

JOURNAL *of* ETHICS
& SOCIAL PHILOSOPHY

VOLUME XIX · NUMBER 2

February 2021

ARTICLES

- 111 Contractualism, Complaints, and Risk
Bastian Steuwer
- 148 Supersession, Reparations, and Restitution
Caleb Harrison
- 168 Resisting Wrongful Explanations
Arianne Shahvisi

DISCUSSION

- 192 Political Disagreement and Minimal Epistocracy
Adam F. Gibbons

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

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

Editor

Mark Schroeder

Associate Editors

Dale Dorsey

James Dreier

Julia Driver

Anca Gheaus

Errol Lord

Tristram McPherson

Colleen Murphy

Hille Paakkunainen

Discussion Notes Editor

Kimberley Brownlee

Editorial Board

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

Managing Editor

Jesse Wilson

Copyeditor

Susan Wampler

Typesetting

Matthew Silverstein

CONTRACTUALISM, COMPLAINTS, AND RISK

Bastian Steuwer

ONE OF the most prominent and forceful objections against utilitarianism is that it fails to respect the separateness of persons. Utilitarianism aggregates all benefits and burdens of an action in order to decide whether or not the action is permissible. It seems as though the utilitarian treats all benefits and burdens an action produces as if they were the benefits and burdens of one entity or one system of ends. In doing so, utilitarianism fails to respect the separateness of persons as individuals and as systems of ends of their own.¹

In response to utilitarianism's failure to respect the separateness of persons, nonconsequentialists have proposed conceptions of morality that are based on the competing claims or complaints that individuals can raise. Placing the commitment to individual claims or complaints at the heart of morality seems a promising route to ensure respect for the separateness of persons. The most systematic of these proposals is contractualism as developed by T. M. Scanlon. Scanlon argues that an act's rightness or wrongness depends on its justifiability to each person. As a test for justifiability, Scanlon proposes that the permissibility of an act depends on whether it follows from a principle that no one can reasonably reject. An act is permissible only when no one can reasonably reject a principle that entails the permissibility of that act. One natural idea is that the individual with the largest complaint has most reason to reject a principle. It then appears that a principle can be reasonably rejected only when the largest complaint is larger than the complaint anyone else could bring forward against any alternative principle.² Recently Scanlonian contractualism has received scrutiny for the way it deals with cases where risks, rather than certainties of harm and benefit, are at stake.³ My discussion in this article will focus on Scan-

1 See Gauthier, *Practical Reasoning*, 125–26; Nagel, *The Possibility of Altruism*, 133–40; Rawls, *A Theory of Justice*, 23–26; Nozick, *Anarchy, State, and Utopia*, 32–33; Nagel, *Mortal Questions*, ch. 8; and Nagel, *Equality and Partiality*, chs. 4–8.

2 See Scanlon, “Contractualism and Utilitarianism,” and *What We Owe to Each Other*, ch. 5.

3 See Reibetanz, “Contractualism and Aggregation”; Ashford, “The Demandingness of Scanlon’s Contractualism”; Lenman, “Contractualism and Risk Imposition”; Fried, “Can

lonian contractualism, but my conclusions may apply more widely to any moral theory that places the idea of justifiability and individual complaints or competing claims at the heart of morality.

The debate around contractualism and risk is typically framed as a debate between two opposing views. *Ex ante* contractualism is concerned with prospects while *ex post* contractualism is concerned with outcomes.⁴ I believe that this framing is unhelpful. What can it mean to say that a theory of risk imposition is concerned with outcomes when it is designed to provide guidance in cases where we are uncertain about the outcome? With the help of a sequence of thought experiments from Michael Otsuka, I provide a more helpful way of understanding what is at stake between different contractualist approaches to risk (section 1).⁵ In addition, the sequence allows me to propose a new view on contractualism and risk, which I call *objective ex ante contractualism* because of the special importance it gives to objective, as opposed to epistemic, probability. My version of contractualism focuses on the complaints of would-be victims whose fates are already determined. After discussing the sequence, I will show that a natural extension of the sequence highlights that two conditions that *ex post* contractualism should ideally fulfill are inconsistent with one another (section 2). In section 3, I will present the defense of my objective *ex ante* view by arguing that it provides us with the best model of the key contractualist idea of acting in ways that are justifiable to each. Section 4 responds to objections.

1. OTSUKA'S SEQUENCE

Dust: A comet is en route to the midwestern United States carrying a pathogen that will soon lead to millions of people being infected and dying. The government is briefed on two alternative ways of containing the pathogen. The first option has the side effect that a different hazard will

Contractualism Save Us from Aggregation?"; James, "Contractualism's (Not So) Slippery Slope"; Fleurbaey and Voorhoeve, "Decide As You Would with Full Information!"; Frick, "Uncertainty and Justifiability to Each Person"; Scanlon, "Reply to Zofia Stemplowska"; John, "Risk, Contractualism, and Rose's 'Prevention Paradox'"; Frick, "Contractualism and Social Risk"; Kumar, "Risking and Wronging"; Otsuka, "Risking Life and Limb"; Horton, "Aggregation, Complaints, and Risk"; and R uger, "On *Ex Ante* Contractualism."

- 4 For the former, see James, "Contractualism's (Not So) Slippery Slope"; John, "Risk, Contractualism, and Rose's 'Prevention Paradox'"; Kumar, "Risking and Wronging"; and Frick, "Contractualism and Social Risk." For the latter, see Fleurbaey and Voorhoeve, "Decide As You Would with Full Information!"; Otsuka, "Risking Life and Limb"; and R uger, "On *Ex Ante* Contractualism."
- 5 Otsuka, "Risking Life and Limb," 77–88.

be released over Florida. It is known that it would cause Bob Johnson, a resident of Boca Raton, to lose one leg. Unfortunately, Bob Johnson cannot be evacuated in time. The second alternative has the side effect that the hazard will have to be released in a dust cloud over California. Each of forty million Californians faces a small risk of death, and it is known that exactly one Californian will die. The Californian who will die has a genetic predisposition that will cause his or her death upon being subjected to the dust.

Intuitively, the right course of action here would be to release the hazard over Florida and cause Bob Johnson to lose a leg. But it appears that contractualism struggles to explain this intuitive answer. Bob Johnson's complaint against choosing to release the hazard is not discounted. It is certain that he will suffer. The complaints of the Californians, however, should be discounted. For each of the forty million Californians, the likelihood of being the one who dies is only a one in forty million. Although death is terrible, a one in forty million chance of death is not altogether that terrible. We often incur similar risks when crossing the road, cooking with gas, or swimming in the ocean. The complaint against the imposition of the risk of death would suddenly be a rather trivial moral complaint. How can such a trivial moral complaint outweigh Bob's quite serious complaint of losing his leg?

One way for contractualism to accommodate the case is by pointing out that all the complaints combined add up to something significant: a complaint of the magnitude of certain death. But this response leads to highly counterintuitive results in other cases.

Transmitter Room: Jones, a worker in a TV transmitter room, has had an accident. He is now lying on the floor and suffering extremely painful electric shocks. There is only one way to save Jones, namely by interrupting the current transmission signal for about fifteen minutes. This in turn will cause millions of viewers who want to see the football World Cup match that is in progress to be upset.⁶

If we add up the complaints due to inconvenience and upset of all the millions of viewers, it seems that they will outweigh Jones's complaint against being subject to pain. But here it is clear that we should not let Jones suffer for the relatively mild loss of missing fifteen minutes of a football match. We should not aggregate morally trivial complaints so that they outweigh serious moral complaints of single individuals.

6 Scanlon, *What We Owe to Each Other*, 235.

Otsuka, in his discussion of Dust, resists this solution and instead points to a different feature of the case. Unlike in Jones's case, in Dust there is one person who will experience grave harm. The aggregated complaints add up to the real-life predicament of one person. We do not need to imagine a social entity that experiences the harms of dying, but there is an individual made out of flesh and blood who will die. It is merely a fact concerning our informational limitations that prevents us from identifying that person in the same manner we were able to identify Bob Johnson. Yet we can still say something about the individual who is going to die. The person who is going to die is "the Californian with the genetic predisposition." The complaint of the Californian with the genetic predisposition is nondiscounted. Her (or his) complaint would outweigh Bob Johnson's complaint.

Now is the complaint of the Californian with the genetic predisposition a complaint *ex ante* or *ex post*? *Ex post* contractualism can account for this complaint. We know that the result of the action will be one person dying. Since the outcome distribution of the action is already known to us, an *ex post* contractualist can peek ahead, anticipate this distribution, and assign complaints to those affected by it.

But can *ex ante* contractualism? I think it can. The Californian with the genetic predisposition is a person with a determinate identity when we make the decision. Regardless of what happens and regardless of our action, the Californian with the genetic predisposition will always be the same person. If we limit our attention to only those possible worlds that are possible outcomes of our action, then we can say that "the Californian with the genetic predisposition" rigidly designates over this restricted domain of discourse. Since only those possible worlds that constitute possible outcomes of our actions are of interest to us, I will simply refer to such descriptions as "rigid designators."⁷ Releasing the hazard over California will impose the certainty of death on this existing person with a determinate identity. From the *ex ante* perspective, the Californian with the genetic predisposition can object to the imposition of a 100 percent risk of death. We do not need to appeal to the outcome of the action *ex post* to make this claim.

This means that our understanding of *ex ante* contractualism should be broader. The classical version of *ex ante* contractualism focuses on the risks as faced by individuals with proper names, or otherwise identifiable individuals. But not all

7 This definition also includes an element of temporality in the *ex ante/ex post* distinction. The possible worlds that are possible outcomes of the action are those possible worlds that coincide in their history until the point of action. Rigid designators are descriptions that refer to information that is contained in the shared history. Nonrigid designators are descriptions that refer to information about the future where the possible worlds no longer coincide.

versions of *ex ante* contractualism focus on these risks. The version of *ex ante* contractualism I defend focuses on the complaints that rigidly designated individuals can raise. The two forms of *ex ante* contractualism differ thereby in whose complaints they focus on. This in turn is linked to a distinction between two kinds of risk: epistemic risks (credences) and objective risks (chances).⁸ The distinction I am relying on here classifies some probability functions as expressing our uncertain degrees of belief or confidence about the world. These are epistemic probability functions, also called credence functions. By contrast, objective probability functions express a mind-independent idea of probability. The objective probability function, a chance function, reflects information about the world and not about our knowledge of the world. If there are nontrivial objective probabilities, then there are truly “chancy” events. While there are various theories on what chances are, the differences between them are not important for my arguments.⁹ What I rely on is solely the contrast between chances and credences.

In Dust, we only have epistemic probabilities for the risks that each identifiable Californian faces. However, we can give objective probabilities for the risk that the Californian with the genetic predisposition faces. This suggests an important link between the question of whose complaints we are interested in and what kind of risk we are interested in. By focusing on rigidly designated individuals, objective *ex ante* contractualism gives primacy to objective risk assessments over epistemic risk assessments. Objective *ex ante* contractualism holds that in a case like Dust where the uncertainty is merely a matter of failing to identify the victim, we should choose descriptions that reveal the objective risks that individuals are facing. This is the “objective” component in objective *ex ante* contractualism.¹⁰

8 I follow here the orthodox tradition in the philosophy of probability dating back to Rudolf Carnap, who distinguished between two concepts of probability (frequentist and evidential), which are examples of the broader approaches of chance and credence. See Carnap, “The Two Concepts of Probability,” 516–25; Eagle, “Chance versus Randomness,” sec. 1; and Hájek, “Interpretations of Probability,” sec. 3.

9 The most common approaches are frequentism, propensity views, and best systems approaches. In addition, some philosophers embrace a “no theory” approach to chances according to which objective probabilities are not reducible to anything else like frequencies or propensities. For an overview, see Hájek, “Interpretations of Probability”; for the no theory approach, see Sober, “Evolutionary Theory and the Reality of Macro-Probabilities,” 148–54. Actual frequentist views are an exception to my claim that my use of objective chance is neutral between the different theories of chance. According to actual frequentist views, objective probabilities only refer to actually occurring frequencies. Under such a view, objective probabilities only represent statistical facts about reference groups and have no obvious moral significance.

10 Importantly, the two kinds of risks are linked in a manner that should guard us from identi-

Let me move on to the next case in the sequence:

Wheel: The case is structurally similar to *Dust*. Again, we have a comet en route and a disaster about to occur. Again, one of our options is to release the hazard over Florida and cause Bob Johnson's loss of a leg. But now our second option changes. As a side effect of averting the disaster, each Californian will be placed under a gigantic roulette wheel in the sky. The wheel will spin indeterministically and release a roulette ball that will kill exactly one person.

Otsuka reports his intuitive judgment that in *Wheel*, as in *Dust*, we should still prefer to release the hazard over Florida, causing the loss of Bob Johnson's leg. But here we cannot rely anymore on the description of "the Californian who is genetically predisposed." Instead, we would need to rely on a description like "the Californian who would be hit by the roulette ball" or "the Californian who would be most harmed by the decision." These descriptions are nonrigid designators since different persons may die due to the falling ball. While the complaints of rigidly designated individuals have to be discounted, the complaints of nonrigidly designated individuals do not. The probability of someone being harmed by the wheel is one. We can peek ahead and assign a complaint to that person. We may think that such statistical persons are still actual persons worthy of respect and with claims that ought to be taken into consideration.¹¹

This cannot be reconciled with the *ex ante* perspective. The complaint of the Californian most harmed by the decision is not a complaint of any person with a determinate identity prior to the action. There is no token individual for whom it is true that she has imposed on her a 100 percent risk of death. Accordingly, my objective *ex ante* view holds that releasing the hazard over California is permissible in *Wheel*.

