.
- 28 For the opportunity to be as well off as Martellus to be *equal* it also must be true that Martellus is able (and we might suppose no generally less inclined) to transfer five units to Michael, should he so choose.

times over the other's favored music. Both need to play the music to a certain volume in order to appreciate it and at that volume it is audible in the other's property, supplanting peace and quiet. When Melody plays jazz it makes Matilda worse off than she would have been with peace and quiet and vice versa when Matilda plays country. Assume that the advantage levels of Melody and Matilda are each ranked as follows:

Playing their preferred music alone > both playing preferred music >
neither playing preferred music > not playing music while the other
plays.

Suppose that Melody alone plays. Is this state of affairs unfair? Although Matilda is person-comparatively worse off than Melody and, as I explained above, worse off in brute-luck terms (she endures jazz, while Melody does not endure country), she is nevertheless *person-comparatively* worse off in consequence of her own choices. She has the opportunity to be as well off as Melody. Now, there might be *non-egalitarian* reasons for thinking it is bad if either or both Melody and Matilda play their music.²⁹ The point of relevance, though, is that the mere fact that one person imposes costs on another does not entail that the worse-off person has an egalitarian complaint of unfairness, even though the person is absolutely worse off through no choice of their own. And in fact we tend to find this thought quite intuitive. Take job applications under conditions of background equal opportunity. If Abi and Isobel both have equal opportunity for a job and Abi has already applied, Isobel's applying renders Abi worse off than she would otherwise be if she did not face competition, now that her prospects for the job are worsened with this new competition.³⁰ If Isobel does not apply and hands a brute-luck advantage to Abi in that regard, there is no unfairness given that it is Isobel's choice that results in her being worse off—she could be as well off as Abi should she have so chosen.³¹

The fact that cost imposition does not itself entail unfairness is significant, and illuminates the sense in which the anti-sharing position held by Casal and

29 And, to complicate matters, any favored principle of stakes could require that music-playing costs should be internalized.

30 Assuming Isobel is a genuine competitor, with relevant prospects of securing the job herself.

31 For a similar point see Olsaretti, "Children as Negative Externalities?" 160–61. Isobel's choice is brute luck for Abi only insofar as it is not reasonably foreseeable for Abi when choosing whether or not to apply. If it is something that is, in fact, reasonably foreseeable, then it figures as a determinant of the option luck (in this case, the prospects of the success of the application) resulting from Abi's choice. Similarly, from Isobel's perspective, Abi's applying makes her brute-luck worse off than had Abi not applied, given that it lessens her prospects for a successful application. I am grateful to an anonymous reviewer for drawing my attention to these aspects of the example.

Williams cannot be supported with the claim that a cost-sharing scheme would necessarily entail unfair inequalities in brute luck. If indeed we ought to hold parents liable for the costs of children, this does not straightforwardly follow from a commitment to egalitarian fairness, even of a responsibility-sensitive variety.³² So if each person has an equal opportunity to bear and parent children, then even though nonparents might be made absolutely worse off by sharing in the costs of caring for children that parents are responsible for creating, even if they are *person-comparatively worse off than parents in brute-luck terms* (because the costs they share are not the result of their own choices but result from parents' choices) this does not entail egalitarian unfairness provided that nonparents had the equal opportunity to be parents themselves.³³ The conclusion toward which we have been moving, then, is that procreative liability cannot be supported by a commitment to responsibility-sensitive egalitarian fairness alone. The anti-sharing view therefore requires going beyond egalitarian fairness, thus understood, and necessitates an appeal to a further claim about the costs persons ought to internalize. To be clear, I am not suggesting that the anti-sharing stance is particularly afflicted here, or that this is a reason itself to reject an anti-sharing view. Rather, I am suggesting that an anti-sharing position in the context of procreative justice is apt to seem appealing on grounds of fairness alone but, in fact, depends on further, and contestable, suppositions as to what persons ought to be held liable for.

4. "FAIRNESS" AS BOTH FAIRNESS AND COST INTERNALIZATION

My claim, then, is that an anti-sharing position (and, indeed, any claim about the liability persons have for imposing costs on others) must go beyond an appeal to fairness (i.e., that no person should be person-comparatively worse off unless it is because of a choice of their own) and must depend on an account as to why any given costs ought to be internalized. To fully defend that claim, let me consider two possible objections. First, it might be argued that my way of framing the distinction between fairness and cost internalization fails to take due account of the notion of fairness with which the advocates of the anti-sharing stance defend their view. I mentioned in the foregoing that the anti-sharing position defended by Casal and Williams takes its inspiration from Dworkin's conception

32 I am of course assuming that each person has an equal opportunity to bear and care for children, which is far from the case in the real world.

33 Crucially, an "equal opportunity to be parents" is intended to include the advantages and disadvantages that accrue from bearing and caring for children. For instance, if one set of parents is likely to find it inherently more burdensome than another set of parents, then they do not enjoy an "equal opportunity" to be parents in the relevant sense. I am grateful to an anonymous reviewer for pressing me to clarify this.

of equality of resources. It seems, though, that under that equality of resources view, “fairness” is conceptualized as something that entails *both* fairness, as I conceive it, and cost internalization. On such a view, the distinction between fairness and cost internalization that I have attempted to draw in the preceding discussion is wrongheaded; “fairness” depends on a given individual’s life not being, in certain ways, more costly than the lives of others. Indeed, Dworkin draws the connection between fairness and cost internalization explicitly when he explains that it is a compelling feature of his equality of resources view that “people decide what sorts of life to pursue against a background of information about the actual cost their choices impose upon other people and hence on the total stock of resources that may fairly be used on them.”³⁴ Dworkin writes this in a context where he defends resources as the appropriate measure of egalitarian concern in preference to one of its foremost rivals, equality of *welfare*. One of the critical deficiencies of the equality of welfare view, for Dworkin, is a failure to take account of the costliness to others of satisfying any given individual’s preferences. On the Dworkin view, then, it seems that cost imposition is unfair because the persons who impose the costs have *an unfairly large proportion of resources devoted to their life* as compared with the lives of others. Those with more expensive preferences gain a greater share of the resources under equality of welfare in a way that is wholly insensitive to the loss in resources felt by others.³⁵

However, if that is the sense in which cost imposition is required on grounds of fairness, then it is just not clear that this is required in respect of *other-affecting choice* at all. After all, if *A* imposes costs on *B*, then a failure to internalize those costs (e.g., requiring *A* to compensate *B*) does not entail devoting more resources to the life of *A*. To be sure, if *A* imposes costs on *B*, it makes *B* absolutely worse off (this is just what it means to impose costs in this sense). It does not, however, involve using more of the total stock of resources (or whatever currency) on the life of *A*. To reflect this back to the procreative justice debate: the Dworkinian-inspired objection to cost sharing looked like there was something wrong with nonparents being required to shoulder the costs created by the choices of parents. But this is not a matter of parents gaining more resources than nonparents, but the instantiation of new claimants on resources (children), which under a cost-sharing scheme diminishes the resources available to nonparents (and parents alike). As such, the Dworkinian position—that we should not devote an unfairly large proportion of resources to some lives and not others—does not constitute a basis for objecting to a cost-sharing scheme, given that such a scheme does not require devoting more resources to

34 Dworkin, “What Is Equality? Part 2,” 288.

35 Dworkin, “What Is Equality? Part 1,” 228–40.

some lives over others. Again, fairness—understood now in these Dworkinian terms—does not necessarily commend an anti-sharing stance.

In response to this last claim, perhaps it might be argued that cost imposition through a cost-sharing scheme *does* involve unfairly devoting more resources to the lives of parents who create those costs, because cost sharing is, in effect, a mechanism for compensating parents for the costs they incur as a result of choosing to have children. It therefore represents a way of giving more to parents who have already spent part of their fair share by choosing to bear children. Clearly, though, this argument will not succeed as it begs the question by presupposing that these are parents' costs to bear in full to begin with.³⁶ They *may* be the parents' costs to bear, but we need an independent and prior argument to establish this.

Nor, indeed, can this Dworkinian opposition to cost sharing be defended simply on the basis that under a cost-sharing scheme the choices of parents render nonparents person-comparatively worse off in a way that is outside of the nonparents' control. Now, it is worth registering that a cost-sharing scheme *need not necessarily* make the nonparents person-comparatively worse off than parents. If we have a scheme that takes account of the benefits that parents enjoy and levies higher costs on them to discount for those benefits, then a scheme will not make nonparents person-comparatively worse off than parents. But more importantly, we have already seen that being worse off in this regard (suffering worse brute luck than parents) is not necessarily unfair. Provided that nonparents have the relevantly equal opportunity to bear (and benefit from) children themselves, nonparents are person-comparatively worse off (in, say, brute-luck terms) *qua* remaining nonparents only to the extent that they refrain from bearing and caring for children themselves—as per the argument above. So although nonparents are person-comparatively worse off, they are not person-comparatively worse off *in a way that is relevantly outside of their control* (as they could have imposed the very same costs on parents).

I have thus far argued that the Dworkinian view under scrutiny fails to establish that fairness alone commends an anti-sharing view. It might be said, though, that under the Dworkinian way of conceiving fairness, “fairness” is simply a more complex conception than the thin egalitarian form of fairness that I am describing. Rather, “fairness” amalgamates both the idea that each

36 Olsaretti offers a similar counter when she explains that “it is unjustified, and importantly misleading, to characterize the demand that nonparents share the costs of children as a demand that they ‘settle for less than their fair shares after resources have been divided justly’ [Rakowski, *Equal Justice*, 153, emphasis added]. . . . [This] assumes what is in question, namely, that what constitutes a fair share for a person” (Olsaretti, “Children as Negative Externalities?” 160).

ought to have an equal opportunity to be as well off as others (thin fairness) and a further principle that requires cost internalization. No doubt this is true. And no doubt one can present a version of Dworkin-inspired equality of resources that supports an anti-sharing position. However, it should be clear that this is no answer to the chief contention of the paper that there is an important difference between fairness and cost internalization. In that way it is quite coherent to present such a conception of “fairness,” but we should also be clear that it is a *complex ideal* with distinct (albeit connected) component elements. That is, Dworkinian “fairness,” thus understood, is an umbrella category that encompasses both fairness (as I have articulated it) and a certain view about the costs that individuals ought to internalize. Unless we are clear about that, we may be in danger of erroneously supposing that a commitment to thin egalitarian fairness (no one should be person-comparatively worse off other than in consequence of their own responsible choices) necessarily commends a cost-internalization principle. And, in virtue of that error, we may miss the need for a positive account of the costs it is just to impose on others

Indeed, to further understand that fairness is different from, and does not entail, cost internalization, note that strict cost internalization is quite implausible. Every student of John Stuart Mill is familiar with the challenge to Mill’s famous harm principle that few if any of our acts do not impact others in some way.³⁷ If nearly all of our actions can in one way or another be said to “harm” others, however marginally or indirectly, then the harm principle lacks efficacy as a way of discriminating between harming and non-harming actions and protecting the latter from state and societal interference. Whatever the merits of this as a criticism of Mill specifically, it is of import for the issue under scrutiny. For our choices routinely impose a variety of costs on others. Whenever there is competition for a scarce resource, for instance, one person’s pursuit of it characteristically diminishes another’s opportunity to obtain it. When other people choose to go shopping on a Saturday afternoon it makes what might have otherwise been a stress-free, leisurely perusal of the shops for me a hectic, busy, and time-consuming affair. If cost internalization is required on grounds of fairness, then each and every one of these cost-imposing activities inflicts unfairness.³⁸ The idea that there is some egalitarian requirement that we internalize all of the costs of our own choices is deeply implausible, and, indeed, is not something that

37 Mill, *On Liberty*.

38 Now, it is worth saying that unfairness can indeed be trivial. The egalitarian conclusion, though, is that, if such costs and benefits (trivial or otherwise) even out across persons (assuming they are commensurable), there is no unfair inequality. As such it is hard to see how grounds of egalitarian fairness could condemn the imposition of any particular set of costs on others *before* we first know whether or not the imposed cost manifests an inequality.

those who affirm the anti-sharing position (on Dworkinian or other grounds) would themselves accept.³⁹ An anti-sharing position therefore depends on going beyond an appeal to fairness and requires support from an account as to why the costs of children specifically are unreasonable ones to impose upon others.

Notice that there is a still stronger reason why, under this Dworkinian egalitarian view, cost internalization is not entailed by fairness (in the sense in which I intend the term). This is that knowing what costs one will be required to bear in consequence of different activities is precisely part of what one needs to know in advance in order to determine the value of the resource in question.⁴⁰ That is to say, the value of any given resource depends partly on what one is permitted to do with it and, in that sense, the extent to which one is permitted to use it in ways that impose costs on others that one is not required to internalize. If I need to compensate my neighbor for the noise that starting my car creates each morning, the value of the car is less than it might have been without my being required to do so. Likewise, the value of procreating also depends on the costs of doing so (the resources used for procreation, the levels at which I might want to insure against the need for treatment to successfully procreate, etc.) and whether one is required to internalize them. This again indicates that cost internalization is not something that should be confused with a principle of fairness. Rather, at least on the Dworkin conception of equality of resources, certain issues of cost internalization should be seen as prior to the determination of what constitutes an equal and fair distribution, given that the value of those resources depends on how far one is required to internalize the costs of their use. For this reason there is no basis for claiming that on the grounds of fairness alone people are required to internalize the costs of the choices.

Before concluding, let me respond to a final reservation concerning my claim that considerations of fairness alone do not support an anti-sharing view. It might be said that parents and nonparents differ, for brute-luck reasons, in their preferences for birthing and rearing children, and that a cost-sharing policy therefore subsidizes the preferences of parents in a way that makes parents unfairly *person-comparatively* better off. Because parents' preferred way of life is

39 I suggest, then, that on the Dworkinian conception of equality, the connection between fairness and cost internalization is best understood as pertaining to the process of arriving at a determinate value for resources, rather than as a matter of requiring, on grounds of fairness, that persons internalize the costs of their choices they would otherwise impose on others. In other words, costliness to others figures in determining the fair distribution insofar as it provides a criterion of value for the metric of equality—that every person's preferences are equally taken into account when determining the value of the resources to be distributed.

40 For this point made in a different context, see Miller, *National Responsibility and Global Justice*, 56–60.

made cheaper than that of nonparents, parents have an unfairly better opportunity for advantage compared with nonparents. We therefore have fairness reasons for requiring parents to internalize the costs of their procreative decisions. Such a claim depends, of course, on a metric of advantage that includes preference satisfaction, but it is not for that reason misguided. Rather, it fails as an objection to the contentions of the paper for following reason—that it does not depend on the principle that *fairness requires persons to internalize the costs of their own choices*, but is rather based on the principle that the *brute option sets available to persons should be no more valuable than any other person*. We can see this by noting that it is contingent on whether subsidizing procreative choices will make parents person-comparatively better off than nonparents in brute-luck terms. Suppose, for brute-luck reasons, that Jennifer has an overwhelming desire to be a parent but Kayla does not, and that being a parent is far more expensive, overall, than any of Kayla's preferred pursuits. On the preference satisfaction view we are considering it seems that, all else equal, *subsidizing* Jennifer's procreative choices helps equalize the value of the option sets she and Kayla respectively face by making her preferred option (to parent) no worse, in terms of overall preference satisfaction, than Kayla's. In this instance, subsidizing certain options is what is *required for* a fair distribution, rather than disruptive of it. So it is simply an open question whether subsidizing the option to parent creates unfair inequality—it depends entirely on whether doing so upsets an extant fair distribution. Again, then, a commitment to fairness (that none should be worse off other than in consequence of their own choices) does not imply a requirement that individuals internalize the costs of their choices.

5. FAIR PLAY, STAKES, AND FAIRNESS

As some of the earlier examples illustrate, the idea that persons ought to internalize all of the costs of their choices is deeply implausible and not a position to which any of the authors defending an anti-sharing stance would subscribe. What is required is a constructive substantive account of which costs resulting from person's choices they ought to be required to internalize themselves. To frame this in some of the language deployed to explain Olsaretti's view, what is required is a principle of stakes—a specification of the costs and benefits for which persons ought to be held substantively responsible as a result of their choices. Now, as I mentioned in the foregoing, much of the current literature on procreative justice focuses on whether the beneficial effects of procreation can ground an enforceable duty on non-procreators to contribute to the costs of procreation. To that extent, much of the basis of the disagreement between pro-sharing and anti-sharing views involves contestation as to whether fair-play

considerations mean that the beneficiaries of procreation ought to reciprocate for such benefits by way of sharing in the costs of procreation.⁴¹ That is to say, do the putative benefits of procreation and parenting that accrue to nonparents ground an obligation to contribute to the costs of procreation and parenting? I want to remain agnostic with respect to the success, or otherwise, of the argument from fair play applied to the procreative question. Rather, I want to suggest that the benefits of children that accrue to society resulting from the procreation and parenting of children are relevant directly to the principle of stakes.

To see how the benefits for society offered by the existence of children might play a different role than they do in the argument from fair play, consider the parallel case of the fire service. Roughly, on the standard rehearsal of the fair-play argument in favor of an obligation to contribute, individuals have fairness-grounded obligations to reciprocally contribute to a scheme that provides them with benefits, where others shoulder burdens necessary to supply those benefits, and provided the benefits accruing to the individual outweigh the burdens of contribution. Applying the fair-play argument to fire service provision specifically, those who supply fire service protection, through funding and working in a fire department, provide a benefit to the recipients of that service, which in turn grounds an obligation on the part of those recipients to reciprocate by way of sharing in the burdens of contribution required to provide the service.⁴² Unless the burdens are shared, those who do not contribute to the provision of fire service are in some way free riding on others—they would be taking unfair advantage of others' labor burdens. Notice that this argument does not trade on the idea that those funding or working in the fire service would be unfairly worse off than others as a result of the burdens they willingly bear. Rather, it claims that as recipients of benefits we owe our *fair share* of the contribution for the provision of this good. As such, it is fairness with respect to the burdens, given that everyone enjoys the good, rather than fairness with respect to how much advantage individuals enjoy overall. The fire service worker is owed public money so that they do not unfairly shoulder too great a proportion of the burden of a general public benefit. However, the beneficial character of fire service provision might be thought important in another way. If we assume a responsibility-sensitive egalitarian view that maintains that inequality is fair *only if* it is the result of responsible choice, we then need to appeal to a principle of stakes to identify which consequences of responsible

41 Hart, "Are There Any Natural Rights?"

42 I offer the case of fire-service provision as an illustration. Fire service is ordinarily provided as a component of a broader publicly funded scheme of social cooperation under which any obligations of contribution may take into account other benefits from the scheme and the costs of their provision.

choices are ones that persons ought to be held substantively responsible for such that they would constitute fair inequalities. When determining a principle of stakes, one might argue that the characteristically beneficial character of fire service provision counsels against holding fire service personnel liable for the costs encountered in the course of providing that service, say, for example, through being injured when attending a fire.

It might seem that there is no difference between the fair-play argument and the principle-of-stakes position I have gestured toward. They both, for instance, ground the fairness of burden sharing on the beneficial character of what is provided. The differences between them, however, are subtle but important. First, under the principle of fair play, the duty of alleviating the burdens on fire service personnel necessarily falls on the recipients of the benefit of the service, whereas it is an open question, on the principle-of-stakes position, who has the responsibility to compensate for any costs incurred by fire service personnel.⁴³ Second, on the principle of fair play, the compensation owed to fire service personnel is contingent on their successfully providing a benefit. Even if the benefit in question is only an expected benefit, on standard versions of the fair-play argument it is the *fact of the benefit* that grounds the obligation to reciprocally share in the burdens of provision. On the principle-of-stakes position, the grounds for compensation are not located in the *actual receipt of benefit* but in the character of the choice. It is the nature of fire service provision as a vocation that it is *characteristically beneficial* to the community that renders it inappropriate, on this view, to hold persons substantively responsible for certain costs of that choice to contribute to that provision. To put it a slightly different way, on the principle-of-stakes account, we owe aid to fire personnel not by way of paying them back for some contribution that they would otherwise unfairly shoulder but, rather, because the choice they make just is not of the right kind to render them liable for the costs they would otherwise incur. Such an approach could be grounded on desert-based considerations—that the putatively valuable character of fire service contribution renders it “unfitting” to hold people liable for certain negative consequences—or grounded on consequentialist considerations—that it is good to incentivize certain generally socially valuable contributions.⁴⁴

Distinguishing the principle-of-stakes account from the fair-play argument is important because the fair-play account, at least applied to the issue of proactive justice, is subject to several strong objections, and it is possible that a principle-of-stakes view sidesteps some of the controversy concerning

43 Though there may be second-order duties on others to ensure that the recipients are held to their duties.

44 For a discussion of such approaches, see Olsaretti, “Responsibility and the Consequences,” 182–86.

the relation between being a beneficiary and owing something reciprocally in return.⁴⁵ A defense of a pro-sharing view based on a principle of stakes—one that, for instance, emphasizes the *characteristically* beneficial nature of procreative choices—reframes the procreative justice question as one that asks “Do procreating choices embody the kind of choices that justify a person being worse off than others?” as opposed to asking “Does the receipt of the benefit produced by the choice oblige recipients to bear a fair share of the burdens involved in its production?” Constructing such an account extends well beyond the purview of the paper, but the space for a fair-play-inspired principle of stakes points toward a possible fruitful alternative basis for defending a pro-sharing view.

Finally, then, let me reaffirm why I think it is important to distinguish cost internalization from fairness as a basis for holding people responsible for their choices. Quite apart from the general desirability of perspicuity concerning the reasons that support different normative positions, it is important because it helps us guard against the temptation to think that a concern for *fairness* commands a commitment to holding persons liable for the costs of their choices on others. Failing to distinguish the issue of cost internalization from the issue of fairness thereby embodies the danger that we go awry when thinking about how those considerations are related to one another. Recognizing their distinctness helps us avoid the misguided supposition that by dint of consistency with our commitment to egalitarian fairness we must also conclude that persons should be required to shoulder costs that they would otherwise impose on others. The issue of which inequalities are fair in fact depends on a prior account of the costs it is reasonable to require persons to internalize, and not the other way around.⁴⁶

University of Oxford
gideon.elford@politics.ox.ac.uk

REFERENCES

Arneson, Richard. “Equality and Equal Opportunity for Welfare.” *Philosophical Studies* 56, no. 1 (May 1989): 77–93.

45 For an account that maintains that, in order for the fair-play principle to apply, the benefits in question must be *accepted*, see Simmons, “The Principle of Fair Play.” For the claim that the benefits must be *intended*, see Casal, “Environmentalism, Procreation, and the Principle of Fairness”; and Casal and Williams, “Rights, Equality, and Procreation.”

46 I am grateful to an anonymous reviewer for some very helpful comments. I am also indebted to Andrew Williams for some eminently helpful thoughts on an earlier draft of the paper.

- Bou-Habib, Paul, and Serena Olsaretti. "Equality, Autonomy, and the Price of Parenting." *Journal of Social Philosophy* 44, no. 4 (Winter 2013): 420–38.
- Casal, Paula. "Environmentalism, Procreation, and the Principle of Fairness." *Public Affairs Quarterly* 13, no. 4 (October 1999): 363–75.
- Casal, Paula, and Andrew Williams. "Equality of Resources and Procreative Justice." In *Dworkin and His Critics*, edited by Justine Burley, 150–69. Oxford: Blackwell, 2004.
- . "Rights, Equality, and Procreation." *Analyse und Kritik* 17 (1995): 93–116.
- Clayton, Matthew. *Justice and Legitimacy in Upbringing*. Oxford: Oxford University Press, 2006.
- Cohen, G. A. "Fairness and Legitimacy in Justice, and: Does Option Luck Ever Preserve Justice?" In *On the Currency of Egalitarian Justice and Other Essays in Political Philosophy*, edited by Michael Otsuka, 124–43. Princeton, NJ: Princeton University Press, 2011.
- . "Luck and Equality: A Reply to Hurley." *Philosophy and Phenomenological Research* 72, no. 2 (March 2006): 439–46.
- . "On the Currency of Egalitarian Justice." *Ethics* 99, no. 4 (July 1989): 906–44.
- Dekker, Teun. "Choices, Consequences and Desert." *Inquiry* 52, no. 2 (2009): 109–26.
- Dworkin, Ronald. "What Is Equality? Part 1: Equality of Welfare." *Philosophy and Public Affairs* 10, no. 3 (Summer 1981): 185–246.
- . "What Is Equality? Part 2: Equality of Resources." *Philosophy and Public Affairs* 10, no. 4 (Autumn 1981): 283–345.
- Elford, Gideon. "Equality and Other-Affecting Choice: Why Luck Egalitarianism Does Not Require Brute Luck Equality." *Ethical Theory and Moral Practice* 16 (2013): 139–49.
- Fleurbaey, Marc. *Fairness, Responsibility, and Welfare*. Oxford: Oxford University Press, 2008.
- George, Rolf. "Who Should Bear the Cost of Children?" *Public Affairs Quarterly* 1, no. 1 (January 1987): 1–42.
- Hart, H. L. A. "Are There Any Natural Rights?" *Philosophical Review* 64, no. 2 (April 1955): 175–91.
- Malthus, Thomas. *An Essay on the Principle of Population*. 1798. Edited by Donald Winch. Cambridge: Cambridge University Press, 1992.
- Mill, John Stuart. *On Liberty*. London: Parker and Son, 1859.
- Miller, David. *National Responsibility and Global Justice*. Oxford: Oxford University Press, 2007.
- Olsaretti, Serena. "Choice, Circumstances, and the Costs of Children." In *Hillel Steiner and the Anatomy of Justice*, edited by Stephen de Wijze, Matthew H.