Anticipating the strongest complaint *ex post* is easy in a case like *Wheel*. We know for certain how the benefits and burdens will be distributed in the outcome. We only lack information about who will be in which position. I now move on to a case where certainty about the resulting distribution is absent.

fying epistemic or objective *ex ante* contractualism exclusively with one kind of risk. Whenever we have an objective probability for a given event (such as Charlotte Williams's being harmed), we should adjust our credence (i.e., our epistemic probability) to match the objective probability. The next case in the sequence is an example of this. This widely accepted claim is an implication of David Lewis's Principal Principle. See Lewis, "A Subjectivist's Guide to Objective Chance."

11 See Daniels, "Can There Be Moral Force to Favoring an Identified over a Statistical Life?" 116; and Otsuka, "Risking Life and Limb," 85–86.

Guns: In this case, we have the option to shoot down the comet with an automated weapons system. Unfortunately, the system also has guns in the sky pointed at each Californian. Each gun is operated by an indeterministic randomizer. The chance for each gun to fire and kill the person it is aimed at is one in forty million. The guns, and thus the risks each gun imposes, operate independently of one another.

The objective risk for each Californian is the same as in *Wheel*, one in forty million. Any assessment of rigid designators that relies on objective risks will be the same between *Wheel* and *Guns*. However, the assessment for nonrigid designators like “the Californian who will be most harmed” changes. Here we move away from certainty about the distribution that will come about and introduce risk as well. There is a 63 percent chance that at least one Californian will die, a 26 percent chance that at least two Californians will die, an 8 percent chance that at least three will die, and so on. What should *ex post* contractualists say about a case like this?

One answer is that *Guns* highlights the limits of *ex post* contractualism. Under this version of *ex post* contractualism, we should draw a distinction between two types of cases. In some cases, like *Dust* or *Wheel*, we know that the risk imposition will lead to harm while in *Guns* the harm is not guaranteed. Anticipating the complaint of the eventual victim is permitted in *Dust* and *Wheel* but not permitted in *Guns* according to this view. Since we do not know for certain that someone will be harmed, we cannot anticipate this complaint.¹²

The problem with this version of *ex post* contractualism is that it relies on a distinction between risky cases that is morally dubious.¹³ Cases with guaranteed harms can easily be transformed into cases without guaranteed harm without changing anything of moral relevance. Take the example of a coin flip with inversely correlated harms and benefits. If the coin lands heads, *A* benefits and *B* is harmed. If the coin lands tails, *A* is harmed and *B* benefits. This is a case of guaranteed harm. *Ex post* contractualism would sometimes rule out this kind of risk even if it is in the antecedent interests of both *A* and *B*. But what if the coin lands on the edge? This would be a freak accident but is nonetheless a possibility. Let us assume that no one will be harmed if the coin lands on the edge. The case is now one without guaranteed harm. If we are not allowed to anticipate any complaint *ex post*, we should do what is in the antecedent interests of both. Sim-

12 Sophia Reibetanz Moreau defends such a view (“Contractualism and Aggregation,” 302–4). Victor Tadros, in a different context, argues that these two kinds of risks are distinct (“Controlling Risk,” 148–54).

13 Otsuka makes a similar argument in “Risking Life and Limb,” 88.

ilar things hold for a version of Wheel. If we allow only a tiny chance that no one will be harmed, the restricted *ex post* view would allow the risk imposition since this case would no longer involve guaranteed harm. Yet if we are convinced that imposing the risk in Wheel is impermissible, it should be impermissible even in this varied scenario. We need a different version of *ex post* contractualism.

Earlier I mentioned that in Guns we only know facts about what distributions of harms are to occur with which likelihood. For example, we know that the chance that at least one Californian will die is about 63 percent. One possibility for *ex post* contractualists is to translate these facts about distributions into complaints. Imagine we specify a ranking of all persons affected. The main ranking criterion is how strong each individual complaint against the action is. In cases where individuals are equally affected, we need other tiebreaking criteria. This way we can assign each individual a unique place in the ranking. Then we repeat this for all possible outcomes. We can now construct fictional characters or “statistical persons” based on these rankings. “The worst-off Californian” refers to the first-ranked person in each of the outcomes. “The second worst-off Californian” refers to the second-ranked person and so on. In cases of objective risk imposition, these designators are nonrigid since they refer to different individuals in different possible worlds. This construction allows us to assign unique complaints to individuals instead of being limited to talking about distributions of harms. Speaking of the complaints of nonrigidly designated persons brings the *ex post* perspective closer to the theoretical core of contractualism. It can provide a model of justifiability to each that an analysis of different distributions of harms cannot offer. *Ex post* contractualists should therefore accept the following principle.

Ex Post Discounting: When assessing the complaints of individuals, we should discount the complaints of nonrigidly designated individuals such as the worst-off, the second worst-off, and so on by the improbability of harm.

As mentioned, in our case of Guns, this means that the complaint of the worst-off Californian is a discounted complaint against death rather than a nondiscounted complaint as in Wheel. The complaint is discounted by the 37 percent probability that the worst-off will not be harmed. But now the second worst-off Californian has a discounted complaint as well, as has the third worst-off, and so on. Should this difference matter?

Victor Tadros believes that it should. He gives the following argument based on an example that is a simpler version of the contrast between Wheel and Guns.¹⁴ Imagine we have two options. If we choose the first option, then it is

14 Tadros, “Controlling Risk,” 153–54.

guaranteed that one and exactly one person will die. If we choose the second option, then there is only a 75 percent chance that someone will die but there is also a 25 percent chance that two persons will die. Whatever we do, the risks to each rigidly designated individual are the same. Under one view, the options are equally choiceworthy. If we choose the second option, there is a possibility that no one will die, but this is balanced by the possibility that more than one will die. Tadros, however, argues that we should choose the second option because we have special reason to prevent a situation where harm will definitely occur. We should not regard the loss of two lives as twice as hard to justify than the loss of one life. This is because the two lives are separate and not part of one aggregate that suffers a double loss.

But it is hard to see why the separateness of persons should give us a special reason to avert definite harm. Tadros's argument implies that we have less reason to prevent an additional second death. Attaching special significance to the fact that harm will occur means attaching special significance to an isolated harm as opposed to a harm that occurs alongside many other harms. Yet deaths should have the same disvalue regardless of whether they are part of an action in which only one, two, or many people die. The death is just as tragic and severe for this person regardless of how many other people have died.¹⁵ Respect for each individual and for her separateness would seem to indicate that we should treat her loss by itself and not accord it more or less moral force because of the number of other people who have died. If this is true, then we should treat both options in Tadros's example as equally choiceworthy. The *ex post* contractualist should then regard Guns and Wheel as equally hard to justify. What should matter to us is the expected number of lives lost and not how the risk is distributed across nonrigid designators. This gives us a second principle that *ex post* contractualism would want to fulfill.

Equal Treatment for Equal Statistical Loss: We should treat cases alike if in both cases there is the same expectation of statistical loss and the only difference is the distribution of possible losses across possible outcomes.

2. A PROBLEM FOR EX POST CONTRACTUALISM

Consider:

Gas: We receive yet another option to prevent the catastrophe. This time we have to release a gas in the air that will travel to California. Scientists tell us that there is the possibility that in California the gas will react by

15 See also Otsuka, "Risking Life and Limb," 88–92.

means of an indeterministic process with another substance and become toxic. If that happens, all Californians will die. However, they assure us that this is very unlikely. The objective probability of this occurring is only one in forty million.

In one way, Gas is a continuation of Wheel and Guns. In all three cases, each rigidly designated Californian faces an objective risk of one in forty million. The cases differ, however, in the distribution of risk across nonrigid designators. In Wheel, the distribution represents one extreme. All risk is concentrated in the likelihood of one person dying. In Guns, the distribution is spread out across all forty million nonrigid designators ranked from the worst-off to the best-off. The risks for those higher up the list are very high; for those lower down the list, they are minute. Now in Gas we face the opposite extreme. The risks are spread out perfectly evenly across all nonrigid designators. All nonrigid designators are tied, because whatever will happen, everyone in California shares the same fate. What is particularly interesting about Gas is that the distribution of discounted complaints is the same for rigid and nonrigid designators. Whether we use rigid or nonrigid designators to determine the justifiability of our action does not matter since both will yield the same result.

This is challenging for the *ex post* contractualist for the following reason: I have argued that *ex post* contractualists should accept the following two principles. They should accept *Ex Post Discounting*. This allows *ex post* contractualism to be applied to cases where harms are not guaranteed, and it provides the *ex post* perspective with a model of justifiability to each. Second, they should accept *Equal Treatment for Equal Statistical Loss*. This means that in Wheel and Guns what matters is the number of expected lives lost. The principle follows from accepting the claim that the disvalue of a given harm should not vary depending on how many other people will be harmed. The possibility that no person may die should be balanced by the possibility that more than one person may die.

My case Gas shows how these two principles can conflict. The number of expected lives lost in Gas is one, just like in the other two cases. If Wheel and Guns are on a par, then so is Gas. But Gas contains only heavily discounted complaints by nonrigidly designated persons. This is because the complaint of the worst-off Californian is based on only a one in forty million chance of death, a morally trivial complaint. Following *Ex Post Discounting*, it should be these discounted complaints that determine the justifiability of the risk imposition. If we want to follow *Equal Treatment for Equal Statistical Loss* and hold that the risk imposition in Gas is impermissible, we would need to aggregate the complaints in Gas. But whichever way we calculate the complaints, the complaints in Gas seem very

close to the complaints by the many in Transmitter Room. The complaint of Bob Johnson resembles the complaint of Jones, the worker in the transmitter room. As it turns out, the strongest version of an *ex post* view leads to a case that is very much like Transmitter Room. If we allow aggregating the complaints in Gas, then why can we not aggregate the complaints in Transmitter Room?

One proposal is that while individual and nonaggregated complaints matter, aggregative considerations can determine whether it is reasonable to reject principles.¹⁶ Following this proposal, it is still individual complaints that matter. But their strength would be magnified by the number of people having the same complaint.

Ex Post Discounting (Multiplied): When assessing the complaints of individuals, we should discount the complaints of nonrigidly designated individuals such as the worst-off, the second worst-off, and so on by the improbability of harm. The strength of their complaint is determined by multiplying the strength of their individual complaint by the number of nonrigidly designated individuals who will be equally affected.¹⁷

According to this proposal, it would be unreasonable for Bob Johnson to insist on his complaint given that there are so many complaints on the other side. The strength of the individual complaint opposing Bob Johnson is magnified by the number of people who would be similarly affected. Yet Jones is equally faced with many complaints on the other side. Why should we not be allowed to multiply the individual complaint of a single football fan by the number of football fans that are equally affected? If we are allowed to magnify this individual complaint, then it would be unreasonable for Jones to reject a principle that allows the World Cup match to be broadcasted. The proposal to allow individual and nonaggregated complaints to be amplified reintroduces aggregative reasoning through the back door. So what could distinguish between Gas and Transmitter Room? Why should we understand Bob Johnson's insistence on his individual complaint as unreasonable while Jones's insistence is reasonable?

Perhaps it is the following: In Transmitter Room, the small complaints stem from mere annoyance. In Gas, the small complaints are derivative of a very serious moral claim, namely the claim not to die. This very serious claim becomes

16 This is suggested by T. M. Scanlon in his "Contractualism and Justification." Véronique Muñoz-Dardé had earlier presented the idea that in some cases agents with strong complaints cannot reasonably reject principles. Muñoz-Dardé invokes the idea of a threshold of reasonable demands that one can make on others. This allows for the possibility that a person with a stronger individual complaint may not be able to reasonably reject a principle ("The Distribution of Numbers and the Comprehensiveness of Reasons," 208–15).

17 I owe this proposed revised principle to an anonymous reviewer.

less important to each individual taken separately, due to the sharp discounting of their complaints by the likelihood of death occurring. Maybe Bob Johnson's insistence is unreasonable while Jones's is not because in Jones's case the opposing complaints are not complaints of the right kind. The trivial joy of watching football is not relevant to Jones's torture, while the risk of death, even if small, is relevant to Bob Johnson's lost leg. This proposal is coherent with what I wrote earlier about the opposition to aggregation. I wrote that "we should not *aggregate morally trivial complaints* so that they *outweigh serious moral complaints* of single individuals" (emphasis added). Trivial complaints should not outweigh serious complaints regardless of the numbers involved. But this leaves open that complaints of similar magnitude or qualitative significance could outweigh each other depending on the numbers.¹⁸

In line with the earlier distinction between the complaints of the Californians and the complaints of the World Cup viewers, we could think of complaints as being qualitatively different for different levels of actual or possible harm. Following this idea, heavily discounting a complaint against being killed does not make this complaint morally trivial. The complaint is still qualitatively on a different level than the complaint against mere annoyance. This allows us to distinguish the aggregation in Gas from the aggregation in Transmitter Room.

One problem with the idea that risks of death are qualitatively different from very small certain harms is that the same answer is available to the *ex ante* contractualist. If we stop believing that heavily discounted risks of death are morally trivial, then we could engage in a limited form of aggregation in cases like Wheel too. And then *ex ante* contractualism can account for the same answer. In other words, once we adopt the view that heavily discounted harms are not morally trivial, we lose a key motivation for adopting *ex post* contractualism.

Second, treating risks of death as qualitatively different from small certain harms fails Equal Treatment for Equal Statistical Loss in a central case. It cannot treat identified victims and statistical victims alike, even though equal respect for identified and statistical victims was one of the key motivations for *ex post* contractualism. Suppose that in a one-versus-one confrontation, a complaint against missing fifteen minutes of a World Cup match is as strong as a complaint against a risk of death of one in forty million. If we can save either one person from missing part of the match or one person from this risk of death, we should

18 The idea that complaints can only be aggregated in some circumstances is called limited aggregation. The view is suggested by Scanlon, *What We Owe to Each Other*, 238–41; and also endorsed and defended by Kamm, *Morality, Mortality*, 1:156–61, and *Intricate Ethics*, 31–40; Temkin, *Rethinking the Good*, ch. 3; and Voorhoeve, "How Should We Aggregate Competing Claims?" I set out my own view of limited aggregation in Steuwer, "Aggregation, Balancing, and Respect for the Claims of Individuals."

be indifferent. If, however, there were two people subjected to this risk of death, we should save them at the expense of the person missing parts of the World Cup match. Now what if there are many people who would be missing fifteen minutes of the World Cup match? It seems that here numbers should matter. Otherwise we would give undue importance to small risks. We should rather spare a million people from missing the World Cup match than to reduce a one in forty million risk of death to a single person. In other words, here we should be allowed to aggregate the complaints against missing parts of the World Cup match. If this is so, then we should be allowed to aggregate both the complaints against the risk of death and the complaints against missing fifteen minutes of the World Cup match. If there are many complaints against small risks, similar to my Gas case, then these might add up to one expected life lost. But since we are also allowed to aggregate the complaints of the World Cup viewers, these complaints might be decisive. However, if we contrast a single identified person with the World Cup viewers, as in Transmitter Room, we are required to save the identified person. Distinguishing between different kinds of harm cannot, therefore, treat cases where a statistical life is lost the same as cases where an identified life is lost.

Third, the idea that heavily discounted complaints against serious harm remain morally significant is also implausible in its own right. One downside of this view is that it has a problem analogous to Kamm's Sore Throat case. In Kamm's original case, we have a choice between saving one life and saving another life *and* saving someone from a sore throat. Kamm wants to say that here we should not decide in favor of saving the second person's life solely on the grounds that we can also save someone from a sore throat.¹⁹ Now imagine that the tiebreaker is not the sore throat but the imposition of a tiny risk of death, for example, by calling an ambulance. Not only is it the case that we would be *permitted* to save the person who does not need the ambulance on the grounds that her rescue does not impose a trivial risk. But also we would be *required* to save her. It would be impermissible *not* to use the trivial risk as the deciding factor. Together with the insufficient motivation for treating equally strong complaints differently, I think this gives us grounds to treat equally strong complaints as either relevant or irrelevant. What we should accept, however, is that complaints can be aggregated when their strength is relevant to the strength of the complaints with which they are competing.

Since the *ex post* contractualist cannot distinguish between the aggregation in Gas and the aggregation in Transmitter Room, she should accept the risk imposition in Gas as permissible. She then cannot accept the principle of Equal

19 Kamm, *Morality, Mortality*, 1:146–47.

Treatment for Equal Statistical Loss. This is bad news for the *ex post* contractualist for two reasons. First, she must reject the plausible claim that harms have the same disvalue regardless of how many other people will also be harmed. The risk that one person will be harmed will receive greater weight than the risk that any additional victim over and above the first victim will be harmed. Second, a version of *ex post* contractualism that accepts the risk imposition in Gas includes a bias against statistical lives, a charge *ex post* contractualists usually raise against their *ex ante* colleagues. In some cases, like Gas, a statistical life will not be saved even though an identified life would have been. This criticism against the *ex ante* view becomes less convincing since the two theories differ only in the degree to which they are biased against statistical lives.

3. WHAT WE OWE . . . TO WHOM?

My discussion of the sequence has revealed two things. First, it has shown that two plausible principles that an *ex post* view would want to fulfill cannot be jointly fulfilled. Second, it has given us a better way of understanding *ex ante* and *ex post* views. We can understand these views as answering the question of whose complaints we should be concerned with as contractualists. Should we appeal to the complaints of identifiable individuals (epistemic *ex ante*)? Should we appeal to the complaints of rigidly designated individuals (objective *ex ante*)? Should we appeal to the complaints of nonrigidly designated individuals (*ex post*)? In what follows I will argue in favor of objective *ex ante*. The concern with the complaints of rigidly designated individuals expresses the best model of acting in ways that are justifiable to each separate person. As I explained earlier, such a concern with rigidly designated individuals means that we should draw a distinction between cases involving epistemic risk and cases involving objective risk. In a second step, I argue that this is a virtue of objective *ex ante* contractualism since it illuminates the distinction between luckless and doomed victims.

3.1. *Justifiability to Each Separate Person*

The core idea of contractualism is that actions must be justifiable to each. Moreover, in order to respect the separateness of persons, our actions must be *justifiable* to each as a *separate* person. This guiding idea, I argue, supports the view that our justification should address rigidly designated individuals rather than identifiable individuals or nonrigidly designated individuals. In other words, the basic idea of contractualism supports objective *ex ante* contractualism.

Consider the difference between the following three statements made by the US president after deciding on which option to take. The three statements mirror

the three options for who the ideal addressee of justification is. In each scenario, the president addresses a victim and tries to justify the imposition of the burden on her.

1. To Charlotte Williams, born on the first of June 1975, resident of Santa Barbara, who is going to die from this measure, I can only say that I am deeply sorry but your complaint against the measure was outweighed by other complaints. Even though it is hard to accept, I am convinced the measure is justifiable to you too.
2. To the Californian with the genetic predisposition, whoever he or she *may be*, I hope that you hear me. I can only say that I am deeply sorry but your complaint against the measure was outweighed by other complaints. Even though it is hard to accept, I am convinced the measure is justifiable to you too.
3. To the Californian who is going to die from the measure, whoever he or she *turns out to be*, I can only say that I am deeply sorry but your complaint against the measure was outweighed by other complaints. Even though it is hard to accept, I am convinced the measure is justifiable to you too.

Should we believe that there is an important moral difference between justification 1 and justification 2? Epistemic *ex ante* contractualists like Johann Frick believe that there ought to be. Frick, for example, holds that our ability to identify a given individual with a complaint makes a difference. Should it be impossible or overly burdensome to identify which person is going to die from the proposed policy, then we ought to treat this as a case of many discounted complaints against killing.²⁰ I disagree. Frick's argument relies on an idea about what we can justify to each person. But this, I think, misrepresents the core idea of contractualism. Contractualism is about *justifiability* rather than *actual justification*. Justifiability is already an idealized concept. It requires us to take into account all effects of actions on everyone concerned and to take into account all complaints everyone may have. It also requires us to take into account complaints that no one in fact has or will raise. The ideal of justifiability is one of acting in accordance with principles that would sustain a hypothetical and ideal form of justification. Since we have already idealized, it is difficult to see why we should not idealize epistemic limitations as well.

Therefore, I believe that we should think of 1 and 2 as equally good justifications. In both cases, the president is justifying her behavior to the victim. Both speeches are meant for one person alone, and address and justify the action to

20 Frick, "Contractualism and Social Risk," 193–94.

one person alone. The only difference is that speech 1 includes more detail that allows us to identify the individual. While identifiability is important for Frick, he does not discuss what is required to identify an individual. Taking a cue from Casper Hare, we can think of “identifying” an individual by knowing more personal information about that particular person.²¹ We might then have identified a victim without knowing their name as long as we know enough distinctive personal information. But whether the president is able to include more detail in the description, such as name, birth date, place of residence, or other identifying information, is morally irrelevant. We are not interested in token individuals because of names or other personal information such as appearance, tastes, or talents that allow us to identify them. This information is morally superfluous. We are interested in token individuals because of their particular situation and predicament. The description “the Californian with the genetic predisposition” conveys everything that is morally important. Objective *ex ante* contractualism bases its complaints only on morally relevant information about a person’s situation. This ensures that we do not confuse justifiability, which is at the heart of contractualism, with actual justification.

Even more so, at times additional information that allows us to identify individuals can even distort our moral reasoning. Imagine a doctor who has to decide on which treatment to administer to two unconscious patients, Deborah and Eric.²² Out of expediency, the doctor has to administer the same treatment for both, even though they have two different diseases, *x* and *y*. On the one hand, the doctor can think of the prospects that Deborah and Eric have. Without any further information, the doctor would assign a fifty-fifty probability that Deborah has either of the two diseases. (And the same for Eric.) The trade-off between the two diseases will then be regarded as an intrapersonal trade-off where Deborah’s and Eric’s interests are the same. On the other hand, the doctor could think of the interests of “the patient with disease *x*” and “the patient with disease *y*.” In this way, she would regard the trade-off as interpersonal. This way of regarding the case is superior. The doctor knows that she is dealing with an interpersonal trade-off; she knows that the interests of her two patients are not aligned. Doing one act will harm one and benefit the other. The doctor should not deceive herself into thinking that this is a choice without a conflict of interests.

Rather than between 1 and 2, we ought to hold that there is an important difference between the justifications in 2 and 3. While the contrast between 1 and 2 has shown the importance of justifiability as opposed to actual justification, the

21 Hare, “Should We Wish Well to All?” 467–71.

22 The case is a variation of one by Anna Mahtani (“The *Ex Ante* Pareto Principle,” 310–11.) Mahtani credits Caspar Hare as her inspiration.

contrast between 2 and 3 shows how important it is that justifications have to be addressed to separate persons. In statement 2 (and 1), the president addresses and talks to one person alone, while in 3 the president does not address any specific person. At the time of the president's address, the words are not addressed to one individual alone. The first two speeches constitute a private channel of communication between the president and the victim. The communication and the justification are one to one. If what the president says is correct, then she would have succeeded in justifying her action to this person.

In the third speech, however, the words cannot address only one person. The justification cannot be private or one to one in the same sense. At best, the president will have addressed a person once the policy is applied, but this does not make it the case that the president did address this person prior to the action or when acting.²³ It is thus difficult to see how the justification in 3 conforms to the contractualist ideal of justifying one's action to each. Justification is owed to each separate person. But the discourse in 3 does not address persons separately. The appeal of a justification like 3 stems from the way we assimilate this thought with justifications given along the lines of my proposed speech 2. In these cases, the "someone" refers to a given individual. But this is not the case in 3. In 3, the justification addresses a compound of different individuals across different possible worlds.²⁴

We can see this even more clearly when we consider cases where the complaint of the Californian who is going to die outweighs the complaint of a rigidly designated individual, such as Bob Johnson. Bob Johnson could rightly ask who the person is that can reasonably reject the proposal that would get him off the

23 The formulation here implies a rejection of the view that future contingents already have truth values. But my argument is not restricted to this metaphysical view. Some philosophers believe that future contingents already have truth values and that this view is compatible with indeterminism (see Belnap and Green, "Indeterminism and the Thin Red Line"; or Lewis, *On the Plurality of Worlds*, 206–9). If this is true, then it is the case that the president's justification does actually address one individual even though the identity depends on the objectively risky event. However, this only holds *if* the president *actually* acts this way. Should the president decide not to act this way, we have to assess a counterfactual rather than a future contingent. Under most standard views of counterfactuals, these counterfactuals will be open counterfactuals without a truth value (see Hare, "Obligations to Merely Statistical People," 380–82). This means that the model of justifiability used in 3 and whether it addresses a person will depend on what the decision maker ends up doing. But this puts the cart before the horse. An action should not be more or less justifiable based on what the agent actually does. The fact that alternative actions will be open counterfactuals also means that the model of justification used in 3 cannot be applied to help decide between different alternatives, since all but one of the alternatives include an open counterfactual.

24 See also Frick, "Contractualism and Social Risk," 196.

hook. It cannot be that we determine the identity of said person only after the fact. Even more so, *ex post* contractualism makes it impossible for us to know or determine who that person would be. It would be morally impermissible to perform the only actions that could determine the identity of this person. It will never be determined who the person was for whose sake we sacrificed Bob Johnson's leg.

Indeed, there is a compelling justification for imposing risks in cases like Wheel, even though we know one person will be harmed. Note that the victim in cases like Wheel would not have been permitted to save herself over Bob Johnson. She was facing only a small risk of death, a risk small enough that she would have been required to bear this risk. We can give the following powerful reason to the victim: you were not allowed to save yourself even accounting for your partiality toward yourself. So, you cannot complain to a third party that was not allowed to be partial toward you that she did not save you.²⁵

The fact that speech 3—and thereby the model of justifiability that *ex post* contractualism employs—fails to address a particular person can also be seen clearly in a different context. By carrying the logic of speech 3 forward, *ex post* contractualism makes the permissibility of risk impositions dependent on mere population size. For this, see the following case:

Water (County Level): There is a toxic pollutant in the groundwater all over California. The pollutant will lead to every Californian losing the small finger of the right hand if nothing is done. Scientists have developed a chemical that will neutralize the pollutant. However, the chemical is still in development and thus is risky. The scientists have reduced the risk of death considerably to only one in forty million. The risks are objective and probabilistically independent for each Californian. While the pollutant affects the groundwater of all of California, the water systems are separate for each county. Each local authority has to make the decision.

Let us take as an example Santa Barbara County, which has only about 450,000 residents in contrast to the forty million residents of California as a whole. The objective risk for each individual to die is still one in forty million. But while the likelihood of at least one person dying is significant across California, the likelihood of at least one person dying in Santa Barbara County is lower. The probability is only slightly over 1 percent. Perhaps discounting the harm of death by 99 percent makes the harm less grave than the loss of the finger. (If you do not believe the harm is discounted enough, just reduce the population size further.)

25 See also Voorhoeve, "How Should We Aggregate Competing Claims?" 74.

If this is the case, then *ex post* contractualism allows releasing the chemical for Santa Barbara County. If all the other counties are of a similar or smaller size than Santa Barbara, the risk imposition would be permissible in those counties too.²⁶

This leads to an absurd conclusion. *Ex post* contractualism needs to hold the following. If the government of California were to decide, releasing the chemical would be impermissible in the contractualist sense; it would not be justifiable to each. If each local government were to decide, releasing the chemical would be permissible in each case. It would be justifiable to each. Even though every single person is affected in the very same manner, the policy would turn out to be unjustifiable to one of them if the decision were taken at a different level. *Ex post* contractualism somehow generates a person with a complaint from a group of persons without a complaint. The absurdity is even clearer if we accept that unjustifiable risk impositions wrong an individual.²⁷ While none of the county governments would be wronging an individual if they released the chemical, the government of California would be wronging an individual. But who would be wronged? This example reveals that *ex post* contractualism fails to give us a model of acting in ways that are justifiable to *separate* persons.