- Kramer, and Ian Carter, 70–84. Abingdon, UK: Routledge, 2009.
- . “Children as Negative Externalities?” *Politics, Philosophy and Economics* 16, no. 2 (May 2017): 152–73.
- . “Children as Public Goods?” *Philosophy and Public Affairs* 41, no. 3 (Summer 2013): 226–58.
- . “Responsibility and the Consequences of Choice.” *Proceedings of the Aristotelian Society* 109, no. 1 (August 2009): 165–88.
- Rakowski, Eric. *Equal Justice*. Oxford: Clarendon Press, 1993.
- Simmons, A.J. “The Principle of Fair Play.” *Philosophy and Public Affairs* 8, no. 4 (Summer 1979): 307–37.
- Steiner, Hillel, and Peter Vallentyne. “Libertarian Theories of Intergenerational Justice.” In *Intergenerational Justice*, edited by Axel Gosseries and Lukas H. Meyer, 50–76. Oxford: Oxford University Press, 2009.
- Stemplowska, Zofia. “Making Justice Sensitive to Responsibility.” *Political Studies* 57, no. 2 (June 2009): 237–57.
- Temkin, Larry. *Inequality*. New York: Oxford University Press, 1993.
- Tomlin, Patrick. “Should Kids Pay Their Own Way?” *Political Studies* 63, no. 3 (August 2015): 663–78.
- Vallentyne, Peter. “Brute Luck, Option Luck, and Equality of Initial Opportunities.” *Ethics* 112, no. 3 (April 2002): 529–57.
- . “Equality and the Duties of Procreators.” In *The Moral and Political Status of Children*, edited by David Archard and Colin M. Macleod, 195–210. Oxford: Oxford University Press, 2002a.

IN SEARCH OF A STABLE CONSENSUS

RAWLS'S MODEL OF PUBLIC REASON AND ITS CRITICS

Cyril Hédoïn

RAWLS's political philosophy figures as the most important contribution to what is sometimes called *public reason liberalism*. The concept of public reason, however, makes its appearance only in Rawls's late writings, especially *Political Liberalism* and "The Idea of Public Reason Revisited." The central place that this concept occupies in these writings is associated with Rawls's "political turn" that marks the transition from *A Theory of Justice* to *Political Liberalism*. The recent scholarship on Rawlsian political philosophy has established that this political turn follows Rawls's attempt to overcome the unsatisfactory treatment of the problem of stability of the well-ordered society in *A Theory of Justice*.¹ The contours of Rawls's later solution to this problem are relatively well known. Principles of justice are merely political and should be publicly endorsable by each person within one's "comprehensive doctrine." This establishes an overlapping consensus that guarantees that each member of the society has all-things-considered reasons to support a shared liberal conception of justice.

A key feature of this Rawlsian account of stability is the requirement that the members of the well-ordered society abide by principles of justice for *shared public reasons*. This "stability for the right reasons" requirement has recently been the target of a series of criticisms. The criticisms are mainly internal to public reason liberalism and come from post-Rawlsian scholars who endorse what I shall call a *diversity-convergence account of stability*. They emphasize in particular the inability of public reason to solve the assurance problem with which the members of the well-ordered society are confronted. This paper discusses the debate over Rawls's model of public reason as an account of stability that these criticisms trigger. I suggest a way to preserve the Rawlsian account of stability based on the possibility that members of the well-ordered society can reasonably make use of what I call *community-based reasoning* to solve the assurance problem. This solution is nonetheless fragile in the presence

1 Weithman, *Why Political Liberalism?*

of unreasonable persons. This result calls for a reassessment of the project of public reason liberalism.

The rest of the paper is structured as follows. Section 1 presents the problem of stability in Rawlsian political philosophy. Section 2 characterizes Rawls's model of public reason as an account of stability. Section 3 discusses the various criticisms of Rawls's account of stability developed by proponents of the diversity-convergence account of stability. Section 4 suggests an alternative solution in terms of community-based reasoning that remains compatible with Rawls's model. Section 5 considers the prospects of establishing a well-ordered society based on a minimal set of shared reasons when this last solution fails. Section 6 briefly concludes.

1. THE PROBLEM OF STABILITY IN RAWLSIAN POLITICAL PHILOSOPHY

Rawls's theory of justice is part of the very large family of social contract theories. As are most—if not all—normative accounts belonging to this family, it is confronted with two related problems: the *justificatory problem* and the *stability problem*. The justificatory problem can be formulated as follows: On what basis can a set P of normative principles about what is right and good be justified to some set N of persons? The stability problem is concerned with a different question: What are the conditions, if any, under which the members of N will abide by P ?

There are several ways to meet these requirements. Contractualist moral and political theories generally solve the justificatory problem by positing a set N^* of idealized individuals placed under idealized circumstances C^* . The characteristics of N^* and C^* are such that it is claimed that the set of principles P over which the members of N^* will agree (either through a unanimous choice or a compromise through a bargain) have a particular normative significance. The stability problem then surfaces immediately. Even if the set of principles P is appropriately justified, the fact that the nonidealized members of N are put under nonidealized circumstances C does not guarantee that the principles P will be respected and implemented. A tension between the justificatory problem and the stability problem thus appears.² There are at least two reasons to impose severe restrictions on N^* and C^* . First, our considered intuitive judgments may lead us to think that only agreements based on specific reasons or motivations and reached under specific circumstances can count as morally relevant. Second, idealization is helpful to restrict the candidate principles potentially belonging

2 See Thrasher, "Agreeing to Disagree," for a similar characterization that emphasizes the tension between what he calls the existence problem and the stability problem.

to P . Insufficiently restricted characterizations of N^* and C^* run the risk of leading to complete indeterminacy. But of course, the more N^* and C^* are idealized, the less likely it is that the members of N (actual members of society), put under nonidealized circumstances C , will actually implement the principles.

As Gaus notes, in Rawls's contractualism, the "members of the justificatory public" belonging to set N^* are idealized in several ways.³ In particular, the members of N^* are assumed to be good-willed persons. In Rawls's words, they are *reasonable* persons moved by a sense of justice, respecting others as free and equal persons. They are, moreover, appropriately motivated: considerations foreign to justice are temporarily bracketed in their practical deliberations. This idealization of the members of N^* is of course closely related to the particular circumstances under which they are put. The original position, viewed as a "procedure of construction," is constructed in such a way that idealized individuals cannot but rationally ignore considerations that are not related to justice.⁴ Rawls's theory of justice thus faces the stability problem outlined above. Once idealized members of the justificatory public have agreed on a set of principles of justice regulating the basic structure of the society, it has still to be established that members of the society will respect the institutions that implement the principles.

Commentators of Rawls's scholarship have recently helped to clarify the nature of the stability problem in Rawlsian political philosophy.⁵ Two threats for stability can be identified once the veil of ignorance characterizing the Rawlsian original position is lifted. The first threat refers to the fact that once the veil is lifted, each person will pursue her plans as conceived by her conception of the good life. While members of N^* only take into account considerations of justice based on a "thin" theory of the good, actual members of the society use their full deliberative rationality to pursue their unrestricted conception of the good. This is the problem of "justificatory instability."⁶ Weithman, following Rawls, characterizes this problem as an n -person prisoner's dilemma.⁷ Each member of N may be tempted to free ride, meaning that one's conception of the good may rationally encourage one to behave wrongly (i.e., not to follow the justice principles agreed on), no matter how the other members are behaving. Suppose that the problem is solved along the lines Rawls suggests: members of N have a "sense of justice" that "leads [them] to promote just schemes and to do [their]

3 Gaus, "A Tale of Two Sets," 305–6.

4 Rawls, "Kantian Constructivism in Moral Theory."

5 Gaus, "A Tale of Two Sets"; Weithman, *Why Political Liberalism?*

6 Gaus, "A Tale of Two Sets," 307–15.

7 Weithman, *Why Political Liberalism?* 48; Rawls, *A Theory of Justice*, 505.

share in them when [they] believe that others, or sufficiently many of them, will do theirs.”⁸ Here comes the second threat: persons who have a sense of justice sufficiently strong to overcome the temptation to free ride will act rightly if and only if they expect others to do so. This is a mutual assurance problem that generalizes to n players the following two-person assurance game (fig. 1).

		<i>Person 2</i>	
		Act rightly	Act wrongly
<i>Person 1</i>	Act rightly	3; 3	0; 2
	Act wrongly	2; 0	1; 1

FIGURE 1 A Two-Person Assurance Game

In such an assurance game, acting rightly is rational as long as person 1 believes that person 2 is sufficiently likely to act rightly. Interactive reasoning further implies that person 1 will have this belief only if she believes that person 2 believes that she is sufficiently likely to act rightly, and so on *ad infinitum*. Solving the assurance problem requires finding the appropriate ground for a full (and thus infinite) hierarchy of beliefs based on which acting rightly is rational.⁹

The stability problem is the topic of part 3 of *A Theory of Justice*. To understand Rawls's proposed solution to it, it is important to acknowledge two requirements that Rawls imposes on any plausible solution. These requirements indicate that Rawls is concerned with a specific kind of stability. First, Rawls's characterization of a well-ordered society points out that any kind of stable state will not do. Only stable *just* states are acceptable—that is, states in which people behave following the principles of justice. Second, principles of justice should be *inherently stable* rather than stable by imposition: “A conception of justice is inherently stable if a society that is well-ordered by it generally maintains itself in a just general equilibrium and is capable of righting itself when that

8 Rawls, *A Theory of Justice*, 236.

9 Generalizing the assurance game to n persons leads to two significant differences. First, each player has to form a belief on the likely number of persons who will act rightly. The hierarchy of beliefs is then defined on each person's belief about the number of persons who will act rightly. Second, we may reasonably imagine that different persons have different belief thresholds above which acting rightly is rational. Persons with a more developed sense of justice will act rightly even if they expect a relatively low number of other persons to do so. This makes the dynamic more complicated, but in the end, such an n -person assurance game would still have two stable equilibria: one where a low number of persons act rightly and one where a high number of persons act rightly. See Granovetter (“Threshold Models of Collective Behavior”) for classical dynamic threshold models of this kind.

equilibrium is disturbed.”¹⁰ Rawls’s proposed solution to the stability problem in *A Theory of Justice* proceeds through an argument for the *congruence of the right and the good*.¹¹ In a nutshell, Rawls contends that the institutions of the just society will solve the stability problem if and only if they are able to elicit in the members of *N* a sense of justice that they have sufficiently strong reasons to keep such that acting wrongly would not be rational. Because members of *N* are assumed to rationally pursue what they consider to be good, institutions will foster inherent stability if and only if the sense of justice is part of the conceptions of the good of members of *N*. Rawls claims that this necessary and sufficient condition is fulfilled as soon as we assume that members of *N* want to live up to a small number of ideals (friendship, personal conduct, association) that are constitutive of justice as fairness. The congruence between the right and the good solves both the justificatory instability and the assurance problems at the same time. It establishes that members of *N* have all-things-considered reasons to live up to their sense of justice and thus to abide by the principles agreed on by their ideal counterparts, the members of *N**.¹²

2. RAWLS’S MODEL OF PUBLIC REASON

Specialists of Rawls’s scholarship disagree over the reasons for and the meaning of Rawls’s political turn. This political turn is fully stated in *Political Liberalism* but was engaged by Rawls in the 1980s, in particular in important essays such as “Justice as Fairness: Political not Metaphysical” and “The Domain of the Political and Overlapping Consensus.” Justice as fairness and more generally all plausible liberal principles of justice are reframed as *political* principles. Their reasonability as principles of justice is presented as being independent of metaphysical truths and more generally of the truth value of any proposition belonging to areas of philosophy other than political philosophy. In particular, Rawls’s political liberalism is sometimes read as an attempt to reformulate justice as fairness by substituting a *political* conception of persons as free and equal citizens for the metaphysical conception on which *A Theory of Justice* is thought

10 Weithman, *Why Political Liberalism?* 45.

11 See Weithman, *Why Political Liberalism?* esp. chs. 2–4, for a detailed discussion of the congruence argument and its problems.

12 In fact, the congruence argument is stronger than that. Rawls argues that persons in the original position would not agree on principles that they do not expect to be stable in the relevant sense. That means that the congruence argument states that the sense of justice and the related ideals are actually part of the thin theory of the good that is shared by all members of *N**.

to build. As Weithman forcefully argues, this reading is at best misleading.¹³ First, it is not clear that the argument for justice as fairness in *A Theory of Justice* depends on a metaphysical conception of the person. Second, and more importantly, it misses the point (stated explicitly by Rawls in several places) that the main problem is the treatment of the stability problem in *A Theory of Justice*.

Rawls became skeptical about the congruence argument as an account of the ability of just political institutions to foster inherent stability because it depends on assumptions that contradict the very nature of a liberal society. Liberal political institutions favor the development of a great diversity of views and opinions. These views and opinions are the product of partial or full “comprehensive doctrines” held by members of *N*. Comprehensive doctrines are constituted by values and trade-offs between these values that express conceptions of the good life. The “burdens of judgments” correspond to the many causes Rawls identifies that explain why, in a liberal society, we cannot avoid a reasonable disagreement over conceptions of the good life. The problem with the congruence argument is that it (apparently) solves the stability problem but at a considerable cost: it disregards the burdens of judgment by assuming that members of *N* all endorse ideals that belong to a specific comprehensive doctrine, justice as fairness. Thus, it fails to establish what it needs to establish: that just and liberal political institutions would generate their own stability. In this sense, justice as fairness needs to be reformulated as a political rather than comprehensive doctrine. The reason, however, is not related to the conception of the person *per se*; rather, this account of justice has to be vindicated without assuming that members of *N* endorse any specific conception of the good. A new solution to the stability problem thus has to be found.

What has to be demonstrated is that in a well-ordered society, members of *N* have reasons to abide by political principles of justice for shared reasons that do not depend on their comprehensive doctrines. If these reasons are sufficiently strong, they will solve the justificatory instability problem. To solve the assurance problem, however, they also have to be *public*. Public reason makes its appearance in Rawls’s political account as a coordination device that helps members of *N* obtain the required assurance that everyone will act rightly because everyone expects everyone else to act rightly, all this being commonly known among the members of *N*. To understand this, let me briefly expose a stylized version of what I call *Rawls’s Model of Public Reason* (henceforth, RMPR).¹⁴

13 Weithman, *Why Political Liberalism?*

14 I call it a stylized version because I ignore several details Rawls discusses at length in “The Idea of Public Reason Revisited.” Also, as Gaus and Van Shoelandt demonstrate, Rawls’s political liberalism has the structure of a baroque cathedral that can be represented by several competing and not entirely consistent models (“Consensus on What?”). My

The RMPR can be seen as an attempt to solve the justificatory and stability problems at the same time. In this model, justification comes first through the agreement over political principles of justice by members of N^* put under circumstances C^* . The original position is a procedure of construction that depends on a political conception of persons as free and equal citizens. Individuals who endorse this conception recognize themselves as endowed with two moral powers: first, they have a capacity for a sense of justice; second, they have a capacity to rationally pursue a conception of the good.¹⁵ This directly leads to the construction of the original position, where members of N^* choose principles of justice in ignorance of their personal positions and characteristics while endorsing the thin theory of the good. Letting RN be the set of all-things-considered reasons that can justify the endorsement of principles of justice in N , the permissible reasons are restricted to a subset $R^* \subset RN$. If P is the set of all possible conceptions of justice, the reasons R^* available to the members of N^* put in circumstances C^* make them select a subset P^* of *liberal conceptions of justice*. These conceptions share a small set of features.¹⁶ This first justificatory stage is already public in the sense that the set R^* is public because it follows implicitly from the public recognition that persons are free and equal citizens. In the contrary case, the original position could not serve as a device of justification among the members of N . The justification of P^* is only a *pro tanto* justification, however.

To achieve *full justification*, we need to consider whether components of P^* are still justified when transitioning from the triple $\{N^*, C^*, R^*\}$ to the triple $\{N, C, RN\}$.¹⁷ The issue is whether each citizen can individually and rationally endorse the conceptions contained in P^* in light of their comprehensive doctrine, especially their conception of the good. In the most permissive, “wide” version of public reason Rawls exposes, comprehensive doctrines can reasonably be introduced at the stage of full justification, provided that public reasons are presented in due course in case of conflict between comprehensive doctrines. More formally, full justification requires that $\{N^*, C^*, R^*\} \Rightarrow P^*$,

characterization of the RMPR is consistent with the two-part version of their model III. Their one-part version of model III and their model IV stretch Rawls’s political liberalism toward the diversity-convergence account of stability, probably too much.

15 Rawls, *Political Liberalism*, 18.

16 Rawls, “The Idea of Public Reason Revisited,” 773. Crucially, though Rawls obviously sees justice as fairness as belonging to P^* , he leaves open the possibility that P^* is not a singleton.

17 “Full justification is carried out by an individual citizen as a member of civil society. . . . In this case, the citizen accepts a political conception and fills out its justification by embedding it in some way into the citizen’s comprehensive doctrine as either true or reasonable, depending on what that doctrine allows” (Rawls, *Political Liberalism*, 386).

then $\{N, C, RN\} \Rightarrow P^*$, with the proviso that $\{N, C, R^*\} \Rightarrow P^*$. If this condition is satisfied, an overlapping consensus is established: the liberal conceptions of justice contained in P^* are all endorsable from within the citizens' comprehensive doctrines. That means that members of N 's all-things-considered reasons RN sufficiently overlap to provide shared support for these conceptions. This solves the first part of the stability problem, the justificatory instability problem.

A last step is required to solve the second part of the stability problem, the assurance problem. Recall here that the problem is not whether persons have reasons to act rightly but whether they can have the assurance that others will do so. As I explain above, the assurance requirement trickles up to the whole hierarchy of belief: Ann must have sufficient reason to believe that Bob will act rightly, which requires that she have sufficient reason to believe that Bob has sufficient to reason to believe that Ann will act rightly, and so on. That implies that the assurance problem cannot be solved unless the existence of an overlapping consensus is itself a *public event* in N . By definition, a public event is commonly known.¹⁸ In this case, it is commonly known among members of N that conceptions in P^* are fully justified for reasons R^* that are at the same time shared and endorsable from within the various comprehensive doctrines. The assurance problem is solved because the public establishment of the overlapping consensus (which solves the justificatory instability problem) makes it common knowledge that liberal conceptions of justice are endorsed for shared public reasons. Stability for the right reasons is achieved because officials and citizens all justifiably believe that everyone has shared sufficient reasons to act rightly.

3. THE DIVERSITY-CONVERGENCE CRITIQUE OF THE RMPR

The literature on public reason is huge, and the Rawlsian account of public reason has been the target of a long list of criticisms.¹⁹ I shall not discuss all of them here. Rather, I will focus on the criticisms that explicitly and specifically challenge the ability of the RMPR to solve the stability problem. These criticisms share at least two features. First, they mostly ignore the justificatory instability problem—implicitly indicating that the RMPR provides a satisfactory answer to it. Instead, they argue that public reason cannot be the assurance device that the RMPR assumes it is. Second, they tend to favor a convergence account

18 For an insightful analysis of the role of public events in fostering common knowledge in a variety of social situations, see Chwe, *Rational Ritual*. For a formal account, see Milgrom, "An Axiomatic Characterization of Common Knowledge."

19 See Quong, *Liberalism without Perfection*, 259–60, for a survey of this list.

of public reason, instead of Rawls's consensus approach.²⁰ Proponents of this account tend to argue that Rawls's consensus approach depends on unrealistic or normatively doubtful assumptions to solve the stability problem in highly diverse societies. I present and explore in this section the various arguments of the diversity-convergence critique against the RMPR.

3.1. *Public Reason Is Cheap Talk*

A first argument against the Rawlsian solution to the stability problem is that public reasons cannot provide the required assurance because they are merely *cheap talk*. Consider a specific relation between two public officials (e.g., a judge and a legislator), between an official and a citizen, or between a set of officials and a set of citizens.²¹ The RMPR suggests that the use of public reasons in debates over fundamental political questions will assure others that one intends to act rightly. It would signal one's intention to sincerely abide by the liberal principles of justice. As several critics note, the use of public reasons cannot, however, serve as a proper signal in the context of an assurance game as depicted in figure 1.²² The reason is that each person has an interest in making the other believe that she will act rightly, independently of her actual intention. Indeed, if person 2 believes that person 1 will act rightly, then person 2 has (assuming that the justificatory instability problem has been solved) all-things-considered reasons to act rightly too. Crucially, this guarantees to person 1 either of her two most preferred outcomes. Moreover, using public reasons in this model is costless. That means that one has absolutely no reason not to *pretend* to act rightly by using public reasons. This information is common knowledge among the players. That means that the use of public reasons does not lead to trustworthy messages and cannot solve the assurance problem.

3.2. *Public Reasoning Is Vulnerable to Noise and Its Amplification*

Thrasher and Vallier highlight a second problem with the RMPR related to its vulnerability to noise and its amplification.²³ Experimental evidence indicates that cheap talk can sometimes slightly favor cooperation in social dilemmas.

20 For the consensus/convergence distinction, see D'Agostino, *Free Public Reason*.

21 Rawls indicates that the ideal of public reason applies to government officials and candidates for public office ("The Idea of Public Reason Revisited," 766–67). Relations between citizens *per se* are outside the scope of public reason. See, however, Quong, *Liberalism without Perfection*, 273–75, for an argument in favor of a broader scope of public reason.

22 Gaus, "A Tale of Two Sets"; Kogelmann and Stich, "When Public Reason Fails Us"; Thrasher and Vallier, "The Fragility of Consensus."

23 Thrasher and Vallier, "The Fragility of Consensus."

This is especially the case in face-to-face settings.²⁴ However, most of the interactions in the well-ordered society would take place under more impersonal settings. In this case, even if a global norm of mutual assurance prevails, the equilibrium (and thus the consensus) is susceptible to being destabilized by occasional—and possibly involuntary—defections. This can be the case because random errors can be interpreted as an intentional unwillingness to act rightly, leading to further defections. The result is an informational cascade leading to a sudden switch from a situation where almost everyone acts rightly to a situation where everyone disregards the principles of justice.²⁵ This problem is even more acute under the wide view of public reason. The wide view indeed tolerates the introduction in the political forum of nonpublic reasons attached to comprehensive doctrines. This makes it even harder to discriminate between insincere uses and sincere uses of private reasons that one intends to back up with public reasons.

3.3. *Public Reason Is Incomplete and Manipulable*

A third criticism underlines the incompleteness of public reason and its related manipulability.²⁶ Incompleteness can have two origins. On the one hand, public reason may be *indeterminate* because it cannot provide an answer to some fundamental political questions—for example, What should the monetary policy of the European Central Bank be? The answer to this question is unlikely to be found within the realm of public reason, at least if one accepts Rawls's requirement that it not appeal to controversial arguments not accessible to all citizens.²⁷ That seems to imply that citizens and officials must appeal to their comprehensive doctrines and nonpublic reasons to adjudicate this kind of issue. In these circumstances, public reason can no longer be used as an assurance mechanism. On the other hand, incompleteness can result from *inconclusiveness*. In this case, public reason provides contradictory and non-dominated justifications on political issues. The worry is not that citizens and officials may appeal to nonpublic reasons but rather that they might choose public reasons that are the most favorable to their private interests. Again, this

24 Wilson and Sell, "Liar, Liar . . ."

25 Game-theoretic evolutionary models moreover show that in games with similar structures to the assurance game, only the suboptimal equilibrium is stochastically stable. See, for instance, Young, *Individual Strategy and Social Structure*.

26 Kogelmann, "Public Reason's Chaos Theorem."

27 Note, however, that Rawls is explicitly requiring the completeness of the *political conceptions* underlying public reason ("The Idea of Public Reason Revisited," 777). But presumably a political conception does not fully cover topics such as monetary policy, though monetary policy has an obvious relevance to distributive justice.

presumably considerably weakens the ability of public reason to serve as an assurance mechanism. Inconclusiveness indeed makes public reason not strategy-proof, and thus vulnerable to manipulation, along lines similar to standard results about strategy-proofness in social choice theory.²⁸

3.4. *Public Reason Depends on a Common Knowledge Condition*

A fourth and last criticism emphasizes the problems related to the fact that common knowledge is a precondition for Rawlsian stability for the right reason.²⁹ As I explain above, the transition from full justification to public justification implies that the existence of the overlapping consensus is common knowledge among the members of *N*. At least, this is required if public reason is to serve as an assurance mechanism. The assurance has to be given across the whole belief hierarchy of each official and citizen.³⁰ This requirement, however, makes public reason very fragile as an assurance mechanism as soon as there is a small fraction of unreasonable persons in *N*. The presence of a fraction ϵ of unreasonable persons in *N* means that reasonable persons are confronted with a Bayesian game: with a probability of $1 - \epsilon$ they play the assurance game of figure 1, but with probability ϵ they are part of an assurance dilemma with an unreasonable person (fig. 2). Acting wrongly is the dominant strategy for the unreasonable person. It is then best for the reasonable person to also act wrongly.