3.2. *The Luckless and the Doomed*

Objective *ex ante* contractualism draws a distinction between cases like Dust in which the risk imposition is epistemic and cases like Wheel in which the risk imposition is objective. This is because in cases of epistemic risk, like Dust, we can identify a rigidly designated individual who is certain to be harmed while in cases of objective risk, like Wheel, we cannot. This distinction may seem suspect, and none of the other authors writing on contractualism has considered it relevant.²⁸ However, I believe that distinguishing between epistemically risky cases and objectively risky cases is far from being a defect of the view. To the contrary, it is a virtue of it. The reason is that this distinction tracks another distinction about the moral relevance of luckless and doomed victims. In epistemically risky cases like Dust, there is going to be one doomed victim, and in objectively risky cases like Wheel, there is going to be one luckless victim. While the effect on

26 Some counties of California are comparatively large, for example, Los Angeles County with over 10 million people. We can imagine that in those counties more local authorities have to make the decision.

27 See, e.g., Oberdiek, *Imposing Risk*, 126–53. Frances Kamm has argued for the more radical claim that Scanlon's account for wrongness should generally be understood as an account of wronging (*Intricate Ethics*, 461–68).

28 Indeed, Frick argues against its relevance in “Contractualism and Social Risk,” 197–201.

both is the same, we can see that there is a significant difference between having doomed a person who ends up dying and having given that person a very favorable chance of survival.

John Broome in his discussion of fairness makes the following remark about persons who lose out in the allocation of a scarce good.²⁹ Whoever loses out has grounds for complaint. But the person would have an even bigger ground for complaint if it were never even in the cards for her to have received the good. We cannot justify our allocation to this person by saying that we gave her a fair shot at receiving the good. Losing out for this person is not tough luck but, worse, an inevitable feature of our decision. The fact that she might have won, that it was once in the cards for her to win, mitigates her complaint against missing out. In short, after the allocation, a luckless loser has a less strong complaint than someone who was doomed to lose. The lottery example shows how the kind of risk that is at play in allocating the good matters for the complaints that individuals can raise. In a lottery that employs epistemic risks, it was never in the cards for anyone other than the winner to win. In an objectively risky lottery, this is not the case. Every person stood a chance of getting the good. The lottery is fair because it is the luck of the draw that decides who gets it.³⁰ Objectively risky lotteries are such that we can say to the person that she could have received the good. We designed the lottery such that it could have easily gone the other way and she might have won.³¹

These points about fairness in allocating goods are not limited to the allocation of benefits. They should also apply to the allocation of burdens or harms. Common examples to illustrate lottery fairness include such cases. The draft lottery to select soldiers for the Vietnam War is a paradigm example. The cases I have discussed are similar. In all cases, harms are avoidable only at the expense of a moral catastrophe. We have to decide about the allocation of harm. This means that we can say to those who are luckless that they could have avoided the harm, whereas those who would have been doomed would not have had any

29 Broome, "Fairness," 98.

30 This idea is even invoked by critics who account for lottery fairness in a different manner. George Sher and Michael Otsuka give accounts of lottery fairness of merely epistemic lotteries since both doubt that lotteries with objective risks exist. Sher mentions the "luck of the draw" interpretation as the most obvious rationale for lottery fairness, but adds that it is incomplete because it cannot account for the fairness of lotteries that do not employ objective risks. Otsuka argues that objectively risky lotteries would be fairer than epistemically risky lotteries, if it were possible to run them. Sher, "What Makes a Lottery Fair?" 203-4; Otsuka, "Determinism and the Value and Fairness of Equal Chances."

31 I owe this point to Kai Spiekermann. He explores this idea in connection to lottery fairness and social risk in "Good Reasons for Losers."

such chance. It is a virtue of objective *ex ante* contractualism that it can distinguish in this manner between luckless and doomed victims.

While the previous considerations on fairness illustrate the importance of the distinction between the luckless and the doomed in giving reasons after the risk materializes, there are also reasons to care about the distinction before the action. Consider the following case narrated by Anatol Rapoport.³² In the Second World War, an allied air base in the South Pacific faced the problem that most of their planes did not survive their allocated missions. The chance of survival was only one in four. An alternative but rejected policy would have increased the chances of survival. Only half of the planes would fly missions with increased bomb load. The increased load would mean that less fuel would be available, and the pilots could not return to safety and would crash. Instead of giving everyone a chance of one in four, the policy would fate half the pilots to certain death. The repulsion against and failure to adopt the policy is best explained by an objection against dooming individuals to death.³³

However, the difference between doomed and luckless victims goes beyond cases where the victims know their fate. Assume a small variation of this case where, in order to ensure compliance with the order to fly, after the selection by lot all pilots board a plane. Neither commanders nor pilots know which planes are loaded and which carry empty loads. Pilots who fly an empty plane have orders to return to a different base when they realize their plane is empty at the first target. At the decision to order the pilots to fly, every pilot faces an epistemic risk of death of 50 percent. This variation is no less objectionable than the initial plan. By distinguishing between doomed and luckless victims, objective *ex ante* contractualism can account for this. The doomed pilots are certain to die whereas under the ordinary protocol all pilots face a three-quarters objective risk of death. By contrast, epistemic *ex ante* contractualism may justify the order to fly given that it reduces the epistemic risk each pilot faces. *Ex post* contractualism in turn would justify the order to fly given that it reduces the number of expected lives lost. Only objective *ex ante* contractualism can account for the answer that is both the actual decision at the base and the intuitively correct one.

One might object to my analysis of the case of the pilots. Assuming that the selection by lot were random, every pilot would have faced a 50 percent objective risk of death under the alternative policy as opposed to a 75 percent objective risk of death under the standard policy. However, it is not accurate to draw

32 Rapoport, *Strategy and Conscience*, 88–90. Rapoport presents this case as a real-life case but could not vouch for its authenticity.

33 Jonathan Glover reports that the horror of certain death motivates the refusal to accept the policy of one-way missions in Rapoport's example (*Causing Death and Saving Lives*, 212–13).

the conclusion that objective *ex ante* contractualism would therefore endorse the alternative policy. The problem here is similar to the problem of medical experimentation discussed by Frick. In the example of medical experimentation, there is an *ex ante* selection of persons to be experimented on. At the stage of selection, the policy of experimenting is beneficial to all, but after the selection is made, severe hardship is imposed on some. Objective *ex ante* contractualism can avail itself of the same reply as epistemic *ex ante* contractualism and adopt what Frick calls the Decomposition Test.³⁴ The Decomposition Test imposes a requirement to always act, in each action, in ways that are justifiable to each. The policy of selecting people at random first and then imposing severe hardships on them does not meet this test. This holds for the case of medical experimentation as well as for the case of the pilots. When sending out the pilots, some pilots are doomed to certain death. Objective *ex ante* contractualism prohibits this.³⁵

Our objection to dooming the pilots to certain death is linked with our intuitions about risk concentration and risk dispersal. Take, for example, our reaction to a now debunked story about the Coventry Blitz, the horrendous bombing raid of Nazi aircrafts on the city of Coventry. According to the story, Churchill knew about the impending devastating attack on Coventry and could have averted it. To avoid revealing military intelligence, Churchill sacrificed Coventry for the sake of the overall war effort and reduction of the overall death toll. When the story was published, it was perceived as a grave accusation and moral flaw for Churchill to have acted this way.³⁶ Distinguishing between doomed victims in Coventry and unlucky victims elsewhere in the United Kingdom can explain why. Rapoport's pilot case and the Coventry Blitz reveal that our intuitions about concentrating and dispersing risks are sensitive to what kind of risk we are talking about. The plan to fly one-way missions disperses and reduces epistemic risks, but this does not make the plan very appealing given that objective risks are concentrated. There is little point in dispersing epistemic risks if we know

34 Frick, "Contractualism and Social Risk," 201–12.

35 Nir Eyal has suggested that what is problematic about Rapoport's case is not that the pilots are doomed, but rather that they are doomed by their commanders. The commanders, as opposed to enemy fire, would be killing the pilots by adopting the policy. See Eyal, "Concentrated Risk, the Coventry Blitz, Chamberlain's Cancer," 105–7. However, I believe that this part of the story is not central. My reaction would not change if some of the planes had insufficient fuel due to sabotage and the commanders had the choice of aborting the mission and calling the planes back. (Imagine that bombs are loaded automatically according to overall weight.) The commanders would still doom some pilots to certain death, even if the pilots would not be killed by the commanders.

36 See Eyal, "Concentrated Risk, the Coventry Blitz, Chamberlain's Cancer," 94–95. Eyal seeks to vindicate Churchill's imagined reasoning.

that it is already carved in stone who will die. However, dispersing objective risks is a genuine sense in which burdens are shared and additional burdens are spread more widely.

Thus far I have argued that part of the reason why the distinction between objective and epistemic risks is meaningful is because it can explain the moral difference between luckless and doomed victims. This allows me to respond to one concern about my view. Imagine a vaccine that we know carries a certain small risk of serious harm. Whether the foreseen harms of mass vaccination are a reason against the mass vaccination will depend, on my view, on the specific mechanism by which the risk manifests itself. If the mechanism is a random mutation, then it is a small objective risk, whereas if the mechanism relies on genetic predispositions, then it is a small epistemic risk but a large objective risk. Why should this mechanism matter? In response: the mechanism matters because in the case of the random mutation, the harmed victim is luckless, whereas in the case of the genetic predisposition, we would doom the victim to be harmed. As I have argued, there is an important moral difference between being luckless and being doomed, and this moral difference makes the otherwise uninteresting-seeming difference in the biological mechanism of the vaccine relevant. While we often do not know with certainty what mechanism applies, we often have information about whether our applied case is more like the case of random mutations or more like the case of genetic predispositions. This, I believe, rightly influences how we ought to act in the case.

The distinction between objective and epistemic risks is also important for another reason. It can illuminate the importance of hypothetical consent. An important and familiar reason for rejecting *ex post* contractualism is that it makes actions impermissible even if these actions would receive the hypothetical consent of all affected parties. For each individual, it is sometimes rational to take small risks of death for moderate gains. For example, it would be rational to take a vaccine against a disease that is not life threatening even if there is a risk of a lethal allergic reaction. If such risks are imposed on a large scale, then we can be virtually certain that some person will die from the risk. Not only are these risk impositions intuitively permissible, but we can give a strong argument in favor of them. Frick has called this the Argument from the Single Person Case.³⁷ If the risk imposition were to affect only a single person, it would be permissible. In such a case, it seems reasonable that we should do what is in that person's rational self-interest. Now in a second step, we learn that there is a second person in

37 Frick, "Uncertainty and Justifiability to Each Person," 133–34; and Frick, "Contractualism and Social Risk," 186–88. Similar arguments are made by Tom Dougherty ("Aggregation, Beneficence, and Chance," 8–11) and Caspar Hare ("Should We Wish Well to All?" 455–67).

an identical position as the original person. The risky treatment is available at no additional cost for that person too. The case is still relevantly similar to deciding for one person. It does not involve any competing claims. We can add more and more people. Individually, we would always favor giving them the treatment. Yet *ex post* contractualism needs to hold that for a sufficiently large group the risk imposition becomes impermissible.

Is there anything the *ex post* contractualist could say to reject the Argument from the Single Person Case? The best response seems to be the following. The hypothetical consent that each person would give is vitiated because each person is imperfectly informed.³⁸ If we knew that a person would only consent because she was insufficiently informed, it is less plausible to assign moral weight to this hypothetical consent. Imagine that you are a guardian charged with that person's interest. If you were fully informed and knew that the risk imposition was in that person's interest only because of imperfect information, you would not assign moral importance to that fact about self-interest. A close variation of this case is a case where you are in charge of various persons' interests. You may not know which person is going to lose out, but you still know the related fact that one of the persons whose interests you look after is going to lose out. As a fully informed guardian, you would therefore object to the action. In epistemically risky cases like the vaccine case, this is the case. Somewhere in the chain, there is a person for whom it is not in their fully informed self-interest that the risk will be imposed. The chain of single person cases is no longer fully symmetrical under conditions of full information. Since we can anticipate this, we have grounds to object to the risk imposition.

The reply to the Argument from the Single Person Case helps us refine the importance of hypothetical consent. Unlike with actual consent, we have no reason to give moral significance to hypothetical consent that arises due to imperfect information. Yet this challenge does not impede giving significance to hypothetical consent that is not tainted in this manner. Such untainted hypothetical consent is at stake in objectively risky cases. Remember the Water case introduced earlier. In Water, every Californian faces the same problem for deliberation. Either they will lose their small finger or they will incur a minute risk of death. The gamble is in the self-interest of each Californian; each would hypothetically consent. In this case, the response that hypothetical consent arises only out of imperfect information has no bite. Even if all Californians knew all relevant facts about themselves, it would nonetheless be in their self-interest to take the gamble. The Argument from the Single Person Case stands. Distinguishing between epistemic and objective risks helps us understand that the Argument from the

38 See Fleurbaey and Voorhoeve, "Decide As You Would with Full Information!"

Single Person Case is compelling in some cases while unconvincing in others. By distinguishing between these cases, objective *ex ante* contractualism retains what is attractive in the Argument from the Single Person Case while avoiding the charge that hypothetical consent is vitiated due to imperfect information. In the revised case, all risk impositions are independent from one another. There is no conflict over the resource that gives everyone a favorable prospect for their lives. Since there is no connection between the risks, there is no reason why it should not be permissible to impose all of them at once. Consequently, objective *ex ante* allows us to impose all of them at once.