		<i>Person 2 (unreasonable)</i>	
		Act rightly	Act wrongly
<i>Person 1 (reasonable)</i>	Act rightly	3; 2	0; 3
	Act wrongly	2; 0	1; 1

FIGURE 2 An Assurance Dilemma

We might assume that knowing the other’s type is sufficient for a reasonable person to choose to act rightly (if the other’s type is reasonable). But this is not the case: the reasonable person has to know that the other person knows that she is reasonable, and so on. Ultimately, public reason can serve as an assurance mechanism only if it is common knowledge that both persons are reasonable.

28 Taylor, *Social Choice and the Mathematics of Manipulation*.

29 Chung, “The Instability of John Rawls’s ‘Stability for the Right Reasons.’”

30 As Chung notes, this is never made explicit in Rawls’s writings (“The Instability of John Rawls’s ‘Stability for the Right Reasons’”). Readers of Rawls who endorse his public reason model recognize, however, the common knowledge requirement. See Hadfield and Macedo, “Rational Reasonableness”; and Weithman, *Why Political Liberalism?*

As Chung formally demonstrates, whenever there is a fraction ε arbitrarily close to zero of unreasonable persons, there is always a Bayesian game where both persons acting wrongly is the unique equilibrium.³¹

4. THE RMPR AND COMMUNITY-BASED REASONING

We may ask whether the RMPR can be adapted to respond to the objections that have been leveled against it by the diversity-convergence account. The questions at stake are the following: Under what condition(s) can public reason serve as an assurance mechanism? Are these conditions likely to apply in diverse liberal societies?

Based on a coordination model that Hadfield and Weingast developed, Hadfield and Macedo argue that public reason can solve the assurance problem by fostering a common logic among the members of the relevant population.³² This common logic provides the basis for believing that everyone will act rightly. In Hadfield and Weingast's model, a third-party institution supplies the common logic in a population of buyers and sellers who want to trade for goods. The third-party institution provides a classification of agents' behavior, in particular whether the action of an agent can be classified as "cheating." Based on this classification, the institution supports a system of belief such that agents are credibly threatened with being punished if they cheat. On this basis, buyers are able to coordinate on boycotting sellers who would cheat, thus deterring actual cheating behavior from the sellers. According to Hadfield and Macedo, this model "helps us understand [that] the importance of public reason is the recognition that in order to coordinate the participants in a community dependent on a decentralized enforcement of rules and principles expressing judgments of right and wrong, the common logic must be *publicly accessible*. Indeed, in game-theoretic terms, it must be *common knowledge*."³³ Hadfield and Macedo go on to emphasize that endorsing the publicly accessible common logic could ultimately lead persons to set aside their personal reasons and inference norms, and accept a form of normativity derived from a "we-mode" of thinking.³⁴

So far, so good. Hadfield and Macedo's account appropriately captures the common knowledge requirement that I emphasize above. However, they do

31 Chung, "The Instability of John Rawls's 'Stability for the Right Reasons.'" This result can be seen as a variant of Rubinstein's "electronic mail game" ("The Electronic Mail Game"). This kind of model shows that common knowledge and approximation of common knowledge have very different implications.

32 Hadfield and Weingast, "What Is Law?"; Hadfield and Macedo, "Rational Reasonableness."

33 Hadfield and Macedo, "Rational Reasonableness," 12–13, emphasis in original.

34 Hadfield and Macedo, "Rational Reasonableness," 39.

not make it clear how public reason makes the common logic publicly available. Hadfield and Weingast's model *postulates* the existence of a third-party institution that makes public announcements. It does not give any indication regarding the circumstances under which this institution emerges and is able to achieve the required publicness. I shall now propose an extension of the RMPR that accounts for a plausible form of practical reasoning through which public reason generates common knowledge. Following my earlier work, I call this form of practical reasoning *community-based reasoning*.³⁵ It can be characterized in the following way:

Community-Based Reasoning: A person i 's practical reasoning is community-based if, in some strategic interaction S , her action A follows from the following reasoning steps:

1. i believes that [she] and all other persons j in S are members of some community C .
2. i believes that some state of affairs x holds.
3. Given 2, 1 grounds i 's belief that all other persons j in S believe that x holds.
4. From x , i inductively infers that some state of affairs y also holds.
5. Given 3 and 4, 1 grounds i 's belief that all other persons j in S believe that y also holds.
6. From 5, and given i 's preferences, i concludes that A is best in S .³⁶

Suitably reformulated in the terms of the RMPR, we can say that members of N are community-based reasoners if they infer that they have all-things-considered reasons to act rightly partly from the fact that they belong to the same *political community*. Community membership indeed sustains two key steps in this logic of reasoning. In step 3, community membership serves as a basis for i to infer that x is mutual knowledge. In step 5, it serves as a basis for i to infer that everyone in N infers y from x . The term "grounds" in both steps singles out the fact that i assumes, based on community membership, that she and other members of N are sharing both privileged epistemic accessibility to some state of affairs x and some form of inductive inference from x to y .

The key point here is that if all members of N are community-based reasoners with respect to some state of affairs x , then it can be shown that y is common

35 Hédoïn, "A Framework for Community-Based Salience" and "Community-Based Reasoning in Games." I leave most of the technical details aside. The interested reader may consult "Community-Based Reasoning in Games," in which an epistemic game-theoretic model is presented.

36 Hédoïn, "Community-Based Reasoning in Games," 4.

belief (or knowledge) in N .³⁷ Substitute “everyone else acts rightly” for y ; x refers to a mutually accessible state of affairs or event—for example, an official’s making an announcement in an assembly. Now, upon observing x , any member of N will first infer that everyone else, as members of the same community, has also observed x . She will also infer from x that y is the case, and that everyone else makes the same inference from x to y . Then, she believes that y is mutual belief in N . Therefore, one can infer from x that y is mutual belief in N , which we denote by y' . Substitute y' for y in step 4 above. From 5, one can infer that y' is mutual belief in N , which we denote y'' . We can reiterate the process indefinitely, thus establishing that y is common knowledge in N . Assuming that the members of N are community-based reasoners then solves the intricated issue of the origins of common knowledge. It thus directly answers the objection discussed in section 3.4 above.

Provided that the justificatory instability problem has already been resolved, community-based reasoning also responds to the cheap talk objection. Once person i has reached the conclusion that it is common knowledge that everyone will act rightly, the preference structure of the assurance game makes acting rightly the strict best response (see fig. 1). Suppose that x denotes some announcement, based on one or several public reasons, made by a public official in an assembly. Two points are relevant here. First, we implicitly have to assume that everyone agrees on what constitutes “public reasons.” This is what lurks behind step 3. This assumption is directly grounded on the fact that because members of N are members of the same community, we consider that *by definition* they share a common conception of what public reasons are. Second, once again by the very fact that they are members of the same community, we assume that members of N inductively infer from x that everyone else will act rightly. Strictly speaking, this inference may be false. It could happen that some persons in N do not act justly even if they observe x . But in this very community, that will not happen. Members of N just share a *lebensform*, a “form of life” in the Wittgensteinian sense. This is sufficient to provide the required assurance. In this community, using public reasons is not cheap talk, as a matter of fact. Quite the contrary, the use of public reasons is a *community-based salient event* in the sense I offer elsewhere: an event based on which everyone correctly infers that some proposition is commonly known among a community.³⁸ This very

37 See Hédoin, “Community-Based Reasoning in Games,” 7–8, for a formal proof. Lewis is the first to have proposed an account that shows how a proposition *becomes* common knowledge in a population of agents who share a form of inductive inference and have mutual epistemic access to a given state of affairs (*Convention*).

38 Hédoin, “Community-Based Reasoning in Games,” 10.

property singles out this event, separating it from the stream of events that any normally rational person can observe.

Community-based reasoning thus provides a sufficient condition for public reason to satisfy the required common knowledge condition. This account adequately supplements Hadfield and Macedo's claim that public reason makes common knowledge a common logic in a population.³⁹ Two remarks should be made in addition. The first remark is related to a hidden assumption behind the above argument: here, I have implicitly assumed not only that everyone in N is a community-based reasoner but also that it is somehow common knowledge that this is the case. This is indeed required to make the iterative steps that successively show that y, y', y'' , and so on, are mutual knowledge. Where does this common knowledge come from? A plausible answer, at least in some circumstances, is just to point out that community membership is a public event *and* that to be a member of a community is to share some form of reasoning, including inductive inference, through common practices. In this case, community-based reasoning can indeed become common knowledge in the relevant population.⁴⁰ This leads to an ostensible Wittgensteinian reinterpretation of the concept of public reason.⁴¹ To make use of public reasons is to follow some rules that, at the bottom, correspond to shared practices that do not have any fundamental justification. Public reason is just what members of a community agree it is through their practices, and to share a conception of public reason and the related practices is just what it is to be a member of some community. This emphasizes an important link between justification and stability. To be fully justified, principles of justice P^* must be agreed on by members of the relevant population, based on public reasons R^* . One is then a member of the community by virtue of agreeing on P^* based on R^* . Agreeing on the justification of P^* based on R^* defines at the same time what it is to be a member of the community. In this Wittgensteinian sense, if principles P^* are properly justified, they must already be common knowledge. The step from full to public justification is merely theoretical; in practice, they are one and the same thing.

39 Hadfield and Macedo, "Rational Reasonableness."

40 See note 18 above.

41 See especially Wittgenstein's remarks on "language games" and "forms of life," pointing out the relationship between meaning and the existence of shared practices among the members of a community (*Philosophical Investigations*). As Forrester has documented, Rawls's early social philosophy has been largely influenced by Wittgenstein's conception of rules and games (*In the Shadow of Justice*). It should be acknowledged, however, that this reinterpretation is only partial. A full Wittgensteinian account of public reason would assert that *all* reasons are public.

The second point is more exegetical. Rawls explicitly states that a “well-ordered democratic society is *neither a community* nor, more generally, an association.”⁴² Rawls notices two differences between a well-ordered society and an association. On the one hand, a well-ordered society is closed in the sense that one cannot enter or exit it in the course of her life. On the other hand, a well-ordered society does not have final ends or aims. Moreover, a well-ordered society “is not a community either, if we mean by a community a society governed by a shared comprehensive religious, philosophical, or moral doctrine.” He adds that this “fact is crucial for a well-ordered society’s idea of public reason. To think of a democracy as a community (so defined) overlooks the limited scope of its public reason founded on a political conception of justice.”⁴³ These remarks can be related to the distinction between the political and the comprehensive that is at the core of political liberalism, and to the exclusion of comprehensive reasons from the realm of public reason. The point is that to solve the stability problem, there must be *something* common and public in the relevant population. Rawls may have been wrong in excluding truth, perfectionist values, and more generally comprehensive reasons from the realm of public reason. On the other hand, from a Rawlsian perspective, the burdens of judgment make unlikely the possibility that reasons based on metaphysical, religious, or moral considerations can serve as a basis for public justification. The difficulty, pointed out multiple times in the literature, is that it is unclear why the disagreement between persons stemming from the burdens of judgment should not concern political principles of justice. Rawls’s answer to this objection is to argue that the idea of public reason is constitutive of a constitutional democratic society governed according to liberal principles of justice. The idea of public reason has then presumably no bearing in societies that are not *de facto* liberal and democratic in this sense. If we follow Rawls in this postulate, we may then argue that the members of *N* belong to a *democratic political* community. This political community is defined by the fact that its members accept a specific political conception of justice, along the lines Rawls identified.⁴⁴

5. SOME CONCERNS AND OBJECTIONS

I shall address in this section some concerns and objections to my proposed interpretation of the RMPR in terms of community-based reasoning. This will also permit making more explicit its implications and filling in some detail.

42 Rawls, *Political Liberalism*, 40, emphasis added.

43 Rawls, *Political Liberalism*, 41.

44 Rawls uses the term “political community” in *Political Liberalism*, but not in this sense.

There are, in particular, two related worries: that reinterpreting the RMPR in terms of community-based reasoning makes too strong a concession to communitarianism and that it depends on a too-demanding form of normalization.⁴⁵ These worries find their roots in the above characterization of community-based reasoning. In particular, it might be suggested that steps 3 and 5 make strong demands on the form of practical reasoning required and that, because of that, it will either fail to solve the stability problem most of the time or do so in a manner that betrays the very point of Rawls's political liberalism. Another way to state this concern is to ask what remains of the idea of public reason as a distinctive account to solve the stability problem under the community-based approach. I shall answer in two steps.

There is no doubt that the proposed revision of the RMPR goes in the direction of communitarianism. This is not only because the account of community-based reasoning obviously relies on the concept of community. This is more fundamentally related to the fact that the idea of public reason, at least in Rawls's political liberalism, is tightly related to legitimacy and justification.⁴⁶ I have acknowledged above that under this revised interpretation of the RMPR, using public reason consists in following rules that do not have a fundamental justification, except that they are constitutive of shared social practices. Now, if the community-based reasoning account indeed entails deflationary concepts of justification and legitimacy, this is not *totally* inconsistent with Rawls's political turn. As I have noted in the preceding section, that turn itself strongly suggests that the idea of public reason is grounded in what can be called the *democratic form of life*, consisting of political and social practices that are themselves a reflection of the *democratic political culture*. In other words, there is already in Rawls's late solution to the stability problem the admission that public reason, and more generally the whole political conception, is related to some way of life that is specific to liberal democratic societies. My argument for the community-based reasoning reinterpretation of the RMPR can then be formulated as follows: as a device that makes it possible for a society to foster stability around a conception of justice, public reason is specific to the democratic form of life of liberal democracies but still has to rely on a form of practical reasoning that is community based—a form of practical reasoning that is not specific to any kind of society but that we may well see at work in any human community. This form of practical reasoning makes it possible to satisfy the requirement of common knowledge that, as we have seen, is needed for public reason to fulfill its job.

45 I thank the two anonymous reviewers who have each and in different ways pushed me to be more explicit about these worries and how they can be addressed.

46 Rawls's "liberal principle of legitimacy" is indeed stated and defended in the context of his account of public reason (*Political Liberalism*, 216).

This leads to the second worry. Is it plausible to think that this revised version of the RMPR can account for the way public reason solves the stability problem—if indeed it does—in modern liberal democracies? I should first mention that community-based reasoning is only a partial solution to the stability problem. It depends on the fact that persons indeed are community-based reasoners. The plausibility of this assumption depends not only on empirical considerations but also on what we regard as the appropriate analytical understanding of the concept of community. The risk is to navigate between the tautology (and thus empirical irrelevance) and the empirically refuted. Regarding the latter, it may indeed happen to be that contemporary societies are too diverse and, as a consequence, that community-based reasoning is not available as a form of practical reasoning. It can be argued in particular that it is unlikely that the members of a nation form a community in the relevant sense. This is of course a relevant and fully justified worry. There is no doubt that in some polities the lack of shared national identity can be a major obstacle to the emergence of a stable overlapping consensus around the constitutional essentials and matters of basic justice that are the subject of Rawlsian public reason.⁴⁷ Even in more culturally homogeneous polities, it is unclear that practical reasoning can be community based. We should, however, remember that, at least in Rawls's version, the idea of public reason is limited in scope and content. It concerns only the "political" aspect of constitutional essentials and matters of basic justice, and it applies only to discussions within the "public political forum": the discourse of judges in their decisions, the discourse of government officials, and the discourse of candidates for public office.⁴⁸ Within this restricted scope, the assumption that persons—for instance, members of parliament—are at least sometimes community-based reasoners seems to be less demanding.⁴⁹

Even if this answers the two major worries I have identified, it should be acknowledged again that community-based reasoning remains only a partial solution to the stability problem. In the preceding section, I argue that community-based reasoning allows the RMPR to solve two objections: that public

47 In a worse case, this lack can escalate into a civil war. But less dramatic examples, such as the case of contemporary Belgium, illustrate the kind of difficulties that may arise and how they can be interpreted from the Rawlsian perspective.

48 Rawls, "The Idea of Public Reason Revisited," 767.

49 Rawls notes that citizens, when they vote on constitutional essentials and matters of basic justice, "are to think of themselves *as if* they were legislators" and so must use the requirements of public reason to assess government officials and candidates for public office ("The Idea of Public Reason Revisited," 769). We may doubt that this corresponds to actual practice in liberal democracies. Indeed, this demand is more akin to what Rawls calls the *ideal* than the idea of public reason.

reason is cheap talk and that it depends on a common knowledge condition. I have said nothing, however, about the other two objections: that public reason is vulnerable to noise and to manipulation. Moreover, as Chung shows, satisfying the common knowledge condition is far less easy if the justificatory instability problem has not yet been fully solved.⁵⁰

This points to a simple but essential fact: stability with too much diversity is compromised. In other words, there must be something common among the members of a society to obtain the required stability—that is, not a mere *modus vivendi*. Ideally, this common minimal basis must be furnished by the political institutions themselves—through education, for instance—as Rawls himself underlined. More generally, by serving the role of correlating devices, social norms governing relationships between public officials, and between public officials and citizens, provide the basis on which public reason operates in a well-ordered society. Particular norms or rules do not select directly a particular political conception. But as consistent systems of norms and rules, institutions create *indirect public reasons* to settle on a particular conception in specific circumstances.⁵¹ But the existence of these norms is itself constitutive of forms of life—that is shared modes of reasoning generating focal points and salient events. As Thrasher and Vallier explicitly admit, their model “assumes that the relevant type of stability [i.e., stability in the Rawlsian sense] is already in place.”⁵² This is a severe limit because as a result it hardly improves on the RMPR.⁵³ Arguably, the more a society is diverse in its practices and beliefs, the smaller the scope of public reason will be. At the extreme, public reason must rely on the lowest common denominator that makes people belong to the same society.

6. CONCLUSION

Proponents of the diversity-convergence approach do not stop at criticizing Rawls’s model of public reason. They have also argued that a convergence account of public reason where persons agree over rules and principles for

50 Chung, “The Instability of John Rawls’s ‘Stability for the Right Reasons.’”

51 Thrasher and Vallier, “The Fragility of Consensus.”

52 Thrasher and Vallier, “The Fragility of Consensus,” 950.

53 As they seem to implicitly admit, Thrasher and Vallier’s model is not incompatible with the RMPR (“The Fragility of Consensus”). There is no obvious reason why the latter would not admit the role played by “choreographers” in fostering social coordination. Thrasher and Vallier underestimate the importance of shared reasoning for the existence of correlated equilibria, which is formally captured by the so-called common prior assumption (see Gintis, *The Bounds of Reason*, ch. 7).

different reasons solves the stability problem.⁵⁴ It is beyond the scope of this paper to assess the merits of the convergence account in comparison with those of the RMPR with respect to the stability problem. To end this essay, however, two brief remarks are in order. First, what I have just said above suggests that a pure convergence account is unlikely to succeed, at least if more than a mere *modus vivendi* is aimed at. As Chung points out using a formal argument, to the extent that diversity promotes extremist views, it is not the case that diversity will strengthen stability, contrary to what, for instance, Kogelmann and Stich argue.⁵⁵ Second, it is not clear what remains “public” in an account of public reason where persons are permitted to bring any beliefs and reasons in support of a law. This is not a mere terminological quibble; it also questions the very nature of the social order in liberal democracies. Overall, this is the whole project of public reason liberalism that must be reconsidered in light of the difficulty in solving the stability problem.⁵⁶

University of Reims Champagne-Ardenne
 cyril.hedoin@univ-reims.fr

REFERENCES

- Binmore, K. G. *Just Playing*. Vol. 2 of *Game Theory and the Social Contract*. Cambridge, MA: MIT Press, 1998.
- Broome, John. *Weighing Goods: Equality, Uncertainty and Time*. Oxford: Blackwell, 1991.
- Chung, Hun. “The Instability of John Rawls’s ‘Stability for the Right Reasons.’” *Episteme* 16, no. 1 (March 2019): 1–17.
- . “The Well-Ordered Society under Crisis: A Formal Analysis of Public Reason vs. Convergence Discourse.” *American Journal of Political Science* 64, no. 1 (January 2020): 82–101.
- Chwe, Michael Suk-Young. *Rational Ritual: Culture, Coordination, and Common Knowledge*. Princeton, NJ: Princeton University Press, 2003.

54 Gaus, *The Order of Public Reason*; Kogelmann and Stich, “When Public Reason Fails Us.”

55 Chung, “The Well-Ordered Society under Crisis”; Kogelmann and Stich, “When Public Reason Fails Us.”

56 This essay was presented at the STOREP conference in June 2021 (held online) and at the international conference “Héritages et usages de Rawls: *Théorie de la justice*, 50 ans après,” held in November 2021 in Paris. I received valuable comments from participants at these events. I thank especially Samuel Ferey and Alain Marciano. I also thank two anonymous reviewers for their remarks.

- D'Agostino, Fred. *Free Public Reason: Making It Up as We Go*. New York: Oxford University Press, 1996.
- Forrester, Katrina. *In the Shadow of Justice: Postwar Liberalism and the Remaking of Political Philosophy*. Princeton, NJ: Princeton University Press, 2019.
- Gaus, Gerald F. *Justificatory Liberalism: An Essay on Epistemology and Political Theory*. New York: Oxford University Press, 1996.
- . "Moral Conflict and Prudential Agreement: Michael Moehler's *Minimal Morality*." *Analysis* 79, no. 1 (January 2019): 106–15.
- . *The Order of Public Reason: A Theory of Freedom and Morality in a Diverse and Bounded World*. New York: Cambridge University Press, 2011.
- . "A Tale of Two Sets: Public Reason in Equilibrium." *Public Affairs Quarterly* 25, no. 4 (October 2011): 305–25.
- Gaus, Gerald, and Chad Van Schoelandt. "Consensus on What? Convergence for What? Four Models of Political Liberalism." *Ethics* 128, no. 1 (October 2017): 145–72.
- Gintis, Herbert. *The Bounds of Reason: Game Theory and the Unification of the Behavioral Sciences*. Princeton, NJ: Princeton University Press, 2009.
- Granovetter, Mark. "Threshold Models of Collective Behavior." *American Journal of Sociology* 83, no. 6 (May 1978): 1420–43.
- Hadfield, Gillian K., and Stephen Macedo. "Rational Reasonableness: Toward a Positive Theory of Public Reason." *Law and Ethics of Human Rights* 6, no. 1 (2012): 7–46.
- Hadfield, Gillian K., and Barry R. Weingast. "What Is Law? A Coordination Model of the Characteristics of Legal Order." *Journal of Legal Analysis* 4, no. 2 (Winter 2012): 471–514.
- Hédoin, Cyril. "Community-Based Reasoning in Games: Saliency, Rule-Following, and Counterfactuals." *Games* 7, no. 4 (December 2016): 36.
- . "A Framework for Community-Based Saliency: Common Knowledge, Common Understanding and Community Membership." *Economics and Philosophy* 30, no. 3 (November 2014): 365–95.
- . "Naturalism and Moral Conventionalism: A Critical Appraisal of Binmore's Account of Fairness." *Erasmus Journal for Philosophy and Economics* 11, no. 1 (Spring 2018): 50–79.
- Kogelmann, Brian. "Public Reason's Chaos Theorem." *Episteme* 16, no. 2 (June 2019): 200–219.
- Kogelmann, Brian, and Stephen G. W. Stich. "When Public Reason Fails Us: Convergence Discourse as Blood Oath." *American Political Science Review* 110, no. 4 (November 2016): 717–30.
- Lewis, David K. *Convention: A Philosophical Study*. Oxford: Blackwell, 2002.
- Milgrom, Paul. "An Axiomatic Characterization of Common Knowledge."

- Econometrica* 49, no. 1 (January 1981): 219–22.
- Quong, Jonathan. *Liberalism without Perfection*. Oxford: Oxford University Press, 2011.
- Rawls, John. “The Domain of the Political and Overlapping Consensus.” *New York University Law Review* 64, no. 2 (1989): 233–55.
- . “The Idea of Public Reason Revisited.” *University of Chicago Law Review* 64, no. 3 (Summer 1997): 765–807.
- . “Justice as Fairness: Political Not Metaphysical.” *Philosophy and Public Affairs* 14, no. 3 (Summer 1985): 223–51.
- . “Kantian Constructivism in Moral Theory.” *Journal of Philosophy* 77, no. 9 (September 1980): 515–72.
- . *Political Liberalism*. New York: Columbia University Press, 1993.
- . *A Theory of Justice*. Cambridge, MA: Belknap Press, 1971.
- Rubinstein, Ariel. “The Electronic Mail Game: Strategic Behavior Under ‘Almost Common Knowledge.’” *American Economic Review* 79, no. 3 (June 1989): 385–91.
- Sugden, Robert. *The Economics of Rights, Co-operation and Welfare*. 2nd ed. New York: Palgrave Macmillan, 2005.
- Taylor, Alan D. *Social Choice and the Mathematics of Manipulation*. Cambridge: Cambridge University Press, 2005.
- Thrasher, John. “Agreeing to Disagree: Diversity, Political Contractualism, and the Open Society.” *Journal of Politics* 82, no. 3 (July 2020): 1142–55.
- Thrasher, John, and Kevin Vallier. “The Fragility of Consensus: Public Reason, Diversity and Stability.” *European Journal of Philosophy* 23, no. 4 (December 2015): 933–54.
- . “Political Stability in the Open Society.” *American Journal of Political Science* 62, no. 2 (April 2018): 398–409.
- Vanderschraaf, Peter. *Strategic Justice: Convention and Problems of Balancing Divergent Interests*. Oxford: Oxford University Press, 2019.
- Weithman, Paul. *Why Political Liberalism? On John Rawls’s Political Turn*. New York: Oxford University Press, 2011.
- Wilson, Rick K., and Jane Sell. “‘Liar, Liar . . .’: Cheap Talk and Reputation in Repeated Public Goods Settings.” *Journal of Conflict Resolution* 41, no. 5 (October 1997): 695–717.
- Wittgenstein, Ludwig. *Philosophical Investigations*. Malden, MA: Wiley-Blackwell, 2010.
- Young, H. Peyton. *Individual Strategy and Social Structure: An Evolutionary Theory of Institutions*. Princeton, NJ: Princeton University Press, 2001.

THE PURPOSE AND LIMITS OF ELECTORAL ACCOUNTABILITY

Finlay Malcolm

THE IDEA that voters hold political rulers to account for their actions while in power is generally taken to be one of the features of electoral democracy that most clearly distinguishes it from non-democratic forms of government.¹ This notion of political accountability is often thought to be advantageous, since it strongly incentivizes rulers to act in the best interests of their citizens.² But to make political accountability work would require that voters know how well their rulers have performed, and base their votes on this knowledge by reelecting rulers who perform well, and de-electing those who perform poorly. However, recent studies show that voters regularly do not know how well their rulers have performed, and even if they do, do not base their vote on this knowledge. If this is correct, then the idea that political accountability comes through the electorate seems to be mistaken.

This paper develops a novel way of formulating this problem for political accountability, arguing that, if these studies of voter knowledge and behavior are accurate, then as presently conceived, political accountability cannot succeed. However, the paper will set out and defend an alternative conception of political accountability that is not susceptible to these limitations on the part of the voter.

The paper first (section 1) briefly sets out the idea of political accountability as it comes through elections—*electoral accountability*—and what is required of voters to make it work. Section 2 argues that political accountability faces a dilemma: either voters do not know how well their rulers have performed, or if they do, they do not base their votes on such knowledge. The first horn of the dilemma is explored through literature on political ignorance, while the second looks at evidence that shows that the more people know about political affairs, the more likely they are to be entrenched in a particular political stance, and

1 See Erkkilä, “Governance And Accountability,” 10–12; Guerrero, “Against Elections,” 137–49; Moncrieffe, “Reconceptualising Political Accountability.”

2 Arneson, “Defending the Purely Instrumental Account of Democratic Authority”; Christiano, *The Rule of the Many*.

hence the less inclined they would be to change a voting decision to a different political party or ruler, even if they have performed well. On either horn of the dilemma, electoral accountability, as standardly conceived, fails. Section 3 considers several responses to this dilemma, through aggregative and heuristic approaches to voter knowledge, and alternative systems of government, such as epistocracy, sortition, and lottocracy. It is argued that these approaches are either deeply problematic or entirely inadequate.

In section 4, the paper offers a new way to conceive of electoral accountability. It is argued that we should focus not on what voters know, nor on how citizens cast their vote, but on what they *would* likely know and how they *would* likely vote under conditions where the actions of rulers have substantially negative consequences that are both pervasive and highly salient. In such counterfactual scenarios, political ignorance and arbitrary voting would seem much less likely. On this new theory, electoral accountability is not about checking how well or poorly incumbent political rulers have performed. Rather, it is about preventing rulers from committing or allowing substantial harms to come upon those they govern. This alternative theory gives electoral accountability a limited, but extremely important, role within a well-functioning democracy, one that can be achieved despite well-known, and often quite rational, limitations within the electorate.

It is worth pointing out two issues from the outset. First, a theory of electoral accountability is compatible with a range of theories about how people vote. For instance, people may vote to express a political identity or affiliation, or to try to secure good outcomes for themselves that have been promised in election pledges.³ The account this paper develops does not deny that any of these reasons are central in determining an agent's voting decision. It simply separates these reasons out from those of *accountability*—the keeping of political rulers to account for their actions while in office. The question to be addressed is: Do elections keep political rulers accountable? The answer this paper gives to this question turns on how we view accountability. On standard theories of accountability, the answer will be no, but on the account offered here, the answer will be yes.

The second issue is that the theory of accountability I develop in this paper is part of a defense of democracy within the “realist” strand of recent democratic theory.⁴ Democratic realists can be taken to make two claims. The first, *epistemic* claim, is that voters are often, or perhaps almost always, ignorant or

3 Achen and Bartels, *Democracy for Realists*; Goldman, *Knowledge in a Social World*.

4 Achen and Bartels, *Democracy for Realist*; Brennan, “Does Public Reason Liberalism Rest on a Mistake?”

misinformed about most areas of political life.⁵ The second, *psychological* claim, is that people tend to vote on the basis of political affiliations that are analogous to sports team loyalties.⁶ This paper accepts both views, and indeed, will use literature that supports both of these claims to develop the dilemma for electoral accountability. However, unlike other democratic realists, I do not propose that accountability is not possible within current systems of representative democracy, nor will I suggest alternative forms of government that may make accountability more successful (I review these proposals in section 3). Rather, this paper proposes that electoral accountability within current systems is still possible if we change the way that we think about what electoral accountability is. Indeed, what I propose is how electoral accountability actually works. Given the realists' claims, electoral accountability may seem impossible. This paper argues otherwise: even given the realists' claims, electoral accountability is possible, but only if we reorient our views toward what electoral accountability actually is.

1. ELECTORAL ACCOUNTABILITY

There are several ways by which rulers are held accountable for their actions and policies. First, there is *horizontal* accountability, where rulers are critiqued by other professional politicians of one's own or another's political faction or party, through such measures as parliamentary debate and joint committees.⁷ Such methods can often include votes for or against bills and policies that may pass into law, and so mark a crucial way of holding rulers to account. Horizontal accountability can also come from rulers of other nations, as is the case with multinational institutions like NATO and the EU.⁸ A second form of political accountability comes externally through analysis and reporting from a free press, which provides information and scrutiny to a range of stakeholders, including politicians, businesses, and the public.

This paper focusses on a third, *vertical* kind of political accountability, which comes through the process of regular, free, and fair elections. The basic idea is that, first, the members of the electorate are given, through their right to vote,

5 See Achen and Bartels, *Democracy for Realists*; Brennan, *Against Democracy*; Caplan, *The Myth of the Rational Voter*; Delli-Carpini and Keeter, *What Americans Know about Politics and Why It Matters*; and Somin, *Democracy and Political Ignorance*. For a general critique of peoples' tendency to believe falsely, see Duffy, *The Perils of Perception*.

6 See Achen and Bartels, *Democracy for Realists*; and Mason, *Uncivil Agreement*.

7 See Laver and Shepsle, "Government Accountability in Parliamentary Democracy." For the distinction between *horizontal* and *vertical* accountability, see O'Donnell, "Delegative Democracy?"

8 Hirst, "Democracy and Governance."

the power to remove rulers from office if the electorate deems those rulers to have performed poorly, or to keep the rulers in power if they are deemed to have performed well. Call this *electoral power*. In turn, incumbent rulers are incentivized to perform well, in order to retain power in elections. Call this *the electoral incentive*. On this view, electoral power generates the electoral incentive: the fact that citizen voting rights empower or disempower political rulers on the basis of their performance generates for the rulers an incentive to perform well. Due to the strength of the electoral incentive, rulers have a significant motive to do all they can to perform well while in office. After all, winning elections is about retaining power, and so if retaining power is about performing well, then to win elections, rulers need to perform well while in power.⁹ Let us call this general description of electoral accountability, involving both *electoral power* and *the electoral incentive*, the *Standard Theory of Electoral Accountability* (STAN).

It follows from STAN that there is both a reward and punishment component to electoral accountability. If rulers are deemed to perform poorly or to fail to meet the demands and expectations of the voters, then voters can sanction or *punish* them by removing them from power.¹⁰ But if they have performed well or have met the required expectations, then the voters can *reward* the rulers by keeping them in power.

Electoral accountability, as conceived by STAN, is implicit both in arguments for democracy and in theories of voting. Consider the argument from J. S. Mill that participation in political affairs is required for citizens' interests to be taken into account:

The rights and interests of every or any person are only secure from being disregarded when the person interested is himself able, and habitually disposed to stand up for them . . . human beings are only secure from evil at the hands of others in proportion as they have the power of being, and are, self-protecting.¹¹

So, to have one's rights and interests considered by rulers, and to secure oneself from evil at the hands of others, Mill insists people need to be given the power (and have the personal interest to use that power) that comes through suffrage.

9 As Manin et al. note in *Democracy, Accountability, and Representation*, "what ultimately matters for accountability is . . . survival in office" (18).

10 See Darby and Martinez, "Making Identities Safe for Democracy," 12. This view treats elections as "a 'contingent renewal' accountability mechanism, where the sanctions are to extend or not to extend the government's tenure." See Manin et al., *Democracy, Accountability, and Representation*, 10.

11 Mill, *Considerations on Representative Government*, 63; Dahl, *Democracy and Its Critics*, 93–95.

In the background to this is electoral accountability, for the idea is that if rulers disregard or even harm the interests of enfranchised citizens, then the citizens will regard the rulers as having performed poorly, and hence will remove the rulers from power. As such, rulers have the electoral incentive to protect and not harm the interests of the enfranchised.

To elaborate on this point, consider the extant case of the Uyghur people in China's Xinjiang region. A recent report by the US government estimates that China's government has detained more than one million members of religious groups in internment camps

and subjected them to forced disappearance, political indoctrination, torture, psychological and physical abuse, including forced sterilization and sexual abuse, forced labor, and prolonged detention without trial because of their religion and ethnicity.¹²

On Mill's Argument, and STAN, if the Uyghur people had fair voting rights, then the Chinese government would have an electoral incentive not to commit the harms that have been reported, and thus to improve the Uyghurs' interests. In effect, the Uyghurs and other Chinese citizens would hold the government electorally to account for their actions. (We will return to the plausibility of this example in section 3.)

Electoral accountability is also assumed within theories of voting. In particular, in his *retrospective theory of voting*, V. O. Key depicted "the electorate in its great, and perhaps principal, role as an appraiser of past events, past performance, and past actions. It judges retrospectively."¹³ The idea here is that enfranchised citizens look back to how well rulers have performed to determine, through voting, whether or not to keep the incumbent ruler(s) in power. This theory of voting also connects with Mill's idea that voting enhances the interests of the electorate: "By basing their votes on evaluations of performance, voters . . . motivate officeholders to pay attention to the interests of the electors."¹⁴ Again, we see the electoral incentive being used to improve the performance of the incumbent political rulers for, if they perform poorly, they will be dispossessed of power.

It is worth noting three clarifications about STAN. First, it is not itself a theory of voting, but a consideration citizens have when making a voting decision. Voters cast their votes on a range of grounds, including offers made in

12 See the 2019 Report on International Religious Freedom: China—Xinjiang: <https://www.state.gov/reports/2019-report-on-international-religious-freedom/china/xinjiang/>.

13 Key, *The Responsible Electorate Rationality in Presidential Voting*, 61.

14 Ferejohn, "Incumbent Performance and Electoral Control," 7.

campaign manifestos and party-political loyalties. Electoral accountability is simply the idea that a significant proportion of the electorate weighs in the performance of the incumbent rulers when casting their votes, and in so doing, holds those rulers to account for their actions. Recent studies of voter behavior confirm this point. According to Jonathan Woon, people involved in making electoral decisions exhibit “a strong behavioral tendency to vote retrospectively, which in turn induces office-motivated politicians to act in the voter’s best interests.”¹⁵ But, second, even if electoral accountability does incentivize rulers to act in voters’ interests, as Mill and the electoral incentive suggest, it does not guarantee this. Rulers have other incentives to act that will be weighed against the electoral incentive of any particular group of voters. For instance, they may ignore the particular interests of one group if it is especially small, or unlikely to return a positive electoral vote even if their interests are supported. Third, a lack of electoral accountability would not prevent rulers from promoting the interests of the citizens they govern. There are other reasons why rulers might promote the interests of their citizens, including considerations of benevolence and justice. Electoral accountability simply provides a strong incentive to act in the interests of those they govern by performing well while in power.

For STAN to work, there are two conditions that must be met. To show this, let us consider a simple analogy. Suppose you hire me to work on your farm under the condition that, if I perform well, you will pay me, and if I perform poorly, you will not pay me. In this case, I have a *pay incentive* to perform well. Now, consider two extensions to the example:

No Knowledge: After a few days of working for you, I realize that you have no idea what I have been doing—you do not check on my work after I do it, and you do not ask anyone else how my performance has been. Despite lacking knowledge of my performance, you pay me some days and not others.

No Knowledge-Basing: You know everything about my performance. But when it comes to paying me, you flip a coin, and if it comes up heads, you pay me, and if it is tails, you do not. Some days I get paid when I have done a bad job, and other days I do not get paid, even when I have done a good job.

It seems that in both cases, I actually lack the *pay incentive* to perform well. In both *No Knowledge* and in *No Knowledge-Basing* I could just sit at home all day and still possibly receive pay. Indeed, it would be against my interests to

15 Woon, “Democratic Accountability and Retrospective Voting,” 927.

perform well in order to receive pay since I could work my fingers to the bone, and do an outstanding job, and still receive nothing in return.

The analogous point also applies to elections.¹⁶ Consider two symmetrical cases:

No Knowledge: The rulers of a nation realize that the voters have no idea what they have been doing—the voters do not check on the rulers' work, and do not ask anyone how their performance has been. Despite lacking knowledge of their performance, some citizens vote for them and some against them.

No Knowledge-Basing: The voters know everything about the rulers' performance. But when it comes to election day, the voters flip a coin, and if it comes up heads, they vote to keep the rulers in power, and if it is tails, they vote to get rid of the rulers. Sometimes, the rulers get to keep power even when they have done a bad job, and other times, they are disempowered even when they have done a good job.

Again, the rulers in both cases would lack the *electoral incentive* to perform well. In both *No Knowledge* and *No Knowledge-Basing*, the rulers could do no work for their entire term and still possibly retain power. That is not to say they have *no* incentive to perform well. It is just that they would not be held *electorally accountable* for their actions, and hence there would be no electoral incentive for those in power to perform well.

What these cases show is that two conditions must be met for STAN to work:

Knowledge Condition (KC): Electoral accountability requires the electorate to know how well their political rulers have performed while in office.

Knowledge-Basing Condition (KBC): Electoral accountability requires the electorate to base their votes on what they know about how well their political rulers have performed while in office.¹⁷

In the next section, I will present a dilemma for STAN. On the one hand, many voters will fail to satisfy KC. On the other, those who satisfy KC will fail to

16 Guerrero makes the same point in "Against Elections": "If people are ignorant about some issue, or about what their representative is doing with respect to that issue, or about whether what their representative is doing is good, they cannot monitor or evaluate what their representative is doing with respect to that issue" (145).

17 While KC and KBC are necessary conditions for electoral accountability to succeed, they are not jointly sufficient. For instance, we still require, among other things, elections to be free and fair, and suffrage to be universal and equal.

satisfy κBC . If this is correct, then neither condition is satisfied, and electoral accountability, as conceived by STAN , cannot succeed.

2. THE LIMITATIONS OF THE ELECTORATE

To evaluate the standard theory of electoral accountability in light of the first epistemic condition (κC), we can ask: Does the electorate know how well their political rulers have performed while in office? This is a vast empirical question. There are many policy areas over which political rulers could be judged to perform well or poorly, including the environment, crime rates, health care, foreign policy, inequality, and immigration. To proceed, I propose to focus on the question of voter knowledge in relation to the one issue that citizens most regularly recognize as being of the highest importance for them: the economy. If we focus first on this issue, we will be able to see that many voters in fact fail to satisfy κC , and from there we can make some salient extrapolations to other issues as well.

Studies of voting behavior have consistently found that citizens vote *sociotropically*—in accordance with their view of the national economic condition. Lewis-Beck and Stegmaier’s broad review of studies on sociotropic voting in the UK, US, and France finds a consistent pattern that

the economy reliably moves voters to hold their government accountable in national elections. When they see prosperity, they give support. When they see business conditions in decline, they withdraw support.¹⁸

For instance, in a study of voting in the 1996 US election, voters who believed the national economy was “better” were 38 percent more likely to vote for the incumbent (Clinton).¹⁹ Other studies show that the importance of the economy for voters in the US was the same in 1992 as it had been in 1996.²⁰

In Great Britain, David Sanders examined the effects of national economic perceptions in the five general elections between 1974 and 1997.²¹ Drawing from the well-established British Election Study, Sanders found that in 1974, odds of a vote for the Conservative Party were doubled when believing the economy had worsened. Across the range of elections, Sanders concludes that “the governing party loses support among those voters who believe that eco-

18 Lewis-Beck and Stegmaier, “Economic Models of Voting,” 530.

19 Alvarez and Nagler, “Economics, Entitlements, and Social Issues,” 1360–62.

20 Norpoth, “Bush v. Gore,” 53.

21 Sanders, “Party Identification, Economic Perceptions, and Voting in British General Elections, 1974–97.”

conomic conditions have worsened.”²² In the 2001 general election, when the incumbent Labour Party won a significant majority of 413 parliamentary seats, 70 percent of voters saw the economic past as the “same or better,” and 68 percent of voters saw the economic future as the “same or better.”²³ For voters, economic conditions ranked second behind the National Health Service in terms of issue priority.²⁴

A similar picture emerges in France as well. Issues salient to the economy, particularly unemployment and inflation, are central issues for French voters. As Lewis-Beck et al. found from French voter surveys of the 1995, 2002, and 2007 elections, unemployment was always the number one issues for voters. They also found that economic concerns of inflation ranked numbers five and three in 1995 and 2007, respectively; inequality ranked numbers three and two, respectively, in 2002 and 2007; and deficits ranked number five in 2007.²⁵

While voters clearly use sociotropic considerations to judge how well their political rulers have performed while in office, do they *know* how well these governments have performed? Have they got their assessments correct about government performance on economic issues? According to some recent studies of sociotropic voting, the electorate in many countries is often misinformed about government performance with the economy. One recent study by Achen and Bartels that bears out this concern focusses on the problem of “end bias,” where “voters seem to evaluate incumbents on the basis of election-year economic outcomes rather than cumulative economic performance.”²⁶ In their study, Achen and Bartels focussed on the influence of incumbent economic performance in the final two quarters of an election cycle on voter behavior in US elections since the mid-twentieth century. They argue that, for “the cumulative rate of real income growth in the 13 quarters leading up to Election Day . . . every additional percentage point of income growth increased the incumbent party’s expected popular vote margin by almost 1.5 percentage points.”²⁷ This confirms the sociotropic claim that the economy is a priority in voting decisions. But in contrast, focussing merely on “Q14 and Q15—the six months leading up to Election Day,” they found that “every additional percent-

22 Sanders, “Party Identification, Economic Perceptions, and Voting in British General Elections, 1974–97,” 261.

23 Clarke et al., *Political Choice in Britain*, 84–85.

24 Clarke et al., *Political Choice in Britain*, 90.

25 See Lewis-Beck et al., *French Presidential Elections*, ch. 5; Nadeau and Lewis-Beck, “French Election Theory,” 57.

26 See Achen and Bartels, *Democracy for Realists*; Huber et al., “Sources of Bias in Retrospective Decision Making,” 725.

27 Achen and Bartels, *Democracy for Realists*, 152.

age point of income growth increased the incumbent party's expected popular vote margin by more than 6 percentage points."²⁸ That is, for each additional percentage point of income growth, the incumbent party was four times more likely to receive voter support when that percentage point occurred in the six months leading up to a general election.

End bias highlights the human propensity to bring to mind more immediately available evidence when making judgments—sometimes called “the availability heuristic.”²⁹ As an epistemic problem, this bias disposes voters to gather evidence in a way that points to incorrect or misleading conclusions. If voters were making a properly retrospective evaluation of incumbent performance, then they would judge incumbents on cumulative performance over an entire term, rather than on the basis of recent economic results alone.

Achen and Bartels say that end bias produces “myopic” voters who unfairly judge economic performance in terms of the final six months of an election cycle. This has direct implications for STAN, since how should we expect incumbents to behave given such myopic retrospection? According to Achen and Bartels, incumbents

should attempt to maximize income growth in the immediate run-up to elections, but care little about what happens to the economy at other times . . . there is little or no *electoral* incentive for presidents to promote myopic voters' well-being during much of their time in office.³⁰

So, because voting is generally myopic, not only does it fail to incentivize incumbents to promote citizen interests through effective economic management for the majority of their time in office, but it can actually do damage to those interests because the incumbents can do economic harm without facing punishment. For, an incumbent could perform poorly with the economy over the full term, but improve in the final six months, and gain reelection on that basis. This would seem to *undermine* the role of electoral accountability, particularly when it is taken to improve citizen interests.³¹

The conclusion to draw from this is that, with respect to sociotropic voting, there seem to be many citizens in the US who, although they believe that they know how well their government has performed economically over its tenure, are in fact misinformed. On this issue, then, the electorate fails to satisfy KC, and hence electoral accountability as standardly conceived fails with it. But

28 Achen and Bartels, *Democracy for Realists*, 153.

29 Tversky and Kahneman, “Availability.”

30 Achen and Bartels, *Democracy for Realists*, 170.

31 E.g., Mill, *Considerations on Representative Government*; Dahl, *Democracy and Its Critics*.

since end bias is a general human disposition, we have some reason to suppose that it occurs in other elections outside of the US as well, and affects other issues, including crime, health care, immigration, the environment, foreign policy, and inequality.³² The extensive literature on political ignorance reveals widespread areas of misinformation among voters on many of these issues.³³ To dwell on this literature is not necessary to make the point already made for sociotropic voting: voters lack knowledge—to at least some significant extent—on key factors determining the performance of political rulers, and because of this, those rulers are less incentivized to perform well on these issues.

Now, perhaps this is incorrect. Perhaps voters *do* have knowledge of how well incumbents have performed while in office. In other words, κC *does not* fail, as suggested by the studies on voter ignorance. After all, even studies that outline the extent of voter ignorance indicate that the electorate does have some knowledge of government performance.³⁴ For example, Somin claims that the “biggest issue in the important 2010 [US] congressional election was the economy. Yet two thirds of the public did not realize that the economy had grown rather than shrunk during the previous year.”³⁵ But this still means that one-third of the public may well have known that the economy had grown rather than shrunk the previous year. So, perhaps governments have *some* electoral incentive to do well with the economy because *some* of the voters have salient knowledge of their performance. And if one-third of the voters have this knowledge, then that is a significant minority, and so the incentive to do well with the economy will also be significant. The remainder of this section considers a response to this point: that even though there are voters who satisfy κC , those knowledgeable voters will fail to satisfy κBC —they will not base their vote on such knowledge—and hence the incumbent government still lacks the electoral incentive.

To begin, it is worth stating that people tend to gather intricate knowledge on issues that interest them. Someone who learns as much as she can about coffee, or the Brazilian football team, or *Game of Thrones* will almost always be someone who cares a lot about, respectively, coffee, the Brazilian football team,

32 Huber et al, “Sources of Bias in Retrospective Decision Making.”

33 See Brennan, *Against Democracy*; Somin, *Democracy and Political Ignorance*.

34 Studies of voter ignorance often note that such ignorance is rational, or at least not irrational. For instance, in “Is Political Ignorance Rational?” Somin acknowledges the view, often attributed to Downs, *An Economic Theory of Democracy*, that since a single vote has only a fractional chance of changing the outcome of an election, the costs of informing oneself swamp the value of that vote. For discussion of this issue in the context of electoral accountability, see Hardin, “Democratic Epistemology and Accountability,” and for criticism, see Barnett, “Why You Should Vote to Change the Outcome.”

35 Somin, *Democracy and Political Ignorance*, 1.

or *Game of Thrones*. Because people care about these things, they also tend to have entrenched preferences toward them. The coffee connoisseur has a keen liking of coffee, and is unlikely to suddenly change her mind and come to dislike it.

A similar point is thought to apply to people who are knowledgeable about politics: those who know a lot about politics tend to also be those who care a lot about politics.³⁶ What is more, caring about politics tends to go along with specific partisan loyalties: if you are more politically knowledgeable, then you will tend to follow a particular political party that aligns to your identity.³⁷ Just as the person who cares about the Brazilian football team goes out to learn more about them and is an avid supporter of that team, so the person who cares about the Labour Party in the UK, or the Democratic Party in the US, goes out to learn more about the workings of government in general, and their own party in particular. Now, it is not that being knowledgeable makes one politically partisan, but that being partisan tends to go along with a keener interest in political affairs, which then leads to the acquisition of knowledge of those affairs: it is the partisan loyalty that motivates people to seek out information that satisfies the interest.

The problem that is often pointed out with this approach to acquiring political knowledge is that it is acquired through motivated, and hence biased, reasoning. That is, people tend to seek out information that tells them what they want to believe, and confirms the beliefs they want to hold. For instance, if someone is already strongly partisan toward the Labour or Democratic Party, then the information they will look for is that which supports positive beliefs and views about them, and opposes looking positively at, say, the Conservative or Republican Parties.³⁸ As Gunn puts it, “the more political knowledge people possess, the more “constrained” by ideology they tend to be,” in the sense that this ideology motivates their reasoning to preserve and reinforce their pre-existing partisan beliefs.³⁹ So, because people tend to have entrenched political allegiances, then the information they acquire on political affairs will

36 See Hannon, “Are Knowledgeable Voters Better Voters?”; Somin, *Democracy and Political Ignorance*, 93.

37 For evidence of this claim, see Converse, “The Nature of Belief Systems in Mass Publics”; Federico, “Predicting Attitude Extremity”; Feldman and Price, “Confusion or Enlightenment?”; Hetherington, “Putting Polarization in Perspective”; and Kalmoe, “Uses and Abuses of Ideology in Political Psychology.” Much of the literature on this issue is collated and discussed at length in Hannon, “Are Knowledgeable Voters Better Voters?” For response to Hannon, see Gibbons, “Is Epistocracy Irrational?”