4. OBJECTIONS

I will consider two main lines of objection to my version of *ex ante* contractualism that discounts objective, rather than epistemic, risk. The first line of objection stems from the possibility that determinism is true. The second line of objection criticizes an identified victim bias in my position.

4.1. Determinism

My view distinguishes between objective risks and epistemic risks. There is a worry that even if this distinction would be of moral importance, it is irrelevant in the real world. If determinism is true, the worry goes, then there is no such thing as objective risk. There might be actually observed frequencies but no objective risk in a robust sense that could be morally relevant. The view that the truth of determinism implies the absence of objective chances was once taken as the orthodox view in the philosophy of probability. Recently, however, there has emerged a growing literature in the philosophy of probability that argues that objective chance or objective probability is compatible with determinism.³⁹

A first reason to think that objective probabilities are compatible with determinism stems from the existence of probabilistic laws in science. To give some examples, classical statistical mechanics, evolutionary theory, Mendelian genetics, meteorology, and the social sciences all include probabilistic laws. In fact, it appears that deterministic laws are largely confined to just one branch of science, namely the physical sciences. The probabilities posited by the laws of the special sciences, including parts of the physical sciences like classical statis-

39 See Loewer, "Determinism and Chance"; Hofer, "The Third Way on Objective Probability"; Glynn, "Deterministic Chance"; Eagle, "Deterministic Chance"; Lyon, "Deterministic Probability"; Strevens, "Probability out of Determinism"; Emery, "Chance, Possibility, and Explanation"; Frigg and Hofer, "The Best Humean System for Statistical Mechanics"; List and Pivato, "Emergent Chance."

tical mechanics, do not appear to be epistemic. For example, the process of ice cubes melting when being put in water is a probabilistic process according to classical statistical mechanics. It appears that classical statistical mechanics can, by virtue of this probabilistic law, *explain* why the ice cube is melting. Indeed, if we believe that special sciences above the microphysical level are able to explain phenomena, then they explain these phenomena by reference to probabilistic laws. This makes it difficult to conceive of such laws as being concerned with epistemic probabilities. The laws of classical statistical mechanics cannot both incorporate our ignorance about deterministic processes and at the same time explain why ice cubes are melting or why the climate system is changing. Our ignorance cannot explain.

So how can we accommodate both the fact that laws of the special sciences posit objective chances and the idea that the universe is deterministic at the microphysical level? One rationale for the compatibility of objective chance and determinism at the microphysical level is that the descriptions of “chance” and “determinism” are level specific.⁴⁰ It is imprecise to talk about whether or not the world is deterministic. The real question is whether or not the world is deterministic *at a specific level*. A helpful test to see whether the world is deterministic at a given level is to ask whether knowing the entire history of the world described at that level determines a future event. Those who argue that the world is deterministic at the microphysical level mean to say the following: if we knew all the laws of nature as well as the initial conditions of the universe described in microphysical language, then the only chances of an event happening are zero or one. But this does not say anything about whether or not the world is deterministic at some macrolevel. It does not follow that, at the macrolevel, the history of the world already determines the event. In other words, determinism at the microphysical level can coexist with indeterminism at some macrolevel. This way, macrolevel events like melting ice cubes or coin tosses will have their own macrolevel chances.

For the purposes of moral theorizing, we are predominantly concerned with the agential level, the level at which we describe agents and their actions. The agential level is the appropriate level for the moral decision-making of agents. What would rule out the possibility of objective chances in the relevant sense is, therefore, not determinism at the microphysical level but rather *determinism at the agential level*. Yet there is no reason to think that our world is deterministic at the agential level. To the contrary, all indications of our best available (social) science at the agential level tell us that the world is *indeterministic* at the agential level. Even if we knew the entire history of the universe described at the level

40 Glynn, “Deterministic Chance”; List and Pivato, “Emergent Chance.”

of agents and macro-objects, like coins, together with all laws of human behavior, we would not be able to predict, say, the outcome of the next presidential election. Arguments for determinism rely on information about microphysical particles and their properties, something that is inadmissible when thinking about whether the world is deterministic at a higher level. The level-specific approach to determinism and chance retains the ability to draw a distinction between objective chance and epistemic credence at each level of description.⁴¹ Imagine an agent is about to toss a fair coin. The odds of the coin landing heads are 0.5. These are objective chances since the prior history of the world, at the level of coin tosses, does not determine this event. After the coin toss, the agent is covering the coin and asks again what the odds are of the coin having landed heads. The answer would seem to be 0.5. But this statement about probabilities is clearly different from the earlier one. The second odds are credences, the first are chances. Thus, the level-specific view can retain the distinction between chances and credences at every level. This distinction in turn means that while agents can create objective chances, they can also create merely epistemic risks. A lottery based on whose birthday is earliest in the year would create epistemic risks if the birthdays of participants are unknown, but it would not create objective risks for the participants.

We can see the point of the level-specific view in another way. Consider again the coin flip. Assume that we hold all other factors constant except for the force exerted on the coin. The following conditionals might all be true:

If I flip the coin with a force between 0.18345 and 0.18348 N, it will land heads.

If I flip the coin with a force between 0.18349 and 0.18352 N, it will land tails.

If I flip the coin with a force between 0.18353 and 0.18356 N, it will land heads.

And so on. But what about the conditional “If I flip the coin, it will land heads”? Or the conditional “If I flip the coin, it will land tails”? The antecedents of these conditionals are underspecified. They do not tell us with which force the coin is flipped, and the deterministic laws of physics tell us that small changes in the force applied to the coin lead to different outcomes. The antecedent of the underspecified conditionals describes a set of possible worlds. In this set, there are some possible worlds where the coin lands heads and some possible worlds where the coin lands tails. What we can give for the underspecified conditional

41 See List and Pivato, “Emergent Chance,” 139–42.

is a probability of how many worlds are head-landing worlds.⁴² The fact that this probability is not merely epistemic can be seen if we consider the case in which the conditional is a counterfactual conditional. Processes like this coin flip are counterfactually open. No head-landing world is relevantly more similar to our actual world than any tail-landing world. Since the process is counterfactually open, there will not be a fact of the matter about what would have happened had we flipped the coin. There would only be a counterfactual probability. Since there is no fact of the matter about what would have happened, this probability cannot be interpreted as referring to our ignorance about what would have happened.

Now why should we be interested in underspecified conditionals as opposed to fully specified conditionals? After all, in a conditional that is specified at the microphysical level, there are no nontrivial probabilities if we assume determinism at the microphysical level. The reason is the link between contractualism and evidence-based criteria of rightness. Risk impositions are only an issue for contractualism if it is interpreted as an evidence-based criterion of rightness. If contractualism is interpreted as a fact-based criterion of rightness, a risk imposition would be wrong if and only if it leads to eventual harm. But a fact-based criterion is unhelpful in guiding the choices of agents. Evidence-based criteria, on the other hand, link moral permissibility to a choice an agent can make. They capture morality as answering deliberative questions for agents. The actions that contractualism is concerned with are therefore those that are in the choice set of an agent.⁴³ As agents, we are unable to choose the option “flip the coin with a force between 0.18345 and 0.18348 N.” This is simply not an option available to us. The option that is available to us is an option at the agential level, namely “flip the coin.” This gives us an argument for specifying conditionals at the agential level. The agential level captures the options that are available, open to the agent, whereas a microphysical level does not.

The argument for the compatibility of lower-level determinism and objective chances has another upshot. A perennial challenge to *ex post* contractualism is that it prohibits many intuitively permissible forms of risk imposition where small risks are imposed on large populations. It would seem that traffic victims have reason to reject principles that allow higher speed limits. Starting major construction works would be impermissible because of the risk of harm to workers. Air traffic may be difficult to justify because it leads to harms to bystanders.

42 See also Hare, “Obligation and Regret When There Is No Fact of the Matter about What Would Have Happened If You Had Not Done What You Did,” 190–94; and Hare, “Obligations to Merely Statistical People,” 380–82.

43 Scanlon, *Moral Dimensions*, 56–62.

The list goes on.⁴⁴ What these divergent risks all have in common is that they appear random in a relevant sense. They contrast with, for example, the risk of a lethal allergic reaction in an individual. Such an individual's death may have been difficult to prevent, but it is not random in the same sense. The aforementioned examples all appear random because none of these events is determined by the previous history of the world at the agential level. The event "person is killed in car accident" is not already determined by the past history of the world. At most, a description of the event in microphysical language is determined. This means that at the agential level, the level that counts, all the familiar examples are objectively risky. Therefore, objective *ex ante* contractualism can appealingly explain why it is permissible to impose such risks.

4.2. Identified Victim Bias

The second objection arises from the discussion concerning identified and statistical lives. *Ex ante* contractualism generally favors a bias toward identified lives and has received criticism for giving too strong an endorsement to saving identified lives over statistical lives. While this observation is broadly correct, the relationship between my version of *ex ante* contractualism and the problem of identified and statistical lives is more complex. Objective *ex ante* contractualism does not place any emphasis on the victim being identified. Rather, what is relevant is whether the victim is already determined. In a case like Dust, we do not have a way to identify the victim, but given that we have a rigid designator for the victim, we should favor her.

Indeed, my proposal can at times account for saving a statistical life rather than an identified life. For this, see a simplified version of a case by Caspar Hare.⁴⁵ You have two options: either you head north or you head south. If you head north, you will save one person for certain. If you head south, you can flip an indeterministic coin. If it lands heads, you will save another person. If it lands tails, you will save yet another person. The two potential southern victims can complain that if you head north they will die. You deprived them of a 50 percent chance to live. They can also complain that you would allocate chances to live more unequally if you were to head north. The potential northern victim can complain that by heading south you deprived her of a 100 percent chance to live. The northern victim cannot raise an additional complaint about the unfairness of the unequal distribution of chances. If we accept limited aggregation, then it seems plausible that a complaint against a 50 percent chance of death is close

44 See Norcross, "Comparing Harms," 159–67; Ashford, "The Demandingness of Scanlon's Contractualism," 298–99; James, "Contractualism's (Not So) Slippery Slope," 268–72.

45 Hare, "Obligations to Merely Statistical People," 382, 385.

enough to a complaint against a 100 percent chance of death. If this is correct, and we are permitted to aggregate the claims of the southern victims, then the added complaints against unfairness would tip the balance. It would follow, on my view, that you ought to head south and save the statistical, rather than the identified, life.

Nevertheless, the general observation is correct. *Ex ante* contractualism retains a bias against statistical lives, even though this bias is substantially weakened due to the permissibility of limited aggregation. Take, for example, the following revision of Wheel: the indeterministic roulette wheel does not release one ball but ten balls that will kill ten different persons. To many, it is difficult to accept that we should prioritize Bob Johnson's leg over multiple statistical victims. However, we should note that the individual risk for each person, while higher than in the standard version of Wheel, is still vanishingly low at one in four million.

On reflection we notice that small risks of serious harms are omnipresent. It is inevitable that large-scale policies will lead to serious harms. In many such cases of social risk, we nonetheless believe that the risk imposition is permissible. Indeed, accounting for these cases is a key challenge to *ex post* contractualism. Take, for example, the following stylized case:

Vaccine: In order to protect the entire population of California from an infectious disease, which everyone would come down with in the absence of any intervention, the government is considering a mass vaccination program. The disease is not life threatening but would cause the Californians to limp for two months, similar to the effects of a sprained ankle. While the temporary limp is much less bad than the impairment due to loss of a leg, it is significant enough that the Californians want to avoid it. In extraordinary circumstances, the vaccine can, however, be lethal, although the chance of death for each Californian is only one in four million. The government is able to administer the vaccine without intrusion on the bodies of any Californian.

Even though the policy in *Vaccine* will also lead to ten expected statistical deaths, we want to account for the permissibility of *Vaccine*. The risk of death is sufficiently small that it is outweighed by the benefit of avoiding the temporary limp. For example, according to the National Safety Council, the odds of a US resident being struck by lightning in their lifetime are a bit over one in 180,000, more than twenty-two times more likely than the harm due to the vaccine.⁴⁶ Rejecting risks of the kind involved in *Vaccine* would make it difficult to pursue many large-

46 See the overview at National Safety Council, "Odds of Dying."

scale policies or practices. The challenge is now the following. In the case of Vaccine, we prefer saving the population of California from the temporary limp over the loss of ten statistical lives. In the revised Wheel case, we prefer saving the ten statistical lives over Bob Johnson's loss of a limb. Now what if we could choose between saving the population of California from the temporary limp or saving Bob Johnson from the loss of a leg? Since the temporary limp is much less bad than the permanent loss of a leg, it is plausible that a contractualist would reject the aggregation of the complaints against the temporary limp. Hence, we should save Bob Johnson. This leads us to a preference cycle over the three options.

It is not clear how we could justify such a preference cycle. One attempt would be to point out that in Vaccine the gamble is in the *ex ante* interest of all, whereas this is not the case in the revised Wheel case.⁴⁷ This may explain why the option of "ten statistical victims when it was in their *ex ante* interest to take the risk" is not the same option as "ten statistical victims." I am not convinced that this explains our intuitions well. While it is true that the gamble is in the *ex ante* interest of all in the stylized Vaccine case, I do not believe that this is necessary to the case. I believe that delivering the vaccine would be permissible even if some small and unidentifiable part of the population were already known to be immunoresistant. The vaccine would, therefore, be neither to the *ex ante* nor the *ex post* benefit of any of them. In fact, it appears that in most cases of intuitively permissible large-scale risks, the benefits are widespread but not universal.