38 See Kelly, “Intergroup Differentiation in a Political Context”; Green et al., *Partisan Hearts and Minds*; Huddy, “From Social to Political Identity”; Mason, *Uncivil Agreement*.

39 Gunn, “Against Epistocracy,” 35.

be sought to enhance those allegiances, rather than to make impartial judgments of the actions of political rulers.

One of the ways we could interpret the relevance of these claims for STAN is to assume that people acquire knowledge of how well or poorly the incumbent rulers have performed, but fail to take this into account when casting a vote. After all, if political loyalties are entrenched, particularly among the politically knowledgeable, then those loyalties will usually go along with electoral support for the party one is loyal toward. Hence, the acquisition of knowledge on the performance of incumbents will not ground a decision about whom to vote for. One could judge that the incumbent rulers have performed excellently, and yet refuse to support them at the ballot box because of political loyalties one holds toward a rival faction. The likely reason for this will be that she interprets the evidence supporting a good performance of the incumbent in a way that means that she still believes her own political party would be better in power. This does not make this agent's vote arbitrary—she still has reasons for not supporting the incumbent rulers—but it does disconnect her vote from any impartial judgments she should make of the performance of the incumbent rulers. Her judgment could also be correct—the party to which she is affiliated could be that which should be in power from the perspective of accountability. However, this may be due to luck rather than rational judgment, and if the situation changes, it might not mean that her judgment would change with it.

If this is in fact the situation, as the literature appears to show that it is, then we have knowledgeable voters who fail to *base* their vote on that knowledge, and hence who fail to satisfy KBC. This has significant consequences for STAN, for the incumbent rulers would then lack the electoral incentive to perform well while in power for fear that by performing poorly they will lose that power. It does not matter if they perform well or poorly, as they still will not convince others of a rival party to back them, and they will not lose support from their own base. Hence, we undermine the key motive that supports STAN.

We seem to have arrived at a dilemma. Recall that, for STAN to work, voters must know how well their political rulers have performed while in office (κC), and must base their vote on this knowledge (κBC). But, either voters do not know how well their rulers have performed, or if they do, then those voters have loyalties that are so entrenched that such knowledge makes no difference to the way they vote. That is, voters either fail to satisfy κC , or if they do satisfy κC , then they fail to satisfy κBC . In either case, electoral accountability, as conceived by STAN, does not work. In the next section, we will consider several ways of responding to this dilemma, each of which suffers from its own limitations. Then, in section 4, I will offer a further way of resolving the dilemma by developing a novel view of electoral accountability.

3. AGGREGATION, HEURISTICS, AND ALTERNATIVE SYSTEMS

There are at least three ways to respond to the dilemma for the STAN version of electoral accountability. First, deflate the impact of the empirical findings I have cited by, for instance, rejecting the claim that voters are broadly politically ignorant, or positing that, even if they are, accountability can still be achieved by pooling the electorate's knowledge as a whole. Second, propose changes to the political system by, for instance, enfranchising only those who have political knowledge, and thus hopefully securing KC; or, by selecting voters by lot to try to restrict the influence of partisan loyalties, in order to satisfy KBC. Third, reconceive of electoral accountability in a way that is not susceptible to the dilemma facing STAN. In this section, I will explore and critique the first two proposals. In the next section, I will develop on the third.

The first response itself can be approached in three different ways. First, one could seek to reject the veracity of the empirical findings I have cited. It is not the aim of this paper to challenge these empirical findings. I take them to be well-established in political science and psychology, dating back to at least the mid-twentieth century in the work of Downs, who argued that political ignorance was rational for the voter given their minute contribution to the overall electoral outcome.⁴⁰ As Friedman summarizes the findings:

That the public is overwhelmingly ignorant when it comes to politics is ... a discovery that has been replicated unflinching by political scientists; indeed, it is one of the strongest findings that have been produced by any social science—possibly *the* strongest.⁴¹

Despite this, rather than attempt to discredit, challenge, or prove this vast body of literature, I want to more modestly suggest that the dilemma I have proposed for STAN be read as a conditional: if the findings from political science and psychology I have cited are at least largely accurate, then electoral accountability, as conceived by STAN, does not work.

One way to push back against this claim—and the second way of deflating the empirical findings I have cited—is to show that voters can make up for their lack of salient political knowledge through the use of heuristics. If that were the case, then electoral accountability may still be possible. But what heuristics would be relevant here? People often use political parties as an effective route from which to infer policy stances. They could also lean on activists who give support to particular parties due to their specific policy commitments. But

40 Downs, *An Economic Theory of Democracy*.

41 Friedman, "Introduction," 397.

neither of these routes would be helpful for STAN, which requires voters to accurately track past *performance*, not current or future policy.

One heuristic that could be helpful is relying on discrete public groups, or bodies who keep track of political performance, so that the general public can turn to them for information at the time of an election. This issue is similar to the idea of “issue publics.”⁴² While this would no doubt be helpful if the information groups were independent, nonpartisan bodies, they are certainly not widely used at present. As noted in section 2, voters increase their support for an incumbent party dramatically when they do well closer to election time, and at the same time disregard older evidence. This suggests that voters have limited knowledge of incumbent performance, and so if there are issue publics on sociotropic measures, the general public rarely uses them. Instead, they take their information on government performance from the media, and their family and friendship groups. Even if reliable public groups exist, citizens appear to be (rationally) ignorant of them, in the same way they are ignorant of political matters more generally. One way to improve on this could be to formalize the idea of nonpartisan, independent, election-time information groups. I will return to this issue shortly when I consider alternative forms of government. But suffice it to say for now that it is not clear what heuristics are in place that can make up for the shortfall in voter knowledge of past performance.

A third, alternative way to deflate the effects of the empirical findings is to show that, even if, individually, voters are largely ignorant, when taken in aggregate, they are capable of making informed decisions. This is partly because of the diversity of people’s knowledge when taken in aggregate. For instance, ten people with one distinct unit of knowledge will know more than one knowledgeable expert with five units of knowledge. Given that the electorate is extremely diverse, then, in theory at least, it can pool what it does know, so that taken together, it is a highly knowledgeable unit.

Perhaps the leading account that has theoretically modelled this idea is by Landemore, who explored the way that problem solving and information pooling work in jury deliberation to deliver the correct result.⁴³ On her account, when deliberating about a decision, jurors (from a fictional case) dedicate themselves to “collectively brainstorming the available information and arguments and putting them through the many filters and lenses of the group.”⁴⁴ This deliberative process helps the jurors to make the best use of their cognitive skills, and the information at their disposal, to arrive at the correct

42 Iyengar, “Shortcuts to Political Knowledge.”

43 Landemore, *Democratic Reason*.

44 Landemore, *Democratic Reason*, 3.

outcome. She adds, though, that the diversity brought by the group will only trump expert individuals when it comes to arriving at good decisions if the group are “relatively smart (or not too dumb).”⁴⁵ And here we see the problem with using an analogy from a jury to an electoral decision to overcome political ignorance. First, jurors normally take their responsibilities seriously because of the weight of their individual vote (one in twelve, say), and so take time to inform themselves and deliberate carefully with others. In contrast, as Somin notes, “most voters spend either little or no time collecting political knowledge, or focus primarily on conversation partners and media that reinforce their pre-existing biases.”⁴⁶ And given the weight of their vote (one in many millions), this behavior seems rational. Second, it is not clear, given this and the other literature cited on political ignorance, that with respect to political knowledge, voters *are* relatively smart (or not too dumb). So, it is not clear that theories of voter aggregation and diversity are able to overcome the problems of political ignorance within the electorate.⁴⁷

The examples we have considered do not seem to provide any clear or obvious way to deflate the impact of the empirical findings cited in section 2. So, a second approach to the dilemma is to entirely accept its force and the findings used to support it, but to propose that we find alternative forms of government that are not susceptible to it. A number of such proposals have been made in recent years, often with the aim of improving the standards of politics, and in some cases, on the basis of failures of electoral accountability.⁴⁸ Here, I will consider three such proposals.

First, there are forms of so-called epistocracy.⁴⁹ These can include *restricted suffrage*, where people who lack salient political knowledge are disenfranchised (perhaps on the basis of a voter exam), or *plural voting*, where people with more salient knowledge are granted additional votes.⁵⁰ In theory, either, or a combination, of these systems could enhance electoral accountability. In either case, and particularly when combined, it would be more likely that the electors would satisfy κC —indeed, depending on the conditions in a test of salient knowledge, it could be virtually guaranteed that the electors have the

45 Landemore, *Democratic Reason*, 102.

46 Somin, *Democracy and Political Ignorance*, 132.

47 For discussion and critique of alternative models of aggregation, see Brennan, *Against Democracy*, 180–94; and Somin, *Democracy and Political Ignorance*, ch. 4.

48 See Guerrero, “Against Elections,” 137–49.

49 Estlund, “Why Not Epistocracy?”

50 See Brennan, *Against Democracy*, 211–14, on restricted suffrage; Mulligan, “Plural Voting for the Twenty-First Century,” on plural voting. The two forms of epistocracy are compatible—and often traced to Mill, *Considerations on Representative Government*.

requisite knowledge. However, these forms of epistocracy face two problems. First, being knowledgeable might make it more likely that the electors would fail to satisfy KBC for, as was argued in section 2, their enhanced knowledge could indicate a rigid partisan affiliation. Second, citizens who tend to be more knowledgeable also tend to have certain demographics, like being white, male, and non-working class, and so, since people from these demographics will be overrepresented within the electorate, governments will be biased in their policies toward them, leading to unfair and potentially corrupt political rule.⁵¹ The first of these problems makes epistocracy an unviable solution to the dilemma of electoral accountability, and the second, an unlikely solution to poor political policy. On this basis, I will leave aside restricted suffrage and plural voting, and instead consider two alternative systems that have been proposed.

Second, there are *sortition* systems, such as López-Guerra's "enfranchisement lottery."⁵² His system has two devices. In the first, the "exclusionary sortition," "there would be a sortition to disenfranchise the vast majority of the population. Prior to every election, all but a random sample of the public would be excluded."⁵³ Although the sample will be random, López-Guerra holds that the lottery "would produce an electorate that would be demographically identical to the electorate under universal suffrage." The second device is a "competency-building process" that has been "carefully designed to optimize [the electorate's] knowledge about the alternatives on the ballot."⁵⁴ Again, in this kind of system, you could virtually guarantee that the enfranchised would satisfy KC by giving them the knowledge salient to evaluate government performance. Moreover, the enfranchisement lottery would be less susceptible to the problems facing the two forms of epistocracy we considered. For the electorate are not selected for their knowledge, which, as we suggested, could well indicate partisan loyalty. Rather, they are given knowledge in much the same way as a jury might, and what's more, since there will be far fewer electors, there will be greater weight given to each elector's vote. So, we might expect that the newly enfranchised would behave more closely to a jury, in the way predicted by Landemore's democratic model. Finally, since the enfranchised are demographically representative, incumbent rulers cannot privilege certain groups in their policy making while in office in the way they could with epistocracy.

51 This has come to be known as "the demographic objection" to epistocracy (See Estlund, *Democratic Authority*, 215–19). For responses, see Brennan, "Does the Demographic Objection to Epistocracy Succeed?"

52 López-Guerra, *Democracy and Disenfranchisement*.

53 López-Guerra, *Democracy and Disenfranchisement*, 4.

54 López-Guerra, *Democracy and Disenfranchisement*.

All of these features of the sortition system seem like a win-win. Indeed, I accept that, in theory, sortition would deliver better electoral accountability, and thus possibly better public policy than universal suffrage.⁵⁵ But I have two misgivings with this system. First, it conflicts with my stated aims in this paper, which were to show that electoral accountability within current systems of democracy is still possible. Showing *that* might make sortition less desirable than current systems of democracy, even if sortition can offer better electoral accountability. This is because changing the current system is impractical and therefore unlikely, and so a conception of accountability that works less well within existing systems of democracy may be preferable to one that works better within an idealistic system. The second misgiving is that sortition may, in fact, be less ideal than it initially appears. For instance, smaller electorates are liable to corruption by being bribed, or even threatened, by external interest groups. There is also the problem of who designs the competency-building process. It could be manipulated by the incumbent to present the voters with skewed data, leading them to think better of them than they perhaps ought to. The fact that alternative systems of government are not even theoretically ideal options gives us another reason to favor seeking a kind of electoral accountability that works within our current system.

Third, there are *lottocratic* systems that remove elections altogether.⁵⁶ In Guerrero's system, members of the public are chosen by lot for a short period of time to stand as political decision makers on single-issue legislation. They learn about the issue from experts and interact with other members of the public to take in a variety of public opinions. Since lottocracy does away with elections, it also does away with electoral accountability. That does not mean, though, that there is *no* accountability—there will still be horizontal forms of accountability for those selected. Now, of course we cannot say whether lottocracy is better for electoral accountability than current systems of democracy, since it has none. But it may well deliver more competent political decision makers, partly because they are not focused on retaining power, and so do not have to waste time cultivating their public image, but also because they will be focused on fewer issues, and so can be more dedicated to the issues at hand. As a result, lottocracy may well lead to better policy outcomes than current democratic systems.

As with sortition, it is only really the aim of this paper to explore how we should conceive of electoral accountability. However, there is one important reason, which is salient to the account I will develop in section 4, why democracy with universal suffrage is preferable to lottocracy. To preempt the

55 See Malcolm, "Epistocracy and Public Interests."

56 Bourcibus, "Democracy through Multi-Body Sortition"; Guerrero, "Against Elections."

account, we might wonder what happens when a lottocracy becomes corrupt, and the people in place refuse to leave power. There is no *electoral* mechanism to remove them—lottocracy lacks elections by definition. This limitation does not stand, however, in the case of democracy with universal suffrage. If political rulers become corrupt, and harm their own people, then the people have a recourse to remove them from power, namely, through elections. So, while elections may well lead to less competent politicians and worse policy, they retain an important device lost by the lottocratic system: the electoral capacity to remove corrupt rulers from power. Now, we could design lottocratic systems in which there are non-electoral ways to remove corrupt rulers from power. And democracy may well produce more corrupt politicians than would lottocracy. But at least to the extent that democracy with universal suffrage has in place electoral accountability, then in that respect, it has a benefit that lottocratic proposals lack.

So, while there are ways of restructuring the political system, each of these alternative systems faces its own problems, so if we want to retain our current system of electoral democracy with universal suffrage, then we need to explore whether electoral accountability is a defensible notion. This takes us to the account I want to develop in the final section. The idea that I will set out and defend accepts the full force of the dilemma identified in section 1, and agrees that electoral accountability, as conceived by STAN, does not succeed. However, rather than trying to deflate the dilemma, or restructure the political system, my approach *reconceives* of electoral accountability. In the next section, I will defend this novel account of electoral accountability, and argue that it can succeed despite the limitations of voter knowledge and voting behaviors.

4. THE COUNTERFACTUAL THEORY OF ELECTORAL ACCOUNTABILITY

The standard conception of electoral accountability (STAN) focusses on what voters know about the actions of their government, and how they behave in light of that knowledge. To see whether this idea has real purchase, we have looked at actual voters in current democracies in the West—principally the US, but also the UK and France. As far as voter knowledge and behavior goes in these states, there seems to be a lack of electoral accountability. But what if electoral accountability is actually occurring, despite limitations in voter knowledge and behavior? Indeed, what if there is genuine electoral accountability being achieved in these states, but it would only be visible if the situation in current democracies was very different? The kind of counterfactual situations we could imagine would be ones in which governments are disposed toward broad-scale tyrannical actions, such as subjecting their own people to widespread rights

violations, willfully allowing them to suffer extreme famine, or committing genocide on huge numbers of the population. These sorts of events have not occurred in recent years in the countries we have been investigating. Perhaps the reason such events do not occur in current democracies is because the electorate *would* hold their government accountable, so because of this, the governments do not allow them to happen. So, perhaps electoral accountability is a counterfactual accountability mechanism that prevents political rule from turning tyrannical. That would explain how it is occurring, but is not visible.

On this alternative account, we could think of electoral accountability as simply being disposed to depower a government if it engages in widespread, significant harms to the people it governs. That would be different from the STAN account, which treats the electorate as a supervisor of the government, who reviews its performance on complex issues like inflation, employment, and international trade. But these are matters that may well be beyond most citizens to understand, let alone keep track of over a term in office. The role of the electorate on the alternative account being proposed is much easier to satisfy, since those governmental actions would presumably be more salient. If this theory is correct, then since it is proposing that accountability is an unrealized disposition in the electorate because governments are not, it seems, engaging in widespread harms, that would make it invisible to social scientists. It would, in a sense, be a victim of its own success. Let us call this idea the *Counterfactual Theory of Electoral Accountability* (COUNT). On COUNT, electoral accountability is a backstop to tyranny and disastrous culpable mismanagement, rather than supervision over a range of complex micro-issues. In the remainder of this section, I will develop an account of COUNT.

We can build COUNT upon the idea that the function of electoral accountability in democracies is the same as having a formal constitution and the separation of powers, namely, that they are there to protect the rights and general liberty of those governed. This idea has been put forward by Rebecca Brown:

The structural feature of accountability for political actors can be understood . . . as a means primarily to minimize the risk of tyranny in government. . . . Accountability serves this goal . . . by allowing the people to check abuse of power at the polls if they detect a threat and wish to eradicate it.⁵⁷

On this view, the role of electoral accountability is the protection of the people from governmental tyranny, especially abuses to rights and liberties. This idea is distinct from the view that electoral accountability is in place to supervise

57 Brown, "Accountability, Liberty, and the Constitution," 536.

or manage the government to ensure they do well with complex issues such as employment and foreign affairs. The position advocated by Brown is simpler: if the government has turned tyrannical by destroying the rights of the electorate, then the electorate can get rid of them.

As with the typical view of electoral accountability, this alternative idea will also require the electorate to know about the tyranny, and to vote accordingly. With this in mind, here is an initial formulation of COUNT:

If governments were to become tyrannical, the electorate would be very likely to (1) know of the government's tyrannical actions, and (2) base their votes on that knowledge by de-electing the government.

If this account is correct, then it would give governments a strong incentive to resist becoming tyrannical, and would thus offer one explanation as to why current democracies do not seem to become tyrannies. The view is *prima facie* plausible, but there is much more we can say to develop and critique it.

First, what is it about tyrannical actions that make them problematic enough that the electorate would likely remove a tyrannical government from power? That is, why does 2 follow? Presumably, it is because these actions have *substantially bad effects* for the people being governed. This was one of the justifications put forward for representative democracy by J. S. Mill. He drew a historical contrast between “the free states of the world,” such as “the Greek cities . . . [and] the Italian republics,” with despotic oligarchies and monarchies, including “the Persian satrapies . . . [and] the feudal monarchies of Europe.” He then claimed that

no amount of disorder which exaggeration itself can pretend to have existed amidst the publicity of the free states can be compared for a moment with the contemptuous trampling upon the mass of the people which pervaded the whole life of the monarchical countries, or the disgusting individual tyranny which was of more than daily occurrence under the systems of plunder which they called fiscal arrangements, and in the secrecy of their frightful courts of justice.⁵⁸

Mill does not say precisely what he has in mind by the actions of these despotic states, but it seems to concern stealing from the people, perhaps through excessive taxation and land ownership, and unjust courts of law and unfair punitive systems. Many such problems occurred under the communist states of the twentieth century, and still continue today in many countries, where governments are corrupt, law courts are unregulated, police engage in brutality, journalists are murdered, and people are denied rights, including to free speech

58 Mill, *Considerations on Representative Government*, 68.

and education. Each of these problems, we can say, falls under the broad category of *substantially bad effects*.

Now, it may be that not every government must be tyrannical to bring about such substantially bad effects on its people. The political rulers could simply be incompetent, or disinterested in the people they govern. Many monarchies bred rulers who were so disconnected from the people they governed that they simply did not have any interest in their lives, nor any idea of how to improve them. Such ruling systems can still bring about substantially bad effects without being tyrannical, such as famines, low wages, and unemployment. So, it is not the fact of tyranny as such that would make people de-elect such a government, but the *effects* of tyranny, which can be felt in non-tyrannies as well, including monarchies and, as we will see, some democracies too.

To make COUNT work, could the bad effects only be felt by a small minority of the voters? Could it be that voters who are unaffected by the terrible harms a ruler does to other people would *not* vote to remove that ruler from power? Take the case of the Uyghur people in China. There are terribly bad effects being felt by these people. But while the group itself is extremely large—more than one million people in total—they are only a fractional minority of the total Chinese population. If the Chinese people could vote, would unaffected citizens de-elect the government to prevent further atrocities to the Uyghur people? We might naively hope that they would, and it seems fair to assume that the Uyghurs themselves would vote to remove their persecutors from power. But voters have often been found to vote egoistically, or in their own self-interest, and in particular, to prioritise economic factors affecting themselves when making their electoral decision.⁵⁹ For instance, when trade policies would harm someone's individual interests, they become less inclined to support them.⁶⁰ More generally, egoistic considerations have been found to affect party choice, preferences over trade and immigration policy, European integration, and the design of tax policy.⁶¹

How does the idea of egoistic voting square with the sociotropic evidence cited in section 2? Or, how can voters support political rulers on the basis of the impact of economic upturn for others, when, as I claim, they also take into

59 Brennan and Pettit, "Unveiling the Vote."

60 Schaffer and Spilker, "Self-Interest versus Sociotropic Considerations."

61 On party choice, see Tilley et al., "When the Pound in People's Pocket Matters." On trade preferences and immigration policy, see Scheve and Slaughter, "Labor Market Competition and Individual Preferences over Immigration Policy"; Mayda and Rodrik, "Why Are Some People (and Countries) More Protectionist Than Others?" On European integration, see Gabel, "Economic Integration and Mass Politics." On design of tax policy, see Ballard-Rosa et al., "The Structure of American Tax Policy Preferences"; Fernández-Albertos et al., "Economic Crisis, Globalization, and Partisan Bias."

account how it affects themselves? Well, both considerations are generally in play for voters, though given the circumstances, and whether the individuals tend to be more egoistic or altruistic, the considerations will carry a greater weight. For instance, Bechtel and Liesch found that “voters are about two times more sensitive to personal income gains and three times more sensitive to personal income losses than to similar changes in the nation’s average income.”⁶² So, while people do factor in the national income, as the sociotropic literature makes clear, they give greater consideration to changes in personal income—that is, they weigh more heavily egoistic than sociotropic considerations. But it is also the case that the economic impact affecting others in one’s nation will often affect oneself, especially if taxes must be increased to pay for more social support. So, it may ultimately not be possible to entirely separate egoistic and sociotropic considerations.

To return to the main point, the problem egoistic voting raises in the hypothetical case of China is that there could be close to one billion eligible voters in China, making the Uyghurs only 0.1 percent of the voting public. With so many voters external to the plight of the Uyghurs, it could well be that unaffected voters would not factor their plight into their voting decisions. One reason we could give for why the wider voting Chinese public would not reject the Chinese government because of the harms they have done to the Uyghurs is because there are current democracies where other harms committed to minority groups are not rejected at the ballot box—where people vote egoistically in spite of these problems. Consider the case of Hungary, which, under leader Viktor Orbán, restricts LGBT rights, but which has a functioning electoral system. In 2012, same-sex marriage was made illegal in Hungary, and yet Orbán was reelected with a considerable majority in both 2014 and 2018.⁶³ So, we have a case of substantially bad effects upon the people governed, but where the government is still retained. Similar issues have been flagged in recent years by scholars documenting “white ignorance” of pervasive injustices against Black people in the US.⁶⁴ Here, racially based injustices in the criminal justice system are overlooked by white people who form a majority group. So, we cannot just assume that because some rulers produce substantially bad effects, that those effects will lead to their de-election, or even to knowledge of those effects.

Now, the examples we have discussed might be different if the legislation removing LGBT rights in Hungary, or racist justice systems, affected a majority

62 Bechtel and Liesch, “Reforms and Redistribution,” 2.

63 A number of additional restrictions have been made to LGBT rights in Hungary since 2012, and in particular since 2020, but the electoral effects will not be evident until the next election. For data, see <https://www.equaldex.com/region/hungary>.

64 Mills, “White Ignorance.”

of the people—for instance, if the majority in Hungary wanted same-sex marriage, the majority in the US were Black, or the Uyghur people were a majority group. If that was the case, then we might expect that these groups would remove the government from power, so long as they had voting rights.

What it seems to take, then, to make 2 plausible, is that a government acts in such a way that those actions bring about substantially bad effects and *pervasive consequences* on those governed, in the sense that they affect either a majority, or at least a substantial minority of people in the state—enough people to alter the course of an election. For instance, if the government revoked the right to free speech, that would affect all people. Or if they added a 50 percent tax to all workers, that would affect a majority of people (only excluding the retired, unemployed, and minors). If such bad effects are felt so pervasively, why then would the people in these circumstances look to remove a government from power in an election? Presumably, because they recognize that their own interests are being severely hampered, and so would seek to preserve those interests. So, when we have substantially bad effects whose consequences are pervasive enough to affect a group large enough to change the course of an election, then we seem to have a situation in which 2 becomes plausible.

Why is this response not available to the objection from section 2, in which those who know about political affairs do not base their votes on such knowledge, but stick to entrenched political affiliations? Because in this counterfactual situation, the government is acting in ways that are far more damaging than in current democracies, and so it is reasonable to assume that voter behavior would change as well. When rights are being revoked across the board, or huge numbers of people are detained, indoctrinated, or murdered, or corrupt governments inflict massive taxation on their citizens to line their own pockets, then the voters will, it seems likely, vote differently from how they currently vote in, say, France or the UK. Political affiliations are stronger when governments are generally working in the best interests of the public, rather than inflicting the sorts of harms I have mentioned.

We can now provide a revised formulation of COUNT that focusses, not on tyrannical governments, but on the bad effects brought about by some political rulers:

If political rulers were to bring about substantially bad effects with pervasive consequences for the people they govern, then the electorate would be very likely to (1) know of the ruler's actions, and (2) base their votes on that knowledge by de-electing the rulers.

So far, we have assumed that the bad actions are likely to be known by the people who are governed—that is, that 1 would be true in these counterfactual

scenarios. We have also assumed that because of these posited negative and wide-ranging effects, the people will acknowledge the harms done to their own interests, and respond by removing their rulers from power. But to assume that 1 would be true is to assume that the bad actions are *high salience*, in the sense that they are easily noticeable to the electorate.⁶⁵ High-salience issues, like crippling tax rises, denials of rights, or widespread forced indoctrination, would seem to be known to the people. It would be unusual if such actions would go by unnoticed. However, in theory, some such events could be hidden from the view of the electorate. For instance, if the government sends its people away to a just foreign war, then even if most of those sent to the war die in battle, the people may accept the bad effects as justified, and so not de-elect the government on that basis. But if they die unnecessarily due to mismanagement, or if the war was unjust, then if the salient facts in either of these cases became known, it would likely show up at the ballot box. That would only be the case, though, if the voters knew about the mismanagement or unjustness, which could theoretically be hidden from the voters through media control and propaganda. In that kind of case, we would have actions with substantially bad effects and pervasive consequences, but which are *low salience*.

In general, though, hiding such issues from the electorate will not be realistic. The kinds of issues we have been talking about include what David Estlund calls “*primary bads*: war, famine, economic collapse, political collapse, epidemic, and genocide.”⁶⁶ Issues such as these have such substantially bad effects and are so pervasively felt that not only (a) will the people almost certainly know about them, but they will also (b) de-elect their rulers in the cases where such issues are brought about either directly by their rulers, or indirectly through political mismanagement. It would require an extremely unusual situation for an electorate to fail to know they are experiencing genocide or famine. But as I have indicated, it could be theoretically possible, though also quite unlikely today, to hide from the government’s mismanagement of an overseas war that affects many of those voters.

So, perhaps where issues would have substantial and pervasively felt consequences *and* be high salience, then in those cases, the voters would very likely know about their ruler’s actions and would likely choose to de-elect them. Although the effects of such a position on electoral accountability are invisible to social scientists for the reason that this view concerns counterfactual scenarios, we can point to current democracies to provide some evidence to justify it. For instance, as Amartya Sen has indicated, there have been no famines

65 Guerrero, “Against Elections,” 149.

66 Estlund, *Democratic Authority*, 163.

in modern democracies, and perhaps the reason why is because, if there was, the electorate would know about it, and de-elect its rulers.⁶⁷ Some political scientists have also touted the role of accountability in preventing different democratic nations from going to war with one another.⁶⁸ So, we can use the fact that these problems have not occurred as support for the invisible effects of counterfactual accountability.

According to the full theory proposed, then:

Counterfactual Theory of Electoral Accountability (COUNT): If political rulers were to bring about (a) highly salient, (b) substantially bad effects with (c) pervasive consequences for the people they govern, then the electorate would be very likely to (1) know of their ruler's actions and (2) base their votes on that knowledge by de-electing the rulers.

This account explains many of the examples considered so far. For instance, consider an unfair justice system that targets only Black people, who form a small minority. This issue would have substantially bad effects for some, but may be *low salience* because of problems that Mills points to, such as “white ignorance,” and would only have *localized* consequences. In this case, conditions a and c fail, and so, being somewhat pessimistic (or perhaps grimly realistic) about voter knowledge and behavior, we would conclude that electoral accountability would not normally succeed. The voters would need to overcome their ignorance, and choose to vote altruistically, to hold their rulers to account in this scenario. That may well happen, but to be confident of making electoral accountability work, we would normally expect all three conditions a–c to be achieved.

The account can also explain why there is, or would be, a lack of accountability when there are even high-salience issues with bad consequences, but that hit minority groups. We considered the possible case of the Uyghurs earlier, but we could also point to actual instances, such as LGBT rights in Hungary, or even Jews in Nazi Germany. In both cases, there might be widespread knowledge of the bad effects, but because they affect minority groups, we have seen elections where the political rulers gain or retain power in spite of the harms they bring. Hence, to make COUNT plausible, the harms it would need to check at the ballot box would need to be felt more widely.

The final point worth noting is that COUNT retains the two features of STAN—*electoral power* and *the electoral incentive*—but narrows the range of

67 Sen, *Development as Freedom*, 178–80, and *The Idea of Justice*, 342–45.

68 Doyle, “Kant, Liberal Legacies, and Foreign Affairs”; Moaz and Russett, “Normative and Structural Causes of Democratic Peace, 1946–1986.”

what it is to perform well or poorly to the bringing about of highly salient, substantially bad effects with pervasive consequences. In terms of electoral power, COUNT holds that voters would very likely de-elect rulers who brought about these effects, and in terms of the electoral incentive, it holds that, since voters would very likely vote in this way, political rulers are strongly incentivized to resist bringing about these effects.

5. CONCLUSIONS

This paper has argued against viewing electoral accountability in terms of voters judging the ongoing performance of their political rulers on a range of complex sociopolitical issues. That position suffers from well-known issues of voter ignorance and motivated reasoning. But that does not mean we should abandon electoral accountability. Instead, we should think of electoral accountability as providing a backstop on political rulers engaging in actions that would have terrible effects on the people they govern. To avoid the problems found in STAN, and with egoistic voting, these actions must be felt pervasively by, and be highly salient to, the voters. This account is supported by the fact that such actions do not seem to happen in current democracies, and so the effects of counterfactual electoral accountability would be invisible to political scientists.

Nevertheless, an important upshot of this proposal is that universal suffrage is still absolutely critical. People who acknowledge the limitations of the electorate, such as those outlined in section 2, often despair at the state of current democracies, and propose alternative systems with stronger electoral accountability. We noted some of these at the end of section 3. That may help to secure accountability under the standard conception (STAN). But it is not necessary for the counterfactual theory, which delivers extremely valuable controls on the actions of political rulers. Without voting rights, political rulers could bring about substantially bad effects on all of their citizens without fear of losing power. So, while electoral accountability can no doubt be improved, it still delivers an extremely positive outcome, in spite of the very real limitations voters have.

But is the counterfactual theory I have developed overly restrictive or pessimistic? Have I cornered electoral accountability to such an extent that it becomes entirely impoverished, or, indeed, meaningless? Not at all. As just noted, it prevents political rulers from bringing widespread terrible harms on the people they govern. This is a significant advantage. But it is also realistic by taking seriously the limitations with voters and their behavior. And further still, there are other forms of accountability that provide checks and balances on political power outside of voting, which were noted at the outset of section 1.

It might also be objected that voters generally take into consideration a range of factors when making their voting decision that are completely unrelated to whether or not their political rulers have harmed them—for instance, manifesto pledges, problems and achievements within one’s local constituency, or the likeliness a candidate will do well on the global stage. My account does not deny that any of these reasons are key to determining an agent’s voting decision. It simply separates these reasons out from those of *accountability*: to keep political rule to account is to check actual or potential abuses at the polls. Beyond that, voting decisions have other effects, like trying to secure a better deal for oneself on local or national policies.⁶⁹ My suggestion is that these are not about electoral accountability and should be distinguished as appropriate.

The epistemic claim that follows from COUNT is that it is not the role of the electorate to keep up to date with everything the government does, and nor would it be rational to do so. That role is fulfilled by other, mainly horizontal controls on accountability. The only epistemic obligation the counterfactual theory places on voters is to know when widespread terrible harms occur, or seem likely to occur, on themselves and others, and reject political rule that enacts, or would enact, these harms. Accountability beyond this would be supererogatory on the part of voters. But such an obligation would not be difficult to satisfy, and is consistent with the limits of the electorate. Electoral accountability may be limited in scope, but its purpose is of deep importance, and is critical for ensuring democracies function justly and effectively.⁷⁰

University of Manchester
finlay.malcolm@manchester.ac.uk

REFERENCES

- Achen, Christopher, and Larry Bartels. *Democracy for Realists: Why Elections Do Not Produce Responsive Government*. Princeton, NJ: Princeton University Press, 2016.
- Alvarez, R. Michael, and Jonathan Nagler. “Economics, Entitlements, and Social Issues: Voter Choice in the 1996 Presidential Election.” *American Journal of Political Science*, 42, no. 4 (October 1998): 1349–63.
- Arneson, Richard J. “Defending the Purely Instrumental Account of Democratic Authority.” *Journal of Political Philosophy* 11, no. 1 (March 2003): 122–32.

69 Goldman, *Knowledge in a Social World*, ch. 10.

70 Thank you to referees from this journal for their helpful feedback.

- Ballard-Rosa, Cameron, Lucy Martin, and Kenneth F. Scheve. "The Structure of American Tax Policy Preferences." *Journal of Politics* 79, no. 1 (January 2017): 1–16.
- Barnett, Zach. "Why You Should Vote to Change the Outcome." *Philosophy and Public Affairs* 48, no. 4 (Fall 2020): 422–46.
- Bechtel, Michael, and Roman Liesch. "Reforms and Redistribution: Disentangling the Egoistic and Sociotropic Origins of Voter Preferences." *Public Opinion Quarterly* 84, no. 1 (Spring 2020): 1–23.
- Bouricius, Terrill G. "Democracy through Multi-Body Sortition: Athenian Lessons for the Modern Day." *Journal of Public Deliberation* 9, no. 1 (April 2013): 1–11.
- Brennan, Geoffrey, and Philip Pettit. "Unveiling the Vote." *British Journal of Political Science* 20, no. 3 (July 1990): 311–33.
- Brennan, Jason. *Against Democracy*. Princeton, NJ: Princeton University Press, 2016.
- . "Does Public Reason Liberalism Rest on a Mistake? Democracy's Doxastic and Epistemic Problems." In *Political Epistemology*, edited by Elizabeth Edenberg and Michael Hannon, 135–52. Oxford: Oxford University Press, 2021.
- . "Does the Demographic Objection to Epistocracy Succeed?" *Res Publica* 24, no. 1 (February 2018): 53–71.
- Brown, Rebecca L. "Accountability, Liberty, and the Constitution." *Columbia Law Review*, 98, no. 3 (April 1998): 531–79.
- Caplan, Bryan. *The Myth of the Rational Voter*. Princeton, NJ: Princeton University Press, 2007.
- Christiano, Thomas. *The Rule of the Many: Fundamental Issues in Democratic Theory*. Boulder, CO: Westview Press, 1996.
- Clarke, Harold, David Sanders, Marianne C. Stewart, and Paul Whiteley. *Political Choice in Britain*. Oxford: Oxford University Press, 2004.
- Converse, Philip. "The Nature of Belief Systems in Mass Publics." In *Ideology and Discontent*, edited by David Apter, 206–61. New York: Free Press, 1964.
- Dahl, Robert. *Democracy and Its Critics*. New Haven, CT: Yale University Press, 1989.
- Darby, Derrick, and Eduardo Martinez. "Making Identities Safe for Democracy." *The Journal of Political Philosophy* 30, no. 3 (September 2022): 273–97.
- Delli-Carpini, Michael X., and Scott Keeter. *What Americans Know about Politics and Why It Matters*. New Haven, CT: Yale University Press, 1996.
- Downs, Anthony. *An Economic Theory of Democracy*. New York: Harper and Row, 1957.
- Doyle, Michael. "Kant, Liberal Legacies, and Foreign Affairs." *Philosophy and*

- Public Affairs* 12, no. 3 (Summer 1983): 205–35.
- Duffy, Bobby. *The Perils of Perception: Why We're Wrong About Nearly Everything*. London: Atlantic Books, 2018.
- Erkkilä, Tero. "Governance and Accountability—A Shift in Conceptualisation." *Public Administration Quarterly* 31, no. 1/2 (Spring/Summer 2007): 1–38.
- Estlund, David. *Democratic Authority: A Philosophical Framework*. Princeton, NJ: Princeton University Press, 2007.
- . "Why Not Epistocracy?" In *Desire, Identity, and Existence: Essays in Honor of T. M. Penner*, edited by Naomi Reshotko, 53–69. Kelowna, BC: Academic Printing and Publishing, 2003.
- Federico, Christopher. "Predicting Attitude Extremity: The Interactive Effects of Schema Development and the Need to Evaluate and Their Mediation by Evaluative Integration." *Personality and Social Psychology Bulletin* 30, no. 10 (October 2004): 1281–94.
- Feldman, Lauren, and Vincent Price. "Confusion or Enlightenment? How Exposure to Disagreement Moderates the Effects of Political Discussion and Media Use on Candidate Knowledge." *Communication Research* 35, no. 1 (February 2008): 61–87.
- Ferejohn, John. "Incumbent Performance and Electoral Control." *Public Choice* 50, no. 1/3 (1986): 5–25.
- Fernández-Albertos, José, Alexander Kuo, and Laia Barcells. "Economic Crisis, Globalization, and Partisan Bias: Evidence from Spain." *International Studies Quarterly* 57, no. 4 (December 2013): 806–16.
- Friedman, Jeffrey. "Introduction: Public Ignorance and Democratic Theory." *Critical Review* 12, no. 4 (Fall 1998): 397–411.
- Gabel, Matthew J. "Economic Integration and Mass Politics: Market Liberalization and Public Attitudes in the European Union." *American Journal of Political Science* 42, no. 3 (July 1998): 936–53.
- Gibbons, Adam F. "Is Epistocracy Irrational?" *Journal of Ethics and Social Philosophy* 21, no. 2 (February 2022): 251–81.
- Goldman, Alvin. *Knowledge in a Social World*. Oxford: Oxford University Press, 1999.
- Green, Donald, Bradley Palmquist, and Eric Schickler. *Partisan Hearts and Minds: Political Parties and the Social Identities of Voters*. New Haven, CT: Yale University Press, 2002.
- Guerrero, Alexander. "Against Elections: The Lottocratic Alternative." *Philosophy and Public Affairs* 42, no. 2 (Spring 2014): 135–78.
- Gunn, Paul. "Against Epistocracy." *Critical Review* 31, no. 1 (2019): 26–82.
- Hardin, Russell. "Democratic Epistemology and Accountability." *Social Philosophy and Policy* 17, no. 1 (Winter 2000): 110–26.

- Hannon, Michael. "Are Knowledgeable Voters Better Voters?" *Politics, Philosophy and Economics* 21, no. 1 (February 2022): 29–54.
- Hetherington, Marc J. "Putting Polarization in Perspective." *British Journal of Political Science* 39, no. 2 (April 2009): 413–48.
- Hirst, Paul. "Democracy and Governance." In *Debating Governance. Authority, Steering, and Democracy*, edited by Jon Pierre, 13–35. Oxford: Oxford University Press: 2000.
- Huber, Gregory A., Seth J. Hill, and Gabriel S. Lenz. "Sources of Bias in Retrospective Decision Making: Experimental Evidence on Voters' Limitations in Controlling Incumbents." *The American Political Science Review* 106, no. 4 (November 2012): 720–41.
- Huddy, Leonie. "From Social to Political Identity: A Critical Examination of Social Identity Theory." *Political Psychology* 22, no. 1 (March 2001): 127–56.
- Iyengar, Shanto. "Shortcuts to Political Knowledge: The Role of Selective Attention and Accessibility." In *Information and Democratic Processes*, edited by John Ferejohn and James Kuklinski, 160–85. Urbana, IL: University of Illinois Press, 1990).
- Kalmoe, Nathan P. "Uses and Abuses of Ideology in Political Psychology." *Political Psychology* 41, no. 4 (August 2020): 771–93.
- Kelly, Caroline. "Intergroup Differentiation in a Political Context." *British Journal of Social Psychology* 27, no. 4 (December 1988): 319–32.
- Key, V. O., Jr. *The Responsible Electorate Rationality in Presidential Voting, 1936–1960*. Cambridge, MA: Harvard University Press, 1966.
- Landemore, H el ene. *Democratic Reason: Politics, Collective Intelligence, and the Rule of the Many*. Princeton, NJ: Princeton University Press, 2013.
- Laver, Michael, and Kenneth Shepsle. "Government Accountability in Parliamentary Democracy." In *Democracy, Accountability, and Representation*, edited by Bernard Manin, Adam Przeworski, and Susan C. Stokes, 279–96. Cambridge: Cambridge University Press, 1999.
- Lewis-Beck, Michael, Richard Nadeau, and  ric B elanger. *French Presidential Elections*. Basingstoke, UK: Palgrave Macmillan, 2012.
- Lewis-Beck, Michael, and Mary Stegmaier. "Economic Models of Voting." In *The Oxford Handbook of Political Behavior*, edited by Russell J. Dalton and Hans-Dieter Klingemann, 518–32. Oxford: Oxford University Press, 2007.
- L pez-Guerra, Claudio. *Democracy and Disenfranchisement: The Morality of Electoral Exclusions*. Oxford: Oxford University Press, 2014.
- Malcolm, Finlay. "Epistocracy and Public Interests." *Res Publica* 28, no. 1 (March 2022): 173–92.
- Manin, Bernard, Adam Przeworski, and Susan C. Stokes. "Introduction." In *Democracy, Accountability, and Representation*, edited by Bernard Manin,

- Adam Przeworski, and Susan C. Stokes, 1–26. Cambridge: Cambridge University Press, 1999.
- Mason, Lilliana. *Uncivil Agreement: How Politics Became Our Identity*. Chicago: University of Chicago Press, 2018.
- Mayda, Anna Maria, and Dani Rodrik. “Why Are Some People (and Countries) More Protectionist Than Others?” *European Economic Review* 49, no. 5 (August 2005): 1393–1430.
- Mill, John Stuart. *Considerations on Representative Government*. 1861. Buffalo, NY: Prometheus Books, 1991.
- Mills, Charles. “White Ignorance.” In *Black Rights/White Wrongs: The Critique of Racial Liberalism*, 49–71. Oxford: Oxford University Press, 2017.
- Maoz, Zeev and Bruce Russett. “Normative and Structural Causes of Democratic Peace, 1946–1986.” *American Political Science Review* 87, no. 3 (September 1993): 624–38.
- Moncrieffe, Joy Marie. “Reconceptualising Political Accountability.” *International Political Science Review* 19, no. 4 (October 1998): 387–406.
- Mulligan, Thomas. “Plural Voting for the Twenty-First Century.” *The Philosophical Quarterly* 68, no. 271 (April 2018): 286–306.
- Nadeau, Richard, and Michael Lewis-Beck. “French Election Theory: Why Sarkozy Lost.” *Parliamentary Affairs* 66, no. 1 (January 2013): 52–68.
- Norpoth, Helmut. “Bush v. Gore: The Recount of Economic Voting.” In *Models of Voting in Presidential Elections*, edited by Herbert F. Weisberg and Clyde Wilcox, 49–64. Stanford, CA: Stanford University Press, 2004.
- O’Donnell, Guillermo. “Delegative Democracy?” *Journal of Democracy* 5, no. 1 (January 1994): 55–69.
- Sanders, David. “Party Identification, Economic Perceptions, and Voting in British General Elections, 1974–97.” *Electoral Studies* 22, no. 2 (June 2003): 239–63.
- Schaffer, Lena Maria, and Gabriele Spilker. “Self-Interest versus Sociotropic Considerations: An Information-Based Perspective to Understanding Individuals’ Trade Preferences.” *Review of International Political Economy* 26, no. 6 (2019): 1266–92.
- Scheve, Kenneth F., and Matthew J. Slaughter. “Labor Market Competition and Individual Preferences over Immigration Policy.” *Review of Economics and Statistics* 83, no. 1 (February 2001): 133–45.
- Sen, Amartya. *Development as Freedom*. New York: Alfred A. Knopf, 1999.
- . *The Idea of Justice*. Cambridge, MA: Belknap Press of Harvard University Press, 2009.
- Somin, Ilya. *Democracy and Political Ignorance: Why Smaller Government Is Smarter*. 2nd ed. Stanford, CA: Stanford University Press, 2016.

- . “Is Political Ignorance Rational?” In *The Routledge Handbook of Political Epistemology*, edited by Michael Hannon and Jeroen de Ridder, 241–52. Oxford: Routledge, 2021.
- Tilley, James, Anja Neundorff, and Sara B. Hobolt. “When the Pound in People’s Pocket Matters: How Changes to Personal Financial Circumstances Affect Party Choice.” *Journal of Politics* 80, no. 2 (April 2018): 555–69.
- Tversky, Amos, and Daniel Kahneman. “Availability: A Heuristic for Judging Frequency and Probability.” *Cognitive Psychology* 5, no. 2 (September 1973): 207–32.
- Woon, Jonathan. “Democratic Accountability and Retrospective Voting: A Laboratory Experiment.” *American Journal of Political Science* 56, no. 4 (October 2012): 913–30.

BARE STATISTICAL EVIDENCE AND THE RIGHT TO SECURITY

N. P. Adams

EVIDENCE is presented at trial to inform a judgment of law. A primary way that we evaluate evidence is its reliability, its propensity to indicate the truth or falsity of disputed facts. However, there is a type of apparently highly reliable evidence that we sometimes refuse to use at trial: statistics. This gap between our theoretical assessment of statistics and our practical treatment of them is the core of the problem of bare statistical evidence.

There are two basic strategies to addressing the problem: debunking or vindication. The rarely pursued debunking strategy aims to resolve the problem by rejecting our current practice: once we appreciate the strength of statistical evidence, we should overcome our prejudices and use it in court, as elsewhere.¹ In contrast, the vindicatory strategy aims to make sense of our refusal to use bare statistical evidence, usually while admitting that statistics are highly reliable.² This is much more commonly pursued because our refusal is widely endorsed in legal practice, theorists' armchair judgments, and a range of empirical findings.

In this article, I defend a new vindicatory strategy based on what I call the right to security. It is widely (though not universally) recognized that our refusal to use bare statistical evidence is moral in nature: finding against the defendant on such evidence would *wrong* them. The right to security explains this wrong. Understood here as a robust good in Philip Pettit's framework, security requires that someone risking harm to another's protected interests adopts a disposition of concern toward the other that controls against wrongfully harming them across an appropriate range of possible worlds.³ Adjudicating disputes via trial risks the defendant's interests, so the state must control against wrongfully

1 For example, see Schauer, *Profiles, Probabilities, and Stereotypes*, 106; Hedden and Colyvan, "Legal Probabilism"; and Papineau, "The Disvalue of Knowledge."

2 Some vindicatory strategies are only partially vindicatory, saving some uses of our practice and rejecting others. I mostly ignore the vindicatory strategy that attempts to dissolve the problem by undermining the statistics' reliability.

3 Pettit, *The Robust Demands of the Good*.

harming those interests. If the state uses bare statistical evidence in making its legal judgment, it fails to realize this control and so violates the defendant's right to security. This vindication is offered as a reconstruction of our settled practice and an explanation of the attached moral judgment. I argue that such an approach is especially apt for legal-political practices and connects to security's role in grounding judicial procedural rights more generally.

Here is the plan. In section 1, I lay out the problem of bare statistical evidence and consider some desiderata for potential solutions. In section 2, I explain Pettit's notion of a robust good and use it to elucidate security. In section 3, I apply the right to security to trial and to the use of statistics. Finally, in section 4, I show that this framework is especially apt for explaining a political practice and I emphasize its explanatory scope.

1. BARE STATISTICAL EVIDENCE

The problem of bare statistical evidence has been present in law for at least seventy-five years and has been the subject of regular debate over that period.⁴ The last decade has seen a noticeable surge in interest from philosophers, especially under an epistemic frame and partly coinciding with increased focus on the practical aspects of knowledge.⁵ In this section, I offer a brief characterization of the problem and the features that will concern us here.⁶

4 Smith v. Rapid Transit, Inc., 317 Mass. 469, 58 N.E.2d 754 (1945); Tribe, "Trial by Mathematics"; Nesson, "The Evidence or the Event?"; Thomson, "Liability and Individualized Evidence"; Wasserman, "The Morality of Statistical Proof and the Risk of Mistaken Liability"; Posner, "An Economic Approach to the Law of Evidence"; and Ho, *A Philosophy of Evidence Law*.