What the response shows, however, is that it is a mistake to frame the problem in the revised Wheel case as either saving ten people from death or saving one person from the loss of a leg. Such a framing already assumes that what matters is the harm that is the result of the risk imposition. In other words, this framing already assumes the *ex post* perspective. If my arguments against the *ex post* perspective are successful, then we should rather phrase this choice as saving the leg of one and reducing the risks of very many by a small amount. So understood, it is more plausible to maintain that it is permissible to impose the risk in the revised Wheel case.

We can give the following justification for our choice. At the time of our decision, there was no person who had as strong of a complaint as Bob Johnson did. We were able to justify our action to each of the forty million persons involved, each of whom faced only a very small risk of death. In fact, none of the forty million would have been permitted to save themselves from such a small risk if doing so had required the loss of Bob Johnson's leg. For example, each would have been required to call an ambulance to save Bob Johnson's leg even if this would have created a one in four million chance of being killed by an ambu-

47 See Walen, "Risks and Weak Aggregation."

lance sliding out of control. We can acknowledge that a better outcome could have been brought about, in which only one person loses a limb rather than ten people losing a life. But that is the sort of thing nonconsequentialists are already willing to acknowledge across a range of familiar cases. Nonconsequentialists accept that oftentimes it is impermissible to do what brings about the best outcome because doing so would violate the claims of a single individual. We can understand deontological constraints in this way.

In line with the analogy to deontological constraints, we can accept a further claim. While nonconsequentialists accept some inefficiency in terms of failing to bring about the best outcome, they typically accept that there are some cases in which deontological constraints can be overridden. Most nonconsequentialists believe that rights may permissibly be violated in cases where doing so is necessary to avoid a moral catastrophe or some other high threshold of weighty moral considerations. In those cases, even deontological constraints such as those that stand in the way of being harmfully used as a mere means can be exceptionally suspended.⁴⁸ In such cases, it can be permissible to do what would otherwise be unjustifiable to the rights holder—for example, violating the right not to be harmed as a mere means. If it is plausible that we can override the individual complaint not to be used as a mere means, then it also seems plausible that we can sometimes override the individual complaint of a determined victim against not being saved. If anything, the complaint against being used as a mere means appears to be a stronger complaint than the complaint against failing to be saved in the cases under discussion in this article.

The analogy is strengthened by a deep theoretical connection that contractualism has with a rights-based morality. Contractualism only covers a part of morality, the part that Scanlon identifies with “what we owe to each other.” This part is concerned with our relations to other persons. A natural thought is that when we act in ways that are not justifiable to a given person, we thereby wrong this person. Similarly, when we violate the right of a person, we thereby wrong this person. This suggests an important theoretical connection between contractualism and a rights-based morality, given that both are concerned with wrongs done to other persons.⁴⁹ Therefore, the idea that there is some threshold of statistical victims at which point we need to depart from contractualist morality is no more problematic than the widely accepted idea that there is some threshold of bad consequences at which point we need to depart from deontological constraints.

48 See, e.g., Nagel, *Mortal Questions*, ch. 5; and Thomson, *The Realm of Rights*, ch. 6.

49 See, e.g., Kamm, *Intricate Ethics*, 461–68.

5. CONCLUSION

In this article, I have argued for a new version of *ex ante* contractualism that focuses on the complaints that rigidly designated individuals can bring forward. Their complaints ought to be discounted by the objective probability that the harm will come about. Unlike other *ex ante* contractualists, I do not believe that we should always discount epistemic risk, nor do I believe that we should be concerned only with individuals that we can identify. Such an objective version of *ex ante* contractualism provides us with a plausible model of justifiability to each. It insists that our actions must be justifiable to everyone at the time that we act. It also insists that justification is owed to separate persons. But it does not require the use of morally superfluous identifying information that would make actual justification to each possible. Objective *ex ante* contractualism is alone in drawing a distinction between cases in which objective risks are at stake and cases in which merely epistemic risks are at stake. But far from being a defect, this is a virtue. We can thereby illuminate the morally relevant difference between luckless and doomed victims. For these reasons, I conclude that objective *ex ante* contractualism is a viable and better alternative that is theoretically superior to both epistemic *ex ante* and *ex post* contractualism.⁵⁰

Rutgers University
 steuwer@cplb.rutgers.edu

REFERENCES

- Ashford, Elizabeth. "The Demandingness of Scanlon's Contractualism." *Ethics* 113, no. 2 (January 2003): 273–302.
- Belnap, Nuel, and Mitchell Green. "Indeterminism and the Thin Red Line." *Philosophical Perspectives* 8 (1994): 365–88.

50 Versions of this article were presented at the LSE Choice Group, the LSE/UCL Risk and Aggregation Workshop, the TIF Conference at the University of Barcelona, the London Graduate Moral and Political Philosophy Workshop at King's College London, and a research seminar at the London School of Economics. I am grateful for helpful comments and suggestions by members of the audiences. I also thank Susanne Burri, David Coombs, Nicolas Côté, Christina Easton, Johann Frick, Joe Horton, Christian List, Anna Mahtani, Chris Marshall, Todd Karhu, Thulasi K. Raj, Tom Rowe, Korbinian Rüter, and multiple anonymous reviewers for very helpful discussions and comments on various drafts. I am especially grateful to Michael Otsuka and Alex Voorhoeve for extensive discussion and comments on countless drafts.

- Broome, John. "Fairness." *Proceedings of the Aristotelian Society* 91 (1991): 87–101.
- Carnap, Rudolf. "The Two Concepts of Probability." *Philosophy and Phenomenological Research* 5, no. 4 (June 1945): 513–32.
- Cohen, I. Glenn, Norman Daniels, and Nir Eyal, eds. *Identified versus Statistical Lives: An Interdisciplinary Perspective*. Oxford: Oxford University Press, 2015.
- Daniels, Norman. "Can There Be Moral Force to Favoring an Identified over a Statistical Life?" In Cohen, Daniels, and Eyal, *Identified versus Statistical Lives*, 110–23.
- Dougherty, Tom. "Aggregation, Beneficence, and Chance." *Journal of Ethics and Social Philosophy* 7, no. 2 (May 2013): 1–19.
- Eagle, Anthony. "Chance versus Randomness." *Stanford Encyclopedia of Philosophy* (Spring 2019). <https://plato.stanford.edu/archives/spr2019/entries/chance-randomness/>.
- . "Deterministic Chance." *Noûs* 45, no. 2 (June 2011): 269–99.
- Emery, Nina. "Chance, Possibility, and Explanation." *British Journal for the Philosophy of Science* 66, no. 1 (March 2015): 95–120.
- Eyal, Nir. "Concentrated Risk, the Coventry Blitz, Chamberlain's Cancer." In Cohen, Daniels, and Eyal, *Identified versus Statistical Lives*, 94–109.
- Fleurbaey, Marc, and Alex Voorhoeve. "Decide As You Would with Full Information! An Argument against *Ex Ante* Pareto." In *Inequalities in Health: Concepts, Measures, and Ethics*, edited by Nir Eyal, Samia A. Hurst, Ole F. Norheim, and Dan Wikler, 113–28. Oxford: Oxford University Press, 2013.
- Frick, Johann. "Contractualism and Social Risk." *Philosophy and Public Affairs* 43, no. 3 (Summer 2015): 175–223.
- . "Uncertainty and Justifiability to Each Person: Response to Fleurbaey and Voorhoeve." In *Inequalities in Health: Concepts, Measures, and Ethics*, edited by Nir Eyal, Samia A. Hurst, Ole F. Norheim, and Dan Wikler, 129–46. Oxford: Oxford University Press, 2013.
- Fried, Barbara H. "Can Contractualism Save Us from Aggregation?" *Journal of Ethics* 16, no. 1 (March 2012): 39–66.
- Frigg, Roman, and Carl Hoefer. "The Best Humean System for Statistical Mechanics." *Erkenntnis* 80, no. 3 (December 2015): 551–74.
- Gauthier, David. *Practical Reasoning*. Oxford: Clarendon Press, 1963.
- Glover, Jonathan. *Causing Death and Saving Lives*. London: Penguin Books, 1977.
- Glynn, Luke. "Deterministic Chance." *British Journal for the Philosophy of Science* 61, no. 1 (March 2010): 51–80.
- Hájek, Alan. "Interpretations of Probability." *Stanford Encyclopedia of Philosophy* (Fall 2019). <https://plato.stanford.edu/archives/fall2019/entries/probability-interpret/>.

- Hare, Caspar. "Obligation and Regret When There Is No Fact of the Matter about What Would Have Happened If You Had Not Done What You Did." *Noûs* 45, no. 1 (March 2011): 190–206.
- . "Obligations to Merely Statistical People." *Journal of Philosophy* 109, no. 5/6 (May/June 2012): 378–90.
- . "Should We Wish Well to All?" *Philosophical Review* 125, no. 4 (October: 2016): 451–72.
- Hoefler, Carl. "The Third Way on Objective Probability: A Sceptic's Guide to Objective Chance." *Mind* 116, no. 463 (July 2007): 549–96.
- Horton, Joe. "Aggregation, Complaints, and Risk." *Philosophy and Public Affairs* 45, no. 1 (Winter 2017): 54–81.
- James, Aaron. "Contractualism's (Not So) Slippery Slope." *Legal Theory* 18, no. 3 (September 2012): 263–92.
- John, S. D. "Risk, Contractualism, and Rose's 'Prevention Paradox.'" *Social Theory and Practice* 40, no. 1 (January 2014): 28–50.
- Kamm, F. M. *Intricate Ethics*. Oxford: Oxford University Press, 2007.
- . *Morality, Mortality*. Vol. 1, *Death and Whom to Save from It*. New York: Oxford University Press, 1993.
- Kumar, Rahul. "Risking and Wronging." *Philosophy and Public Affairs* 43, no. 1 (Winter 2015): 27–51.
- Lenman, James. "Contractualism and Risk Imposition." *Politics, Philosophy and Economics* 7, no. 1 (February 2008): 99–122.
- Lewis, David. *On the Plurality of Worlds*. Oxford: Basil Blackwell, 1986.
- . "A Subjectivist's Guide to Objective Chance." In *Studies in Inductive Logic and Probability*, vol. 2, edited by Richard C. Jeffrey, 263–93. Berkeley: University of California Press, 1980.
- List, Christian, and Marcus Pivato. "Emergent Chance." *Philosophical Review* 124, no. 1 (January 2015): 119–52.
- Loewer, Barry. "Determinism and Chance." *Studies in History and Philosophy of Modern Physics* 32, no. 4 (December 2001): 609–20.
- Lyon, Aidan. "Deterministic Probability: Neither Chance nor Credence." *Synthese* 182, no. 3 (October 2011): 413–32.
- Mahtani, Anna. "The *Ex Ante* Pareto Principle." *Journal of Philosophy* 114, no. 6 (June 2017): 303–23.
- Munoz-Dardé, Véronique. "The Distribution of Numbers and the Comprehensiveness of Reasons." *Proceedings of the Aristotelian Society* 105, no. 1 (June 2005): 191–217.
- Nagel, Thomas. *Equality and Partiality*. Oxford: Oxford University Press, 1991.
- . *Mortal Questions*. Cambridge: Cambridge University Press, 1979.

- . *The Possibility of Altruism*. Princeton: Princeton University Press, 1970.
- National Safety Council. “Odds of Dying.” Accessed November 15, 2020. <https://injuryfacts.nsc.org/all-injuries/preventable-death-overview/odds-of-dying/>.
- Norcross, Alastair. “Comparing Harms: Headaches and Human Lives.” *Philosophy and Public Affairs* 26, no. 2 (April 1997): 135–67.
- Nozick, Robert. *Anarchy, State, and Utopia*. New York: Basic Books, 1974.
- Oberdiek, John. *Imposing Risk*. Oxford: Oxford University Press, 2017.
- Otsuka, Michael. “Determinism and the Value and Fairness of Equal Chances.” Unpublished manuscript.
- . “Risking Life and Limb: How to Discount Harms by Their Improbability.” In Cohen, Daniels, and Eyal, *Identified versus Statistical Lives*, 77–93.
- Rapoport, Anatol. *Strategy and Conscience*. New York: Harper and Row, 1964.
- Rawls, John. *A Theory of Justice*. Rev. ed. Oxford: Oxford University Press, 1999.
- Reibetanz, Sophia. “Contractualism and Aggregation.” *Ethics* 108, no. 2 (January 1998): 296–311.
- Rüger, Korbinian. “On *Ex Ante* Contractualism.” *Journal of Ethics and Social Philosophy* 13, no. 3 (June 2018): 240–58.
- Scanlon, T.M. “Contractualism and Justification.” Unpublished manuscript.
- . “Contractualism and Utilitarianism.” In *Utilitarianism and Beyond*, edited by Amartya Sen and Bernard Williams, 103–28. Cambridge: Cambridge University Press, 1982.
- . *Moral Dimensions*. Cambridge, MA: Belknap Press of Harvard University Press, 2008.
- . “Reply to Zofia Stemplowska.” *Journal of Moral Philosophy* 10, no. 4 (January 2013): 508–14.
- . *What We Owe to Each Other*. Cambridge, MA: Belknap Press of Harvard University Press, 1998.
- Sher, George. “What Makes a Lottery Fair?” *Noûs* 14, no. 2 (May 1980): 203–16.
- Sober, Elliott. “Evolutionary Theory and the Reality of Macro-Probabilities.” In *The Place of Probability in Science*, edited by Ellery Eells and J.H. Fetzer, 133–61. Dordrecht: Springer, 2010.
- Spiekermann, Kai. “Good Reasons for Losers: Lottery Fairness and Social Risk.” Unpublished manuscript.
- Steuwer, Bastian. “Aggregation, Balancing, and Respect for the Claims of Individuals.” *Utilitas* (forthcoming). Published ahead of print, August 25, 2020. <https://doi.org/10.1017/S0953820820000217>.
- Strevens, Michael. “Probability out of Determinism.” In *Probabilities in Physics*,

- edited by Claus Beisbart and Stephan Hartmann, 339–64. Oxford: Oxford University Press, 2011.
- Tadros, Victor. “Controlling Risk.” In *Prevention and the Limits of the Criminal Law*, edited by Andrew Ashworth, Lucia Zedner, and Patrick Tomlin, 133–55. Oxford: Oxford University Press, 2013.
- Temkin, Larry. *Rethinking the Good*. Oxford: Oxford University Press, 2012.
- Thomson, Judith Jarvis. *The Realm of Rights*. Cambridge, MA: Harvard University Press, 1990.
- Voorhoeve, Alex. “How Should We Aggregate Competing Claims?” *Ethics* 125, no. 1 (October 2014): 64–87.
- Walén, Alec. “Risks and Weak Aggregation: Why Different Models of Risk Suit Different Types of Cases.” *Ethics* 131, no. 1 (October 2020): 62–86.