5 Many recent epistemically inclined discussions build on Redmayne, "Exploring the Proof Paradoxes." In the last decade, see Enoch, Spectre, and Fisher, "Statistical Evidence, Sensitivity, and the Legal Value of Knowledge"; Buchak, "Belief, Credence, and Norms"; Blome-Tillman, "Sensitivity, Causality, and Statistical Evidence in Courts of Law," "More Likely Than Not," and "Statistical Evidence, Normalcy, and the Gatecrasher Paradox"; Littlejohn, "Truth, Knowledge, and the Standard of Proof in Criminal Law"; Gardiner, "Legal Burdens of Proof and Statistical Evidence," "The Reasonable and the Relevant," "Profiling and Proof," and "Relevance and Risk"; Di Bello, "Trial by Statistics," "Proof Paradoxes and Normic Support," and "When Statistical Evidence Is Not Specific Enough"; Pardo, "Safety vs. Sensitivity"; Smith, "When Does Evidence Suffice for Conviction?" and "More on Normic Support and the Criminal Standard of Proof"; Pritchard, "Legal Risk, Legal Evidence and the Arithmetic of Criminal Justice"; Moss, "Moral Encroachment" and "Knowledge and Legal Proof"; Bolinger, "The Rational Impermissibility of Accepting (Some) Racial Generalizations"; Di Bello and O'Neill, "Profile Evidence, Fairness, and the Risks of Mistaken Convictions."

6 The problem is related to a variety of issues that have also seen recent uptake that I set aside—for example, general legal probabilism and profiling.

Consider a hypothetical, drawn from one of the original court cases. Smith's car is damaged and we can tell from the markings that a bus caused the damage, but no other identifying features of the bus are available. Smith sues Blue Bus Company and can provide statistics showing that it operates 80 percent of the buses in the city, while the remainder are operated by Red Bus. Can we find Blue Bus liable for the damage to the car merely because of the company's share of buses? In a civil suit such as this, the standard of proof is preponderance of the evidence, often glossed as "more likely than not" or as greater than 0.5 probability. The statistics seem to ground a 0.8 credence that a Blue Bus caused the damage (or, at least, greater than 0.5), and so would seem to support a finding against Blue Bus. In the actual case, the trial court did not even let the jury hear the case; the evidence was deemed insufficient, a ruling upheld by the Massachusetts Supreme Court. Armchair judgments concur and studies of hypothetical jurors have found that large majorities are unwilling to make a finding on this sort of statistical basis.⁷

Contrast this with a parallel bus case where we have an eyewitness who identifies the bus that damaged Smith's car as belonging to Blue Bus. Eyewitness testimony is notoriously imperfect; we estimate this eyewitness's reliability to be 80 percent. Can we find Blue Bus liable for the damage to the car based on the eyewitness identification? At the very least it is clear that this evidence would be sufficient in the legal sense and so presented to a jury for evaluation. More strongly, we would be comfortable if a jury found against the company on this basis and hypothetical jurors expressed a willingness to base legal findings on eyewitness testimony. There seems to be something missing from bare statistical evidence. Even when it is as reliable as other kinds of evidence, we resist using it. Even when it is more reliable than other kinds of evidence (perhaps Blue Bus operates 90 percent of the buses), we resist using it. This is true even when we judge that the evidence is sufficient to ground a belief that the defendant is guilty.⁸ In my view, the notion of use is key: in court, evidence is used against someone.

The general form of a vindication has been widely recognized for some time. Bare statistical evidence lacks a "direct" or "individualized" connection to the subject, yet this is what is required for a finding against.⁹ We can distinguish

7 Wells, "Naked Statistical Evidence of Liability." Also see Wright et al., "Factors Affecting the Use of Naked Statistical Evidence of Liability"; Niedermeier, Kerr, and Messé, "Jurors' Use of Naked Statistical Evidence"; Arkes, Shoots-Reinhard, and Mayes, "Disjunction between Probability and Verdict in Juror Decision Making."

8 Wells, "Naked Statistical Evidence of Liability," 739. Niedermeier et al. and Arkes et al. manipulate this gap between respondents' belief in guilt and willingness to find liability.

9 Thomson, "Liability and Individualized Evidence."

epistemic and practical approaches to characterizing the nature of this requirement and corresponding lack. Epistemic approaches usually argue that bare statistical evidence is not enough for knowledge; they less commonly argue that it is not enough for full belief or justified belief. Practical approaches argue that using bare statistical evidence cannot serve some practical goal of the trial system—for example, deterrence, economic efficiency, respecting rights, or upholding the sociological legitimacy of the judicial system.

Many recent approaches hybridize practical and epistemic concerns. Consider one such hybrid approach, which has some features I draw on below. David Enoch, Levi Spectre, and Talia Fisher note the parallels between modal conditions on knowledge and the problem of bare statistical evidence.¹⁰ Modal conditions capture the idea that even having a justified true belief cannot count as knowledge if that state is reached in a lucky or circumstantial way. Enoch, Spectre, and Fisher pick out the modal condition called Sensitivity: the belief must be sensitive to the truth, roughly meaning that if the proposition were false, the agent would not have the belief. A purely epistemic approach would simply take Sensitivity and apply it to the law, arguing that bare statistical evidence is insensitive and therefore cannot be knowledge (adding some story about why knowledge is required for the legal finding). Instead, Enoch, Spectre, and Fisher argue that something akin to Sensitivity applies to law because insensitive evidence distorts legal subjects' incentives. If subjects of the law know that they might be convicted on the statistical likelihood that they performed some act because of a reference class they belong to rather than on the direct evidence that follows from a particular violation, they have less incentive to avoid committing the violation. It makes sense for us to refuse to use bare statistical evidence because it is insensitive. This justification for refusing to use bare statistical evidence is practical but draws on some conceptual resources from contemporary epistemology.

Enoch, Spectre, and Fisher's approach shares a flaw with many others that appeal to epistemic resources. Our refusal to use bare statistical evidence has a specific character. The problem is not merely one of the fact finder's irrationality or the overall structure of the legal system. The problem is that using bare statistical evidence in these cases would *wrong* their subjects.¹¹ This comes through especially clearly in earlier discussions and those that take a more holistic perspective on the trial process. Judith Jarvis Thomson argues that using bare statistical evidence makes the finding a matter of luck and that this

10 Enoch, Spectre, and Fisher, "Statistical Evidence, Sensitivity, and the Legal Value of Knowledge," 220–23.

11 Gardiner, "Legal Burdens of Proof and Statistical Evidence," 187.

is “unjust.”¹² Laurence Tribe holds that it amounts to sacrificing the defendant’s rights as a person for public safety.¹³ Hock Lai Ho gives the strongest contemporary expression of this sentiment, noting the “special repugnance” of basing a finding on bare statistical evidence, saying that “we intentionally subjected the defendant to an open risk of injustice: we gamble on the facts at his expense.”¹⁴

Theories that incorporate a moral judgment of this kind have immediate advantages over theories that explain the problem by appealing to nonmoral features. The moral judgment explains why people refuse to use the evidence to convict even when they hold that it is sufficient grounds for believing that the defendant performed the act and why they offer moral explanations of this refusal. Incorporating a moral judgment also directly explains why our legal procedures should be constrained in this way, i.e., why we should potentially sacrifice the accuracy of the court by rejecting this evidence.

Epistemic approaches have a more difficult task here. The question is why courts should concern themselves with whether the court (or the jury, or a specific juror) obtains a relevant doxastic state or meets some kind of epistemic qualification. Taken to an extreme, this becomes what Enoch, Spectre, and Fisher call epistemic fetishism.¹⁵ Many constraints on court procedures, including rules of evidence, come at the cost of accuracy—ignoring probative evidence will mean more guilty people are acquitted and innocents convicted. Balancing such costs against epistemic gains seems to fetishize epistemic values.¹⁶

The same must be said of approaches that justify such rules on grounds of rationality. There are many plausible yet conflicting theories of rationality, especially once we get to the level of specific decision rules, and different theories of rationality play different explanatory and evaluative roles.¹⁷ Showing that some court rule is irrational according to one or another of these theories is incomplete at best. Such theories need an account of rationality in law considered as a political practice, without which declaring some practice irrational and therefore unjustified smacks of rationality fetishism.

A striking feature of this problem is that practice precedes theory. The object of vindication (or debunking) is a settled, public, legal practice with a moral

12 Thomson, “Liability and Individualized Evidence,” 214.

13 Tribe, “Trial by Mathematics,” 1374; for similar thoughts, see Wasserman, “The Morality of Statistical Proof and the Risk of Mistaken Liability”; and Pundik, “Statistical Evidence and Individual Litigants.”

14 Ho, *A Philosophy of Evidence Law*, 142.

15 Enoch, Spectre, and Fisher, “Statistical Evidence, Sensitivity, and the Legal Value of Knowledge,” 213.

16 Some do connect their epistemic concerns to morality, as with Littlejohn’s focus on blame.

17 Bermudez, *Decision Theory and Rationality*.

character. Approaches that make the problem simply contiguous with parallel problems of rationality or epistemology are incomplete. They must explain why we would expect this to be realized in a political practice and why it is appropriate to the specific context of law.¹⁸

With the core problem before us, we can identify two important boundaries. First, the problem seems to dissipate in the case of bare DNA evidence. DNA evidence is explicitly statistical: laboratories compare samples on comparatively few alleles and then extrapolate the statistical chances of a random match based on prevalence of allele patterns in the population.¹⁹ Cases based solely on DNA are increasingly commonplace and generally accepted, so rejecting bare statistical DNA evidence is a serious theoretical cost.²⁰ There seems to be something about DNA that distinguishes it from other kinds of bare statistical evidence, perhaps its sheer level of certainty.²¹

The second boundary is terminologically apparent: the problem of statistical evidence is a problem when it is “bare” or “naked.” The relevant contrast is the use of the same kinds of statistical evidence as corroboration. If Smith presents the rate of bus ownership alongside eyewitness testimony or other kinds of direct evidence, it seems that the statistical evidence would appropriately be taken to bolster the overall case. The exact same statistics can be appropriately used in court in a different context. Statistics are not categorically excluded; rather, it is the specific use that they are put to that is thought to be objectionable.²²

Calling this “bare” statistical evidence is only appropriate once we have narrowed our focus to a very specific role for evidence. Other, nonstatistical evidence is always present in these cases. Often, this other evidence is not contested at trial or indeed is a prerequisite on the trial occurring at all. This may put such evidence outside the remit of certain rules of evidence or burdens of demonstration but at the level of justification we need to be clear about the specific role that statistical evidence is playing when we reject its use. What is this role?

18 While certainly possible, the difficulties of such explanations are often overlooked. For example, who is the relevant agent who needs to know or believe in the guilt of the defendant? The reasonable doubt standard arose historically because the prior “moral certainty” standard, which required jurors’ personal belief in guilt, was restrictive enough to impede courts’ functions. See Roth, “Safety in Numbers,” 1161.

19 For an overview, see Roth, “Safety in Numbers.”

20 Some are willing to accept this cost; see Pritchard, “Legal Risk, Legal Evidence and the Arithmetic of Criminal Justice”; and Smith, “When Does Evidence Suffice for Conviction?”

21 Roth, “Safety in Numbers.”

22 Ross, “Rehabilitating Statistical Evidence.” This would rule out some solutions, such as a hypothetical extension of the argument from Pundik, “Predictive Evidence and Unpredictable Freedom.”

In the Blue Bus case, a specific violation is identified and there is already enough proof to show that a bus did the damage. The role of the statistical evidence is to identify which agent is responsible for the damage. Similarly, in cold-hit DNA cases, the DNA sample must be obtained in such a way as to connect it to the harm done. The sample, for example, is found on the murder weapon or on the person harmed. Without this connection between the sample and the harmful outcome, the connection that the statistics make between the sample and the individual is irrelevant. In general, such cases involve established facts of some specific harm (often a preliminary requirement of the possibility of a legal case) as well as some causal story about how the harm occurred (even if this story is quite general, e.g., we know that some human caused the harm). The statistics' role is to identify the responsible agent as grounds for a finding of culpability. It is performing this role without other evidence that we (sometimes) find objectionable. Statistics have been used by courts, apparently less objectionably, to fill in different elements of this story.²³

2. THE RIGHT TO SECURITY

My central claim is that the use of bare statistical evidence to assign legal culpability is impermissible because it violates individuals' right to security. In this section I present the right to security, applying it to the trial context in the next. My presentation proceeds in three steps: robust goods, security as a robust good, and a right to security. The idea of a robust good is the centerpiece of a conceptual framework proposed and defended by Philip Pettit.²⁴ Understanding security as a robust good is an extension proposed by Seth Lazar in a separate context.²⁵ I follow the general contours set by both Pettit and Lazar but differ in some details.

2.1. Robust Goods

Pettit proposes the notion of a robust good to better characterize how we value relationships and social life. A useful example is the robust good of friendship. Robust goods begin with more familiar thin goods. For example, friendship involves a variety of goods such as favor, care, camaraderie, and various pleasures. We realize these thin goods in the course of being friends—for example, while having a nice dinner and conversation. But friendship is more than an

23 *Summers v. Tice*, 33 Cal. 2d 80, 199 P.2d 1 (1948); *Sindell v. Abbott Laboratories*, 26 Cal.3d 588, 607 P.2d 924, 163 Cal. Rptr. 132 (1980).

24 Pettit, *The Robust Demands of the Good*.

25 Lazar, "Risky Killing."

accumulation of thin goods. Friendship is about how those goods are brought about in the context of a relationship. Fair-weather friends are not real friends because they bring about the thin goods only when it is convenient. A real friend's concern for you is realized in inconvenient circumstances as well.

This gestures toward the key role of modality. Real friends provide each other the thin goods of friendship across different possible worlds. This is a fine starting point but it is a bit odd to think that the goods in this world are dependent on the goods of other worlds. It is not quite that you are friends because of what would have happened. You are friends because the care and favor you display in your friendship arises in a specific way, one aspect of which is that you would also give each other care and favor under different circumstances. The key is what Pettit calls a disposition. When you are someone's friend, you build them into your life by structuring your attitudes and decision-making to account for their interests and your shared relationship. Friends have a disposition to favor each other. This disposition *controls* for being friends: as we encounter challenges and changing contexts, our disposition adjusts our behavior so that we still favor our friends. When you favor someone in the characteristic ways of friendship out of a disposition for doing so, you also give them the great additional good of being their friend. In general, robust goods are thin goods realized by dispositional control. For Pettit, they include romantic love, respect, and republican freedom, among many others.

One way of thinking about the modally robust provision of thin goods would be via expected value: you get a higher expected quantity of thin goods from people who would give you the thin goods across a wider range of possible worlds. But Pettit emphasizes that this is the wrong way to think about it, as should be particularly clear for friendship. Treating friendship instrumentally is to miss the point of friendship. The disposition of concern is valuable not because it makes the provision of thin goods more reliable, although it has that effect. Dispositions of concern constitute important parts of a relationship. To form a disposition of concern for someone is to build them into your life, to value their interests precisely in the sense that their interests become choiceworthy in your attitudes and decision-making, and do so by default. Due to the disposition, friends do not have to consider whether to favor on every occasion—that would be one thought too many, repeatedly.

One more detail from Pettit's complex framework is relevant for our purposes here: setting the range of possible worlds. Friends must control for providing favor over some broad range to count as friends at all. If we provide favor too narrowly, we are fair-weather friends or just acquaintances. But friendship also does not demand favor in all possible worlds: friends can take time for themselves, friendships can end appropriately, have limits on sacrifice, and so

on. The key is that friendship is a social practice that involves shared understanding. The appropriate range is something that we collectively construct in our cultural understandings of friendship, that can change over time, that individual friendships negotiate explicitly or implicitly, and so on.²⁶ The appropriate range depends on how we understand and value friendship; there is no way of determining it solely through definitions or transcendental deductions. The scope of friendship is something we understand and work out, together.

2.2. Security

With the robust goods framework in hand, we turn to security. Security vaguely involves protections against risk and harm but, in Jeremy Waldron's assessment, the concept "has not been properly analyzed" and its discussion is in a "sorry state."²⁷ Thus adapting the robust goods framework to the context of security may be fruitful. Let's see.

Security as a robust good must involve the provision of some thin good out of a disposition of concern that secures that provision across an appropriate range of possible worlds. Lazar argues that security is the

robust avoidance of pro tanto wrongful harm. To enjoy security, one must not only avoid wrongful harm in the actual world, but also do so across relevant counterfactual scenarios: those in which the victim does not get lucky. We are insecure to the extent that others make our avoidance of wrongful harm depend on luck.²⁸

The thin good, then, is a negative one: the absence of wrongful harm, where wrongfulness correlates to a violation of someone's rights. We make each other secure when we avoid wrongfully harming each other out of a disposition of concern for avoiding such harm.

The profound interdependence of life in community means our rights-protected interests are constantly under the influence of others. Giving and receiving security is therefore a pervasive feature of our social lives. Driving, for example, risks severe harm to others but when we adopt dispositions of care, we can (arguably) give others sufficient security even while we drive. Without the widespread co-provision of security, driving would be unmanageably risky. Security is particularly important in practices that harm constitutively; if we can harm without wrongfully harming, we can be secure in harm. Surgery intentionally and often grievously harms but surgery patients can have security

26 Pettit, *The Robust Demands of the Good*, 17–20.

27 Waldron, "Security as a Basic Right (After 9/11)," 210.

28 Lazar, "Risky Killing," 8.

when hospitals and surgical teams adopt dispositions of concern for avoiding negligent, unnecessary, and other kinds of wrongful harm.

Security in this sense is thus a very great good. It has instrumental benefits such as peace of mind and, of course, avoiding harm. As Lazar notes, it is also intrinsically valuable: it is essential to community standing.²⁹ My community only accepts and respects me as a member if they value me, establishing norms for controlling against wrongfully infringing my core interests. Dispositions of concern are realized not only in individual psychological states but also in shared social practices, including laws. Security is also essential for individual autonomy and flourishing, and probably also as a social basis of self-respect.³⁰ If I am under constant threat, I can hardly plan for anything other than securing my basic interests.³¹

Robust goods' modal nature is especially apt for security. If my landlord does not keep smoke detectors working in my building, I am insecure even if there is never a fire. The harms that would result from a fire are distinct from the ongoing insecurity against fire that I suffer in the absence of smoke detectors. Since the thin good is an absence, it would be bizarre to ignore security's modal character. If my building has caught fire every day for a week, the fact that it has not caught fire today does not make me secure against fire. I am desperately insecure against fire because of the closeness of worlds where my building catches fire.

Security against fire involves more than smoke detectors; it also includes, for example, fire fighters. My right to security against fire involves different sorts of responsibilities that are distributed among different levels within my community: some accrue to public institutions, some to building owners, some to neighbors. An acceptable distribution of fire-related responsibility could take different forms. The authority to determine such a distribution is one of the political powers that rests with my community precisely because control against fire requires socially settled standards, acknowledged responsibilities, settled expectations, sufficient resources, and so on.³²

2.3. *A Right to Security?*

We now turn to the question of a right to security as a robust good. The fact that robust goods are very important does not show that we have a right to

29 Lazar, "Risky Killing," 9. Also see Wolfendale, "Moral Security."

30 Rawls, *Political Liberalism*, 106.

31 Waldron, "Security as a Basic Right (After 9/11)."

32 Security, then, is helpfully paired with the harms of patterned, repetitive risk—for example as explicated in Bolinger, "The Rational Impermissibility of Accepting (Some) Racial Generalizations."

them—romantic love is important and good but we do not have a right to it, partly because nobody in particular can have a duty to be in love with us.

It may be thought that security is different because it involves rights as part of the thin good. But while it is true that we have rights not to be harmed in certain ways, security is about the separate issue of robustly controlling for avoiding such harms. Some rights are relatively minor and, while they should not be violated, may not constrain our deliberations and attitudes in the way that robust provision from a disposition of concern entails. I have property rights in my pens and perhaps a coworker occasionally takes a pen from my desk, violating those rights. He also does not concern himself with not taking my pens. Taking my pen violates my rights but it is not clear that his lack of disposition against such takings *further* wrongs me, given that my pen-related interests are quite insubstantial. I may not have a claim-right against others to provide me this kind of security.

Resolving the precise conditions under which we have a right to security would require detailed forays into various theories of rights—for example, will or interest theories. However, pursuing these options is unnecessary given our concern with the context of trial. In general, we can say that we have a right to security when our basic or core interests are at stake. Control against the setback of those interests is going to be sufficient to ground a right on either the will or interest route. Human rights, for example, protect such interests. Following Henry Shue, enjoying such fundamental rights requires more than circumstantial noninfringement; it requires the robust (institutional) provision of noninfringement.³³ Trials involve basic interests such as bodily integrity, community standing, and property; significant interests are at play by stipulation (presuming *de minimis non curat lex*). Given that, it is plausible that trial is a context where the value of security translates into a claim-right on the provision of security.

Two final comments about robust goods and security. First, robust goods are a conceptual framework for understanding the nature of some kinds of value, not an axiology *per se*. The conceptual framework is consistent with many different (though not all) positive value commitments. Grounding the right to security in the robust goods framework is relatively ecumenical.

Second, on this construal, security is a fundamentally social notion. Security is helpfully paired with the notion of vulnerability: we are vulnerable to others when they have discretion over our interests.³⁴ We cannot avoid being

33 Shue, *Basic Rights*. Shue notes that physical security is a precondition on enjoying rights; we are concerned with a broader sense.

34 Baier, "Trust and Antitrust."

vulnerable to others, both physically and socially; interdependence implies mutual vulnerability. Security is not invulnerability but living in social relations where others recognize our vulnerability and concern themselves with us by adopting dispositions that control for not exploiting our vulnerability.³⁵ Recognizing that they have discretion over our interests, they ensure that they will not use that discretion to wrongfully harm us. This is a basic relation of respect in community.³⁶

3. SECURITY AT TRIAL

Trials are intrinsically constructed to potentially result in harm. Courts' dispute-resolution function rests on their authority to change legal status, especially to impose costs. Even when the case is rejected or the result is no further change, there is the costly closing of potential avenues of relief since the state claims final authority. The intrinsic possibility of change in interests is clearest in the case of criminal trials where punishment looms, but civil proceedings also involve setback interests or harms, such as fines, settlements, and loss or transfer of legal status. It is precisely when harms are constitutively at stake that security is most relevant.

I argued above that we have a right to security whenever our fundamental interests are at stake. Since risking such interests is constitutive of courts' dispute-resolution function, individuals plainly have a right to security at trial. Avoiding harms altogether is not possible in this context, so security must concern avoiding wrongful harms. Wrongfulness could concern a variety of factors, including the permissibility of imposing harms or disproportionate harms, going far beyond the basic concern of not punishing the innocent. Trials attempt to ensure that the harms they mete out are not wrongful by testing and assessing the claims of rights and desert that the parties make. This is the point of fact finding: to figure out what happened in order to apply the law's deontic framework to draw conclusions about how people deserve to be treated. In central ways, the process of trial is a security mechanism.

Security requires a disposition of concern that controls for avoiding wrongful harm. On my view, the state bears the primary correlative duty to provide security, so we are looking for security realized via state dispositions. This is not like

35 Understanding security as a robust good therefore connects basic material rights with their social provision and recognition.

36 It is unclear where security would fall in Pettit's schema of robust goods, where he distinguishes between the general virtues, attachments, and respect. It may be that respect and security fall under a more general category of status-based goods or community membership.

friendship, which is primarily about affective and psychological states. We are concerned with the state as an institutional and group agent. Its decision-making is primarily a matter of institutional rules and procedures that define individual roles and direct the conduct of participants, combining individual inputs in such a way as to constitute collective behavior. It is also a matter of more informal elements—for example, institutional culture and implicit norms of professional conduct. The disposition of concern that controls for avoiding wrongful harm to parties must be realized institutionally, whether in standards of proof, rules of evidence, established precedent, or constitutional protections.

If we identify the duty bearers as individuals, dispositions are a matter of individual psychology. This seems misleading, not least because there is no way to guarantee that any individual juror has this disposition.³⁷ The court should function with average citizens in official roles, not require moral excellence. So while the role-specific norms of juror behavior will matter for security, the disposition of concern is primarily realized through an institutional process that individual jurors take part in.³⁸

Defendants at trial have a right to security and states have a duty to provide that security by adopting dispositions for controlling against wrongfully harming them. This grounds a host of procedures that aim to reduce the risk of wrongfully harming, including some rules of evidence such as the requirement that forensics be based on established science. But we cannot understand the refusal to use bare statistical evidence on these grounds if we understand this risk in purely probabilistic terms since the problem arises from our assessment that the statistics are reliable. Robustness avoids this trap; Pettit introduced the notion specifically in part to reject the reduction of relational goods to expected values.

As noted above in the case of friendship, the range of possible worlds can only be determined by a shared understanding of the values at stake. We should not imagine the range of control as extending to possible worlds in an even bubble of likelihood with some numerical threshold. The disposition will control for the provision of the good in some quite distant (very unlikely) worlds while not controlling for the provision in some relatively close worlds, more like an amoeba with quite different length and size appendages. Being secure against some harm is not about that harm being 1 percent likely or any other number.

37 Ho's version is particularly demanding: jurors must have empathic care; see *A Philosophy of Evidence Law*, 209–10.

38 Notice that the right to security is only violated if the agent threatening harm fails to control for not wrongfully harming. This is consistent with wrongfully harming since control is imperfect, so mistaken convictions are consistent with security. These are distinct injustices: to be convicted without committing a violation and to be convicted via a process that did not account for your rights-protected interests.

Instead, control follows the contours of our shared understanding of security, especially in its role in enabling community life and giving mutual respect.

We may still use numerical thresholds or standards like beyond a reasonable doubt for some kinds of security. The point is that security extends beyond this—for example, into protections against certain kinds of wrongful harm.³⁹ Consider the risk of a corrupt conviction, where evidence is falsified by officials to secure their preferred outcome. We do not need to measure how corrupt convictions affect the overall accuracy of the trial process to say that falsified evidence should be excluded. A smaller risk of a corrupt conviction may make us insecure while a larger risk of coincidental conviction does not.⁴⁰ The difference is that falsified evidence directly disregards security: it fails to control for avoiding wrongful harms because it circumvents the procedures that are testing for determining liability and so the wrongfulness of potential outcomes. Its impact on overall systemic accuracy is irrelevant; defendants' protected interests should not rely on those kinds of vicissitudes.