SUPERSESSION, REPARATIONS, AND RESTITUTION

Caleb Harrison

IN “SUPERSEDING HISTORIC INJUSTICE,” and in subsequent articles, Jeremy Waldron proposes and defends what he calls the Supersession Thesis.¹ According to the Supersession Thesis, circumstances might be such that the demands of justice in the present can in some sense override the demands of justice arising from cases of historical injustice.² Waldron applies the Supersession Thesis to the appropriation of aboriginal lands by white settlers throughout North America, Australia, and New Zealand, focusing on the history of wrongful appropriation of Maori lands in his home country of New Zealand. He argues that even if it is incontrovertibly true (as he thinks it is) that an injustice occurred when Maori land was wrongfully appropriated, current circumstances are such that the justified claim to reparations possessed by aboriginal groups may be superseded by the claim to a just distribution of resources possessed by the world’s existing inhabitants.³ While the central claim of the Supersession Thesis—that determinations of justice depend on circumstances—seems to be straightforwardly true, it is less clear what conclusions about reparations are entailed by this fact. Waldron’s suggestion that the Supersession Thesis entails that reparations may be superseded seems to conflate claims to restitution (a strong claim) with claims to reparation (a much weaker claim). The Supersession Thesis might entail that claims to *restitution* can be overridden by changes in circumstance, but I will argue in this paper that the thesis does not entail that claims to reparation are overridden by changes in circumstance; to the contrary, claims to reparation are quite robust to changes in circumstance.

It is worth noting that while this essay focuses on the Supersession Thesis as presented by Waldron, my primary aim is not to respond to Waldron *per se*. Rather, my primary aim is to examine the conflation of restitutive claims and

1 Waldron, “Superseding Historic Injustice,” 26.

2 Exactly in what sense historic injustice can be “overridden” by current circumstances will be explored in detail below, but the general idea should suffice for now.

3 Waldron, “Superseding Historic Injustice,” 26.

reparative claims. Waldron's presentation of the Supersession Thesis is helpful toward this end, both because it jump-started discussions of supersession and because it exemplifies the extent to which the distinction between reparative claims and restitutive claims has not been adequately addressed.

In the first section of this paper, I will examine Waldron's Supersession Thesis in detail, suggesting how we might understand the sense in which claims of historic injustice can be overridden by changes in circumstance. In the second section, I suggest that we can distinguish between *restitution* and *reparation*, and that the scope of the former is much more restricted than the scope of the latter. In the third section, I argue that the difference in scope between restitution and reparation can explain why the Supersession Thesis is unlikely to override claims to reparation, though it may help us to understand what restrictions there may be on claims to reparation. I conclude with some remarks on how this discussion might bear on contemporary issues.

1. SUPERSESSION

Imagine that three groups of interstellar explorers—*As*, *Bs*, and *Cs*—crash land on a small planet with a single landmass.⁴ Imagine further that the planet has three bundles of resources—call them *x*, *y*, and *z*—distributed throughout the landmass, and each resource bundle is sufficient for the flourishing of only one group. Given this initial setup, let us envision a few cases.

Case 1. Suppose that initially (at time t_1), each group stakes a claim to the bundle of resources immediately available to it: *A* claims *x*, *B* claims *y*, and *C* claims *z*. Given that a bundle of resources is only sufficient for the flourishing of one group, in claiming a bundle of resources each group excludes the other two groups from use of those resources.⁵ The circumstances here are not self-evidently unjust. Each group has sufficient resources for flourishing, and no group is injured by its being excluded from another group's bundle. Given the circumstances at t_1 , it would be unjust for any group to appropriate the resources of another group without that group's consent, and it is compatible with the demands

- 4 This case is adapted from similar "watering hole" cases presented in Waldron, "Superseding Historic Injustice," and his subsequent articles on supersession, e.g., "Redressing Historic Injustice," and "Settlement, Return, and the Supersession Thesis."
- 5 In saying that each group excludes other groups from their resources, I intend to be indifferent between their actively excluding individuals from the other groups and their standing ready to exclude others who encroach. I am primarily interested in isolating a case where groups have clear claims on the separate resource bundles, such that the appropriation of another group's resources would clearly be seen as such (and not merely an incidental encroachment).

of justice that any group exclude any other group from using the resources without consent.

Suppose that at some far later time (t_2), due to an unpredictable and unavoidable natural disaster, group B 's resource bundle y disappears.⁶ Without access to their resource bundle, group B will not survive.⁷ Group C 's bundle z is too far away for the B s to reach before dying, but group A 's bundle x is within reach. So, the B s encroach on the A s' resource bundle, taking enough of it to ensure their own survival. Although x is sufficient to ensure the survival of all the members of A and B combined, it is only sufficient to ensure a subsistence lifestyle—neither group will flourish as they had before. Were group A to exclude group B from their bundle, as they had at t_1 , then group B would not survive. Given the circumstances at t_2 , it would be permissible for group B to appropriate some of group A 's resources without A 's consent; after all, group B 's very survival is at stake, and appropriating resources will not threaten the survival of group A . Not only is it just for group B to appropriate some of A 's resources, but it would be an injustice for A to try to exclude B from the resources. What was a just state of affairs in the circumstances of t_1 — B s being excluded from the resources of A s—is now an unjust state of affairs in the circumstances of t_2 .

Case 2. Suppose the initial conditions at t_1 are the same as in Case 1: there are enough bundles of resources for each group to flourish while claiming exclusive rights to a bundle. Despite the conditions of plenty, suppose that group B wrongfully appropriates some of group A 's resource bundle. Now A is forced to share (at a subsistence level) x with group B ; B shares x with A , while maintaining exclusive control over their own resource y ; and group C maintains exclusive control over their own resource z . Given the circumstances here at t_1 —there are resources sufficient for each group, and each group can use their resources

- 6 We can assume in both cases that the duration between t_1 and t_2 is no less than a few generations—or whatever length of time would be necessary to ensure that no individual alive at t_2 has a direct connection to any individual alive at t_1 .
- 7 In talking about the “survival” of a group throughout this essay, I am referring to the continued living of the members of the group and not to the continued existence of the group *qua* group. This distinction is important given the nature of some of the restitutive claims made by groups. For many indigenous groups, for example, certain lands are metaphysically tied to the group identity, such that the group may not be said to survive, despite the survival of individual members of the group, if the individual members of the group are no longer able to bear important relations to the lands in question. There may be a sense in which a member of a group may not be said to survive *as herself* if she cannot bear certain relations to particular lands. For the purposes of this essay, I will count this as survival, though I do so with the intent of leaving open the question of what justice may demand in a conflict between the survival of a life and the survival of a life of *one's own*, where this latter concept may include important ties to particular geographies.

exclusively without injuring another group—*B*'s appropriation of some of *A*'s resources is an injustice.

Suppose, as with Case 1, that at some far later time (t_2), an unpredictable and unavoidable natural disaster causes *B*'s resource y to disappear. Again, *B* needs resources to survive, and *A* is the only group within survival range for *B*. Given the circumstances at t_2 , it seems to be permissible for *B* to appropriate some of *A*'s resources, regardless of the injustice of *B*'s appropriation of *A*'s resources from t_1 until t_2 .⁸ Furthermore, it also seems plausible that it would be unjust for the *As* to try to exclude the *Bs* from resource x . What was an unjust state of affairs in the circumstances of t_1 —*Bs* appropriating the resources of *As*—is now an otherwise just state of affairs in the circumstances of t_2 . Additionally, as an analysis of the permissibility of certain acts and policies, we might arrive at similar conclusions. *B*'s appropriation of *A*'s resources was previously impermissible. Now, in t_2 , similar such acts would be permissible.

Case 1 and Case 2 are both cases that are intended to demonstrate how it is that justice is tied to circumstance and how it is that circumstances can affect the ways in which we weigh temporally distant claims in our judgments of justice. In Case 1, we start with a just initial condition, and a change in circumstance changes an (otherwise) identical state of affairs—*As* excluding *Bs* from a resource—from being a just state of affairs to being an unjust one.⁹ Case 2 is more or less the converse of Case 1 and ought to make clearer what is required for supersession to occur. In Case 2, our initial condition is an unjust state of affairs: *Bs* wrongfully appropriate *As*' resources. However, a change in circumstances at t_2 changes an (otherwise) identical state of affairs—*Bs* appropriating *As*' resources—from being an unjust state of affairs to being a just state of affairs.¹⁰ Given that the initial condition is unjust, the demands of justice seem to require some form of redress. The exact nature of the required redress, whether it be restitution or reparations (or something else), will likely be determined on a case-by-case basis, but the redress in this case will need to involve at least a transfer of resources from *Bs* to *As*. However, in the circumstances of t_2 , *Bs* and *As* are both making use of the same bundle of resources, and are doing so at a subsistence level; necessarily, any transfer of resources from *Bs* to *As* will threaten the survival of *Bs*. It seems

8 At the very least, the skeletal details of the case—including the fact of *B*'s past appropriation—seem insufficiently relevant to defeat *B*'s claim to access to some of the resources.

9 Alternatively, a change in circumstances changes an (otherwise) identical case—of *Bs* appropriating *As*' resources without consent—from being unjust to being just.

10 If end-state talk is suspect here, we can arrive at similar conclusions by analyzing the permissibility of various acts and policies. The changed circumstances between t_1 and t_2 affect the permissibility of similar acts: where the *Bs* appropriating the *As*' resources at t_1 is impermissible, the *Bs* appropriating the *As*' resources at t_2 is permissible.

plausible that justice does not demand the sacrifice of the lives of individuals who are themselves innocent of wrongdoing (though they may be the unwitting beneficiaries of past wrongdoings), and so it seems plausible that justice at t_2 would not demand, and may in fact preclude, that Bs transfer resources to As.

In such cases, Waldron argues that the demands of justice in the present supersede the claims of historic injustice. As should now be clear from the cases above, supersession can be understood to take place when present demands of justice preclude what is required by the demands of justice in remedying historic injustice. Supersession is particularly salient in the context of the aforementioned appropriations of aboriginal lands by white settlers in North America, Australia, and New Zealand. While it might seem that the historic injustices of wrongful appropriations call for redress, Waldron argues that if the supersession argument goes through, it is equally likely that what would be required by the demands of justice in remedying the wrongful appropriation of aboriginal lands is superseded by the present demands of justice in ensuring the survival of innocent beneficiaries of past wrongdoings.

The Supersession Thesis seems quite plausible, and the argument that Waldron makes for it is quite strong. Interestingly, though, Waldron applies his Supersession Thesis to claims of reparations, concluding that it is reparative claims that are superseded by present demands of justice, particularly as in the aforementioned cases of wrongful appropriations of aboriginal lands. He dedicates very little space to discussion of what, exactly, reparations means. It seems, however, that a strong case can be made that supersession is best applied in cases where there exist claims of restitution, rather than reparation—or so I will argue in the penultimate section. First, however, we should examine the difference between claims of restitution and claims of reparation.

2. REPARATIONS AND RESTITUTION

The predominant grounds for reparation come from John Locke's views in *The Second Treatise of Government*. In it, he lays out a theory of reparations that falls out of his views on property and punishment, nicely summarized as follows:

Besides the Crime which consists in violating the Law and varying from the right Rule of Reason, . . . there is commonly injury done to some Person or other, and some other Man receives damage by his Transgression, in which Case he who hath received any damage, has besides the right of

punishment common to him with other Men, a particular Right to seek Reparation from him that has done it.¹¹

Three points in this passage are of particular relevance to my purposes in this paper, each of which helps us to distinguish cases where reparative claims are warranted from cases where other forms of redress might be warranted. I will highlight them here, and then address them in turn below. First, reparative claims require injury done to persons. Second, reparative claims require that the injury done to persons results from transgressions—that is, from the wrongful actions of some other person. Finally, reparation calls for redress of some sort, and this redress is owed to the injured and must be settled by the transgressor.¹²

The first point of separation between reparation and other forms of redress is that reparation necessarily involves one party injuring another. Given our set of cases above, we can imagine (contrary to the stipulation of the case) that the Bs' appropriation of As' resources (in either case) causes no harm to the As; resources are sufficiently abundant that at both t_1 and t_2 the Bs and the As are flourishing. In such a case, it may be justified to punish the Bs for the "Crime which consists in violating the Law and varying from the right Rule of Reason," but the Bs need not be required to directly redress the As in any way.