So my central claim is that using bare statistics to find a defendant culpable violates their right to security because, in an objectionable way, it fails to control for avoiding wrongfully harming their protected interests. This immediately raises the question: In what way, and why is it objectionable? In what follows I articulate the wrong using the security framework and I appeal to the resources of the ongoing literature. Ultimately, however, the security account is a reconstruction of our collective judgment as realized in a public practice. The account does not provide an independent criterion in the sense that you can assess the security features of some use of statistics and make a judgment about permissible use in the abstract. In the next section, I further defend this methodology.

As noted above, the general form of a solution to the problem of bare statistical evidence has been acknowledged from the start: statistics do not tie the individual to the culpable act in the right kind of way. The boundaries of the problem that I articulated show that the problem is about ascribing culpability. For that reason, I think previous attempts that emphasized the defendant's autonomy got quite close, tying the problem to the concern with a contemporary legalistic

39 This kind of thought explains why even a very low probability of interpersonal, community-based harm is (correctly) perceived as more threatening (to our security) than higher probabilities of impersonal harms such as natural disasters.

40 The implications of this may be quite revisionary. My claim is not that courts as currently constituted provide sufficient security on the whole. However, my claim is also not that any risk of corruption renders trials illegitimate, since that is a practical impossibility. The point is that defendants have claims to protection against different kinds of harms, many of which are not reducible to likelihood of harm. This protection also extends beyond any specific court to legal processes on the whole. Thanks to an anonymous reviewer for pushing me to clarify this.

understanding of justice.⁴¹ The law cannot treat us as mere members of potentially offending classes. It must ascribe specific violations to us. It must accuse us of acting in a specific way that its individualized deontic framework can apply to.⁴² Statistics show only that someone similarly positioned had some likelihood of acting in some way, not that the defendant acted. The absence of more direct evidence that necessarily accompanies bare statistics makes the possibility of innocence salient.⁴³ Finding against a defendant when the only basis of their culpability is bare statistical evidence wrongs them because it fails to control against the possibility that the defendant is innocent in the way that statistics necessarily leave open, and so not liable to the harms imposed. Using bare statistical evidence makes defendants wrongfully insecure.

Appealing to a security framework raises some concerns.⁴⁴ It may be unclear what distinct work the appeal to security is doing; the previous paragraph, for example, mostly consists of arguments others have made about statistical evidence presented in security language. But that is my intention. As noted, I think we have had a good grip on the main contours of the debate since the problem was recognized. People have tried to explain it by appealing to autonomy and related moral rights. My argument is that security as a robust good is a better way of filling in something we already understand in outline.

The robust goods framework emphasizes that defendants deserve protections against certain risks of harm that cannot be understood purely probabilistically. The high reliability of statistics in comparison to other kinds of evidence does not end the conversation because different kinds of risks realize different kinds of threats to our standing in community. These threats change, as I am inclined to think we see in the case of statistics. The threat to standing of being found culpable purely on a statistical basis was very different in 1945 than it is in 2021, in the age of big data and high-powered algorithms. The threat from DNA statistics is different from the threat of market-share statistics. These statistical threats share the lack of individualized identification but what that threat means in a social context can change drastically over time. Security as a robust good gives us a distinct and, in my view, better way of understanding what is at stake with statistical evidence.

Security here is not reducible to a psychological sense of security. It is mainly constituted by facts about how exposed to harm we are. But what counts

41 Tribe, "Trial by Mathematics," 1374.

42 Wasserman, "The Morality of Statistical Proof and the Risk of Mistaken Liability," 942–43.

43 Pardo and Allen, "Juridical Proof and the Best Explanation"; Gardiner, "The Reasonable and the Relevant."

44 Thanks to an anonymous reviewer for pushing me to clarify the issues raised in the next four paragraphs.

as a sufficient level of security that we have a *claim* on others to *provide* in a particular context is largely determined by our values and attitudes about which sorts of risks and harms are relevant to our lives. As above, the modal shape of security is something we work out, together. The point of security is to enable our individual and collective lives. How we understand those lives and how we choose to prioritize certain forms of life therefore directly and dramatically affect what counts as sufficient security and so what counts as wrongful infringements of the right to security.

Using a right-to-security framework does not settle the bare statistical evidence problem on its own. Security rights pull in opposite directions. The state has a duty not to harm innocent citizens by wrongfully punishing them, but it also has a duty to protect citizens from others' harmful actions.⁴⁵ The security rights of a defendant at trial can be in tension with the security rights of the public outside of the courtroom. Perhaps the public's right to security trumps defendants' right to security such that statistical evidence should be admitted. So the right-to-security framework on its own does not vindicate our current refusal to use such evidence. I am still inclined toward vindication because I think a security approach emphasizes the obligations of the state in its imposition of harm. But this openness is a virtue of the approach. As emphasized above, it gives us a way of understanding the choice we are making with our rules of evidence, presenting it as a matter of security and status in the community.

One significant advantage of appealing to security is how it handles not just the problem's core cases but its boundaries. It makes sense of the striking boundary that statistics are only problematic when used to ascribe culpability. Security is about avoiding wrongful harm. A non-culpable person should not be punished but once culpability has been established, the possibility of permissibly punishing is opened.⁴⁶ Further, although I cannot fully argue this here, culpable actions open one to a range of harmful responses. This explains why it can be appropriate to use statistics to apportion harms done by identified culpable agents but still not appropriate to use statistics to determine culpability.⁴⁷ The burdens of evidence shift around the focal point of culpability because culpability is a main determinant of permissible harm, so the demand for security takes a different form as culpability changes. Establishing culpability moves the burden away from the plaintiff; if the circumstances are such that the plaintiff cannot be expected to produce individualized evidence of the

45 Dworkin, *Taking Rights Seriously*, 12. Cf. Laudan, "The Rules of Trial, Political Morality, and the Costs of Error."

46 That is, there is a weak rights forfeiture assumption; see Wellman, "The Rights Forfeiture Theory of Punishment."

47 *Summers v. Tice*; *Sindell v. Abbott Laboratories*.

causal story, then statistical evidence can be permissible because the culpable act opened the defendant up to such possibilities. What about DNA? To explain that boundary, we must turn to the question of political judgment.

4. POLITICAL EXPLANATIONS

In this section, I connect two lingering issues: the nature of the right to security and the type of explanation it provides. I have already tried to show that the right to security is ecumenical both in axiology and at the level of rights theory. Sometimes philosophers pursue this kind of broad scope to secure a wider audience but here it is more fundamental to the type of right in question. This issue has remained mostly hidden in the discussion of bare statistical evidence up to this point. The unifying concern of this section is methodological clarification: What kind of problem is presented by bare statistical evidence and so what kind of solutions are appropriate to that problem?

Consider an appeal to the defendant's autonomy that relies on a Kantian understanding of autonomy. Even if such explanations are plausible on their own terms, they appear inappropriate to the phenomenon they are trying to explain. Vindication of our refusal is vindication of a shared, public practice. A vindictory strategy that says we were correct because we adhered to the underlying moral truth as articulated by Kant misses something important. We should be able to explain and vindicate our refusal to use bare statistical evidence on public and political terms. This does not rule out autonomy explanations, since autonomy is plausibly a core public value, but it does rule out those that rely on too narrow interpretations thereof.

This shares some features with Rawls's public reason demands, but my claim is not about legitimate constitutional law or other features of the basic structure. The object of analysis here is a long-standing and widely affirmed legal practice. Therefore, the right to security is best understood as a political right in the following quasi-Rawlsian sense: a right included in and following from the commitments of political doctrines, which applies to the public sphere and abjures deep metaphysical commitments.⁴⁸ It is thus explicable to and endorsable from major public perspectives. It is political, rather than legal, in the sense that it is not merely a feature of positive law and should be respected by law where it is not. It is moral, rather than a matter of etiquette or theoretical reason, in the broad sense that it arises from the conditions on shared living,

48 In order to understand the right this way, we do not need to buy into the entire Rawlsian public reason framework. I am also understanding the political domain to be concerned with public interest in some sense, not relying on the idea that there is a private, apolitical domain that defines the contrast.

applying to persons in virtue of those conditions, and being sufficiently weighty to constrain the pursuit of even important goods like accuracy at trial.

Explaining the problem of bare statistical evidence by appealing to the right to security understood in this moral and political sense has three virtues. First, as already articulated above, such a right is apt for the phenomenon being explained. Explanations that appeal to comprehensive moral doctrines, or indeed theories of epistemology or rationality with nonpublic metaphysical or evaluative commitments, owe additional explanations for how this connects to a political practice endorsed from many different public perspectives over many decades.

It may seem that my appeal to the idea of a robust good conflicts with my insistence that we give a political explanation. After all, robust goods are a novel framework, not widely accepted, and certainly not the subject of widespread political agreement. But these two elements of my argument are not in conflict. The notion of a robust good is a rational reconstruction and clarification of widely agreed upon values and relationships, such as friendship, freedom, and respect. New concepts help us understand these common features of our social life. My claim is that we also better understand the problem of statistical evidence by bringing this framework to bear. The terminology is unfamiliar but the justification can be put in terms that demonstrate its grounding in shared political values. The reason that we will not use bare statistical evidence in certain ways is because defendants have rights to protections of their core interests and the state has a correlative obligation to control for protecting those interests. This is why the solution to the problem has been widely acknowledged from the outset. The right to security gives a coherence and structure to this solution.

The second virtue of appealing to a political right is its consistency with many of the other explanations being offered and so its ability to use their novel resources. We may be insecure when evidence does not meet Safety or Sensitivity constraints, for example. More promisingly, I think, recent work has appealed to the notion of normality to explain our reasoning in trial contexts. Martin Smith argues that statistical evidence is flawed because it is normal for such evidence to be consistent with the innocence of the victim, in contrast to direct evidence.⁴⁹ Similarly, Georgi Gardiner argues that juries considering reasonable doubt can rule out some alternatives because they are abnormal, violating our expectations of what social life is actually like.⁵⁰ What counts as abnormal is conventional and “can be domain-specific.”⁵¹ An account that explains the political practice by appeal to a political right can keep the practice

49 Smith, “When Does Evidence Suffice for Conviction?” 1209.

50 Gardiner, “The Reasonable and the Relevant,” 305.

51 Gardiner, “Relevance and Risk,” 10.

even while alternative, deeper accounts provide new resources for understanding our practice.

Such normality notions help explain the range of worlds over which security requires controlling against wrongfully harming. This is going to be partly a matter of convention: we have simply decided that security requires protection against these kinds of normal threats but not abnormal threats.⁵² We can see this conventional stipulation already at work in the trial context. A Massachusetts laboratory technician was recently found to have been under the influence of illicit drugs, including some potentially cognitive performance-enhancing drugs such as amphetamines. This led to more than twenty-four thousand cases being dismissed.⁵³ This is not based on any specific reliability or accuracy claims but on what risks we are willing to accept. We accept the risks of participants in the court process being under the influence of a variety of licit drugs. These distinctions are not based on reliability measurements but on social conventions about what kind of risks are acceptable to impose on defendants and which are not.

This returns us finally to the issue of DNA evidence. Normality is likely connected to the degree of certainty DNA evidence apparently licenses and to courts' reliance on established science. It may simply be that DNA evidence moved into the category of normalized risks that we accept as not infringing on our security: statistics we can live with.⁵⁴ Similarly, Lewis D. Ross has recently argued that we are less hostile to using multiple sources of inculpatory statistics despite our overall evidence remaining statistical.⁵⁵ In my view, there may be no principle that distinguishes single from overlapping statistics and yet our practice may not be unjustifiably incoherent. The practice is a collective judgment about what sorts of statistics make us *wrongfully* insecure and that judgment is (largely) up to us. Our understanding of the nature of these risks and our willingness to tolerate them can change over time and be an appropriate part of our security. As Rawls notes, many political issues of this sort "have no precise answers and depend . . . on judgment. Political philosophy cannot

52 This can be fruitfully connected to the notion of standard threats from Shue, *Basic Rights*.

53 See McDonald, "24,000 Charges Tossed Because They Were Tainted by Former Amherst Lab Chemist's Misconduct."

54 This contradicts Smith's application of his normic standard to DNA evidence. My appeal to normalization does not rely on the particularities of Smith's theory. That said, I agree that, in some cases, depending on how the DNA sample is related to the crime scene, the possibility of randomly shedding DNA requires no extra explanation. But in other cold-hit cases, some explanation is required to explain how the defendant's DNA ended up, for example, on the murder weapon. Thanks to an anonymous reviewer for pushing me to clarify my appeal to normalization here.

55 Ross, "Legal Proof and Statistical Conjunctions."

formulate a precise procedure of judgment; and this should be expressly and repeatedly stated.”⁵⁶

The third virtue of appealing to a political right to security involves explanatory scope. Epistemic accounts have emphasized that they can explain both the problem of bare statistical evidence and related epistemological problems such as the lottery paradox. This scope is taken to be a theoretical virtue; solutions that apply only to the trial context may appear ad hoc. But there is explanatory scope in other directions as well. As Gardiner, Ho, and others have argued, the solution to the problem of bare statistical evidence may be connected to explanations of other aspects of the trial or legal procedure.⁵⁷

The right to security does much more than explain our refusal to use bare statistical evidence. In my view, it grounds the right to a fair trial and the class of judicial, procedural rights more generally.⁵⁸ Security is the *main* constraint on courts’ pursuit of accurate outcomes. Courts do not merely produce beliefs or evidence for some proposition, they also are used to justify actions. In virtue of their dispute-resolution function, those actions always threaten costs, sometimes up to and including social and physical death. Setting up a whole institutional apparatus that by its nature metes out costs, supporting it with the glamour and brutality of the state, and forcing individuals to submit is an incredibly risky enterprise. The right to security structures trials fundamentally, explaining a wide range of features of law and many features of evidence law, including exclusionary rules and our refusal to use bare statistical evidence. This account has explanatory scope within a range of related political practices.

The right-to-security account also has some explanatory scope into epistemological questions, albeit at greater remove. I borrow this idea from Enoch, Spectre, and Levin, who argue that while courts should not concern themselves with epistemology per se, there are underlying reasons why courts would be concerned with something like the Sensitivity condition. Unlike other political-legal explanations, the right to security intrinsically involves a modal element. As stressed, the value of security arises from the provision of the avoidance of harm out of a disposition that controls for outcomes across possible worlds.

Modality is relevant to knowledge and to courts because of underlying concerns with uncertainty, stability across time, and other constitutive elements

56 Rawls, *Lectures on the History of Political Philosophy*, 135.

57 Gardiner, “The Reasonable and the Relevant”; Ho, *A Philosophy of Evidence Law*; Schauer, *Profiles, Probabilities, and Stereotypes*.

58 Adams, “Grounding Procedural Rights.”

of human life.⁵⁹ Life under uncertainty requires ruling out some alternatives but not others, and drawing a line between the alternatives that are relevant for our actions or beliefs and those that are not. This is why knowledge is more than true belief, security more than the absence of harm, freedom more than noninterference, and honesty more than speaking truly. Modality is increasingly the tool that philosophers use to understand these important features of our lives. Robustness also incorporates agency, intention, and responsibility for modally secure outcomes, relating to the motivating concerns of virtue epistemology. It is no surprise that problems with modal provision and parallel solutions arise across domains. This kind of related but still distinct explanation treats epistemology and legal-political practices as importantly different. Even if the extra robustness of true belief adds nothing in epistemology, it would not follow that the extra robustness of culpability determinations adds nothing to our social lives.⁶⁰

University of Virginia
npadams@virginia.edu

REFERENCES

- Adams, N. P. "Grounding Procedural Rights." *Legal Theory* 25, no. 1 (March 2019): 3–25.
- Arkes, Hal R., Brittany Shoots-Reinhard, and Ryan S. Mayes. "Disjunction between Probability and Verdict in Juror Decision Making." *Journal of Behavioral Decision Making* 25, no. 3 (July 2012): 276–94.
- Baier, Annette. "Trust and Antitrust." *Ethics* 96, no. 2 (January 1986): 231–60.
- Bermudez, José Luis. *Decision Theory and Rationality*. Oxford: Oxford University Press, 2009.
- Blome-Tillman, Michael. "'More Likely Than Not'—Knowledge First and the Role of Statistical Evidence in Courts of Law." In *Knowledge First: Approaches in Epistemology and Mind*, edited by J. Adam Carter, Emma C. Gordon, and Benjamin W. Jarvis, 278–92. Oxford: Oxford University Press, 2017.
- . "Sensitivity, Causality, and Statistical Evidence in Courts of Law." *Thought* 4, no. 2 (June 2015): 102–12.
- . "Statistical Evidence, Normalcy, and the Gatecrasher Paradox." *Mind*

59 On the desirability of durability in knowledge, see Lord, "Epistemic Reasons, Evidence, and Defeaters."

60 Papineau, "The Disvalue of Knowledge."

- 129, no. 514 (April 2020): 563–78.
- Bolinger, Renée Jorgensen. “The Rational Impermissibility of Accepting (Some) Racial Generalizations.” *Synthese* 197, no. 6 (June 2020): 2415–31.
- Buchak, Lara. “Belief, Credence, and Norms.” *Philosophical Studies* 169, no. 2 (June 2014): 285–311.
- Di Bello, Marcello. “Proof Paradoxes and Normic Support: Socializing or Relativizing?” *Mind* 129, no. 516 (October 2020): 1269–85.
- . “Trial by Statistics: Is a High Probability of Guilt Enough to Convict?” *Mind* 128, no. 512 (October 2019): 1045–84.
- . “When Statistical Evidence Is Not Specific Enough.” *Synthese* 199, nos. 5–6 (December 2021): 12251–69.
- Di Bello, Marcello, and Collin O’Neill. “Profile Evidence, Fairness, and the Risks of Mistaken Convictions.” *Ethics* 130, no. 2 (January 2020): 147–78.
- Dworkin, Ronald M. *Taking Rights Seriously*. Cambridge, MA: Harvard University Press, 1977.
- Enoch, David, Levi Spectre, and Talia Fisher. “Statistical Evidence, Sensitivity, and the Legal Value of Knowledge.” *Philosophy and Public Affairs* 40, no. 3 (Summer 2012): 197–224.
- Gardiner, Georgi. “Legal Burdens of Proof and Statistical Evidence.” In *The Routledge Handbook of Applied Epistemology*, edited by David Coady and James Chase, 179–95. London: Routledge, 2018.
- . “Profiling and Proof: Are Statistics Safe?” *Philosophy* 95, no. 2 (April 2020): 161–83.
- . “The Reasonable and the Relevant: Legal Standards of Proof,” *Philosophy and Public Affairs* 47, no. 3 (Summer 2019): 288–318.
- . “Relevance and Risk: How the Relevant Alternatives Framework Models the Epistemology of Risk.” *Synthese* 199, no. 1–2 (December 2021): 481–511.
- Hedden, Brian, and Mark Colyvan, “Legal Probabilism: A Qualified Defence.” *Journal of Political Philosophy* 27, no. 4 (December 2019): 448–68.
- Ho, Hock Lai. *A Philosophy of Evidence Law*. Oxford: Oxford University Press, 2008.
- Laudan, Larry. “The Rules of Trial, Political Morality, and the Costs of Error: Or, Is Proof Beyond a Reasonable Doubt Doing More Harm than Good?” In *Oxford Studies in Philosophy of Law*, vol. 1, edited by Leslie Green and Brian Leiter, 195–227. Oxford: Oxford University Press, 2011.
- Lazar, Seth. “Risky Killing: How Risks Worsen Violations of Objective Rights.” *Journal of Moral Philosophy* 16, no. 1 (February 2019): 1–26.
- Littlejohn, Clayton. “Truth, Knowledge, and the Standard of Proof in Criminal Law.” *Synthese* 197, no. 12 (December 2020): 5253–86.

- Lord, Errol. "Epistemic Reasons, Evidence, and Defeaters." In *The Oxford Handbook of Reasons and Normativity*, edited by Daniel Star, 600–630. Oxford: Oxford University Press, 2018.
- McDonald, Danny. "24,000 Charges Tossed Because They Were Tainted by Former Amherst Lab Chemist's Misconduct." *Boston Globe*, September 25, 2019. <https://www.bostonglobe.com/metro/2019/09/25/charges-tossed-because-they-were-tainted-former-amherst-lab-chemist-misconduct/MUPgdHeLy8bdrzlj5KGtvIN/story.html>.
- Moss, Sarah. "Knowledge and Legal Proof." In *Oxford Studies in Epistemology*, vol. 7, edited by Tamar Szabó Gendler, John Hawthorne, and Julianne Chung, 176–213. Oxford: Oxford University Press, 2023.
- . "Moral Encroachment." *Proceedings of the Aristotelian Society* 118, no. 2 (July 2018): 177–205.
- Nesson, Charles. "The Evidence or the Event? On Judicial Proof and the Acceptability of Verdicts." *Harvard Law Review* 98, no. 7 (May 1985): 1357–92.
- Niedermeier, Keith E., Norbert L. Kerr, and Lawrence A. Messé. "Jurors' Use of Naked Statistical Evidence: Exploring Bases and Implications of the Wells Effect." *Journal of Personality and Social Psychology* 76, no. 4 (April 1999): 533–42.
- Papineau, David. "The Disvalue of Knowledge." *Synthese* 198, no. 6 (June 2021): 5311–32.
- Pardo, Michael S. "Safety vs. Sensitivity: Possible Worlds and the Law of Evidence." *Legal Theory* 24, no. 1 (March 2018): 50–75.
- Pardo Michael S., and Ronald J. Allen. "Juridical Proof and the Best Explanation." *Law and Philosophy* 27, no. 3 (May 2008): 223–68.
- Pettit, Philip. *The Robust Demands of the Good*. Oxford: Oxford University Press, 2015.
- Posner, Richard A. "An Economic Approach to the Law of Evidence." *Stanford Law Review* 51, no. 6 (July 1998): 1477–546.
- Pritchard, Duncan. "Legal Risk, Legal Evidence and the Arithmetic of Criminal Justice." *Jurisprudence* 9, no. 1 (2018): 108–19.
- Pundik, Amit. "Predictive Evidence and Unpredictable Freedom." *Oxford Journal of Legal Studies* 40, no. 2 (Summer 2020): 238–64.
- . "Statistical Evidence and Individual Litigants: A Reconsideration of Wasserman's Argument from Autonomy." *International Journal of Evidence and Proof* 12, no. 4 (November 2008): 303–24.
- Rawls, John. *Lectures on the History of Political Philosophy*. Edited by Samuel Freeman. Cambridge, MA: Harvard University Press, 2007.
- . *Political Liberalism*. New York: Columbia University Press, 1999.
- Redmayne, Mike. "Exploring the Proof Paradoxes." *Legal Theory* 14, no. 4

- (December 2008): 281–309.
- Ross, Lewis D. “Legal Proof and Statistical Conjunctions.” *Philosophical Studies* 178, no. 6 (June 2021): 2021–41.
- . “Rehabilitating Statistical Evidence.” *Philosophy and Phenomenological Research* 102, no. 1 (January 2021): 3–23.
- Roth, Andrea. “Safety in Numbers: Deciding When DNA Alone Is Enough to Convict.” *New York University Law Review* 85 (October 2010): 1130–85.
- Schauer, Frederick F. *Profiles, Probabilities, and Stereotypes*. Cambridge, MA: Harvard University Press, 2003.
- Shue, Henry. *Basic Rights*. 2nd ed. Princeton: Princeton University Press, 1997.
- Smith, Martin. “More on Normic Support and the Criminal Standard of Proof.” *Mind* 130, no. 519 (July 2021): 943–60.
- . “When Does Evidence Suffice for Conviction?” *Mind* 127, no. 501 (October 2018): 1193–218.
- Thomson, Judith Jarvis. “Liability and Individualized Evidence.” *Law and Contemporary Problems* 49 (Summer 1986): 199–219.
- Tribe, Laurence H. “Trial by Mathematics: Precision and Ritual in the Legal Process.” *Harvard Law Review* 84, no. 6 (April 1971): 1329–93.
- Waldron, Jeremy. “Security as a Basic Right (After 9/11).” In *Global Basic Rights*, edited by Charles R. Beitz and Robert E. Goodin, 207–26. Oxford: Oxford University Press, 2009.
- Wasserman, David T. “The Morality of Statistical Proof and the Risk of Mistaken Liability.” *Cardozo Law Review* 13 (1991): 935–76.
- Wellman, Christopher Heath. “The Rights Forfeiture Theory of Punishment.” *Ethics* 122, no. 2 (January 2012): 371–93.
- Wells, Gary L. “Naked Statistical Evidence of Liability: Is Subjective Probability Enough?” *Journal of Personality and Social Psychology* 62, no. 5 (May 1992): 739–52.
- Wolfendale, Jessica. “Moral Security.” *Journal of Political Philosophy* 25, no. 2 (June 2017): 238–55.
- Wright, Edward F., Lora MacEachern, Elaine Stoffer, and Nancy MacDonald. “Factors Affecting the Use of Naked Statistical Evidence of Liability.” *Journal of Social Psychology* 136, no. 6 (1996): 677–88.