The second point of separation between reparation and other forms of redress is that reparation requires that the injury done to persons be the result of wrongful actions. Returning again to our set of cases above, we can imagine that the Bs' appropriation of As' resources *does* harm the As, but that the harm is the result of a justified action. This is the case in Case 1 at t_2 , when the Bs' very survival requires that they appropriate the As' resources, and in Case 2 at t_2 , when the circumstances have changed such that the Bs' very survival requires that they continue to appropriate the As' resources.¹³ In both cases, the As were made worse off than the Bs as a result of the Bs' actions, but in Case 2, the Bs' actions were in no way wrongful. Likewise, when a jaywalker sprints into traffic and is injured by a vehicle, so long as the driver is abiding by the appropriate traffic laws and norms, we do not think that the driver owes the jaywalker any sort of redress,

11 Locke, *Two Treatises of Government*, secs. 10–11.

12 While these three points are relevant to my argument in this paper, they are not all that can or should be said about reparations. For a more involved discussion, see Boxill, "A Lockean Argument for Black Reparations."

13 The second case mentioned here is a bit more complicated in that circumstances alter a state of affairs from including a wrongful action to including a just action. The proper analysis here is to see the state of affairs at t_1 as one that calls for reparations, while the state of affairs at t_2 is not.

because we do not think that the driver's actions were in any sense wrongful, though they were injurious.

The final point of separation between reparation and other forms of redress is that reparation requires that the redress (in whatever form it takes) be owed to those injured by the transgressor.¹⁴ When the *Bs* wrongfully appropriate the resources of the *As*, the *As* are owed reparations, and the reparation must come from the *Bs*. Suppose that at some later time (t_3) after the *Bs* wrongfully appropriated the resources of the *As*, as they did in Case 1 (at t_2), the *Cs* helped the *As* push out the *Bs*, and the *Cs* gave the *As* enough resources to place them somewhere near the level they would have been had the *Bs* not made use of the *As'* resources. In such a case, the *As* would have been compensated for their loss, certainly, but their claim to reparations would still be valid: the *Bs* would still owe the *As* something, and no amount of compensation from third parties innocent of the *Bs'* wrongdoing would change this.

It is worth pausing here to make a few last comments on the preceding points. First, it is important to note that reparations are essentially backward-looking. As we see in Locke's treatise, claims to reparation are generated by injuries resulting from wrongful action that has already occurred. What is required to satisfy a reparative claim will depend on what has transpired between the past wrong and present conditions. This can be contrasted with such forward-looking claims as compensatory claims, which will be discussed below. Forward-looking claims arise from considerations about what might be necessary to attain some future good, rather than from considerations about what might be necessary to restore someone after they have been wronged.¹⁵ That reparative claims are backward-looking leads us to a second important note regarding the content of reparative claims. Locke notes that the injured party may recover from the transgressor "so much as may make *satisfaction* for the harm he has suffered."¹⁶ One thing that might be necessary to make satisfaction for the harm one has suffered is a transfer of material resources sufficient to restore one's own material losses resulting from the wrongful action. It is not the only thing, though. An important feature of the injustice of a wrongful action is the underlying (false) assumption that the injured party has been treated in a befitting manner and that the injured party is not equal to the transgressor in worth or dignity. As such, one

14 Note that reparations might not be in the form of payment. One of the primary purposes of reparations is to restore the harmed, and this restoration might take the form of a social restoration by way of apology, rather than simply a financial restoration by means of a transfer of resources. For more, see Boxill, "Black Reparations."

15 See Boxill, "The Morality of Reparation," 117–18.

16 Locke, *Two Treatises of Government*, sec. 11, emphasis added.

other thing that might be required to make satisfaction for the harm that one has suffered is a sincere acknowledgment of error on the part of the transgressor. A claim to reparation, if it is to satisfy the injured party for the harm that they have suffered, is likely to include both material and social restoration.¹⁷

With these points in mind, it seems in the set of cases above that the *As* have a claim to reparation and that reparation is owed to the *As* by the *Bs*: the *Bs* have injured the *As*, the injury was the result of a wrongful action, and so the *Bs* must in some fashion redress the *As* for the injury the *Bs* caused. Of course, redress can come in a variety of forms, but I will focus on reparation and restitution. One way to understand the difference between reparation and restitution is in terms of their restrictions. The least restrictive form of redress is compensation, which refers broadly to any effort to offset loss. One example of compensation would be a payment by an innocent third party to a party that has suffered a loss (e.g., an insurance company paying homeowners after a flood). Claims of compensation need not necessarily imply blame, or the rightness or wrongness of an action, but merely call for the remediation of the loss of something of value. Reparation is slightly more restrictive and encompasses those forms of redress that involve some kind of transfer from the transgressor to the injured party—for example, West Germany's payment to Israel after World War II. As noted above, claims of reparation are typically restricted to the injured party and can only be fulfilled by the transgressing party. Restitution is in some sense the most restrictive form of redress.¹⁸ Restitutive claims can only be held by the party that

17 An anonymous reviewer helpfully pointed out that the distinctions I draw between restitution and reparations suggest a difference in the source of the normative force of each concept. The reviewer noted that on my account, reparation seems to address harms to well-being, while restitution seems to redress violations of legitimate title. I do not have space here to address this insight, other than to note that I did not intend to imply such a difference, nor did I explicitly rely on it in my analysis. But the reviewer helpfully unearthed what had been lurking below the surface in my thoughts—namely, that restitution and reparations can serve different roles in responding to relational breaches. Insofar as a system of title serves to distribute the rights to enjoy certain goods—and to exclude others from them—restitution serves to redress an error in that system and can redress such errors with little to no relationship between members of that system. A car thief need not even engage with me in order to return my stolen car. Well-being, on the other hand, is deeply dependent on relationships, and so redressing harms to well-being will typically require interpersonal engagement. I admit that I am not quite sure what to make of this difference, nor am I sure that I have satisfactorily articulated it, but I thank the reviewer nonetheless for challenging me to think more on this, and I hope to address it in the detail it deserves in further work.

18 Restitution might be considered less restrictive than reparation in that restitutive claims can be held by parties other than the party from whom a valuable was initially taken, and satisfied by parties other than the party that initially took the valuables, whereas reparative claims can only be held by the injured party and satisfied by the transgressing party.

has a legitimate right to that which was wrongfully taken, and restitutive claims require for their satisfaction the return of the very thing that was taken to the party that has a legitimate right to the taken thing. An example here can help illustrate the requirements of restitutive claims.

Suppose we are in Case 2 (at t_1) above, where the *Bs* have wrongfully appropriated the resource of the *As*; suppose further that this resource is land. Given that the resource in question is land, and land can (for the most part) be returned in full, it seems reasonable to think that the *As* are owed restitution in the form of a full return of the wrongfully appropriated land. Moreover, in Case 2 (at t_1), the *Bs* have full and exclusive access to their own resources, the use of which allows the *Bs* to flourish, as well as access to the resources of the *As*, the use of which by the *Bs* causes the *As* not to flourish. Given that the *Bs'* use of the *As'* land causes the *As* to fall from a flourishing lifestyle to subsistence living, it seems reasonable to think that in addition to returning the original lands to the *As*, the *Bs* owe the *As* some sort of reparation. Of course, restitution may not always be an option. Suppose we are in a modified version of the above case, where everything is identical except for the nature of the resources: instead of land, the resource is oil. Given that oil is nonrenewable, and so cannot be returned in full, the *Bs* cannot owe the *As* restitution.¹⁹ It is still the case, however, that the *Bs* are flourishing off their own oil supply while depleting the *As'* oil supply (and causing the *As* to maintain a bare subsistence lifestyle in the meantime), and so it seems reasonable to think that the *Bs* owe the *As* some sort of reparation.

One final difference between reparations and restitution lies in their responsiveness to changes in the involved parties. To be entitled to reparation, it must be the case that one is injured as a result of the wrongful action of another. For the *As* (in Case 2, say) to be entitled to reparation at any time after t_2 it must be the case that the *As* are harmed by the wrongful appropriation of the *Bs*. Suppose that by the time t_3 rolls around all the resource bundles have been depleted to the point of supporting only a subsistence lifestyle—even the previously flourishing *Bs* and the faraway *Cs* are only scraping by. Given that there is no scenario in which the *As* at t_3 are not living a subsistence lifestyle, it cannot be said that the *As* from t_3 and on are entitled to reparation.²⁰ To be entitled to restitution, however, it must simply be the case that one is the legitimate rights holder for the value in question and that the value in question can be returned in full. For instance, if we consider the *As* in Case 2 (at t_3), then so long as they are

19 Perhaps it could be said that the *As* still hold a legitimate claim to restitution, but if so, it is a claim that is conditional on the possibility of its being fulfilled.

20 If any of the *As* from t_1 or t_2 are still around, however, then they would still be entitled to reparation.

the legitimate rights holders for the wrongfully appropriated land, they are entitled to restitution—regardless of whether they have been injured as a result of the past wrongful appropriation of land by the *Bs*. This would be the case for any other legitimate rights holders too. Perhaps the *As* were in the process of selling the land to the *Cs* when the *Bs* wrongfully appropriated the land at t_1 . If so, then the *Bs* would owe the land to the *Cs*, rather than the *As* (though they may owe the *As* reparations as well, to make up for anything the *As* lost as a result of the delayed sale).²¹ Essentially, the difference is this: reparative claims are restricted to those injured by wrongful action and can be satisfied only by the transgressor, while restitutive claims are restricted to those legitimate rights holders who have been wrongfully separated from something of value and can be satisfied by whoever currently possesses the wrongfully separated valuable.

I am not the first to suggest that the notion of reparations operating in Waldron's work seems different from the notion of reparations at stake in discussions of repairing historic injustices. Over the course of a series of illuminating essays, Rodney Roberts sets out a conception of "rectification" that he argues can help make sense of the lack of interest among white Americans in combating the pernicious effects that historic injustices have had on Black Americans.²² While our accounts share some interesting similarities, they have different background assumptions and respond to different lines of inquiry.

In "Why Have the Injustices Perpetrated against Blacks in America Not Been Rectified?" Roberts notes that the notion of "reparation" that Waldron discusses is similar to the "rectification" that he is concerned with. While acknowledging that he is not "attempt[ing] anything like a full account of rectification," Roberts proffers an account that is situated within a framework in which justice has two aspects: distribution (of goods, rights, and duties) and rectification (of unjust distributions).²³ He identifies three typical features of rectification: (1) restoration, where possible (i.e., the return of that which was unjustly appropriated, or what I refer to as "restitution"); (2) compensation, where necessary (i.e., counterbalancing an unjust loss with something equivalent in value to that loss); and (3) an apology that acknowledges wrongdoing and reaffirms the moral standing of the injured. For Roberts, then, rectification is a component of rectificatory

21 Again, it may be that the restitutive claim is conditional on the possibility of its being fulfilled, but that its fulfillment is currently impossible.

22 For example, see Roberts, "Why Have the Injustices Perpetrated against Blacks in America Not Been Rectified?" "Criminalization and Compensation," and "Another Look at a Moral Statute of Limitations on Injustice." Thank you to the anonymous reviewer who recommended Roberts's work.

23 Roberts, "Why Have the Injustices Perpetrated against Blacks in America Not Been Rectified?" 357.

justice, which itself is a response to failures of distributive justice. By situating his conception of rectification within the scope of rectificatory justice, Roberts is able to employ concepts like “rectificatory compensation” to describe the efforts to replace the value of a loss unjustly suffered with something of like value.²⁴ He contrasts this with “distributive compensation,” which is called for when the distribution of rights and duties obstructs a person’s or group’s opportunity to participate in the benefits of social cooperation.²⁵ With these conceptual tools in hand, Roberts’s account seems well suited for identifying the different purposes that may underlie different compensatory policies (e.g., making up for a past injustice contrasted with correcting a maldistribution), which positions him to explain why white Americans have declined to support policies redressing the historic injustices suffered by Black Americans.

While I do not see the arguments that I am offering in this essay to be necessarily in conflict with, or duplicative of, the arguments that Roberts proffers, I do think our alternative accounts serve different purposes and consequently have different features. Where Roberts is concerned with explicating a conception of rectification that can illuminate why Black folks in America have not yet received, and are unlikely to receive, redress, my intent is to show more broadly that changes in circumstances subsequent to historic injustices are primarily—if a problem at all—a problem for what I call restitutive claims. In contrast to Roberts’s account of rectification, my notion of restitution is an alternative to, rather than a modifier of, compensation. For example, on my account, universalist efforts to close the racial wealth gap like those proposed by the People’s Policy Project are compensatory.²⁶ Such policies function by measuring the gap between Black and white wealth and implementing financial policies in such a way that the gap closes over time. These policies need not—and typically do not—make any reference to historic injustices or to replacing a good that was lost with that very good (i.e., Roberts’s “restoration” and my “restitution”). Likewise, on my account, particularized efforts to symbolically address historic injustices can be restitutive without being compensatory. For example, the return of a mis-

24 Roberts, “Why Have the Injustices Perpetrated against Blacks in America Not Been Rectified?” 357–58.

25 Roberts, “Criminalization and Compensation,” 143.

26 Author Matt Bruenig notes:

A dividend-paying social wealth fund provides a natural solution to [the problem of wealth inequality by] ... reduc[ing] wealth inequality by moving wealth out of the hands of the rich who currently own it and into a collective fund that everyone in the country owns an equal part of. It then reduces income inequality by redirecting capital income away from the affluent and parceling it out as a universal basic dividend that goes out to everyone in society. (*Social Wealth Fund for America*, 52–53)

appropriated artifact of historic or religious significance—by the museum that held it to the nation from which the artifact originated—may restore the good that was taken without serving to address any sort of imbalance the artifact's absence has caused since its initial misappropriation. Such actions would be restitutive without being compensatory. By distinguishing between various forms of redress, my account positions me to explain why changes in circumstances have different impacts on these various forms of redress.²⁷ Given these distinctions, we can reexamine Waldron's Supersession Thesis to see whether he is right to conclude that reparative claims can be superseded by circumstance.

3. CAN REPARATIVE CLAIMS BE SUPERSEDED?

The Supersession Thesis argues that the demands of justice in the present may preclude the satisfaction of reparative claims involving past harms resulting from wrongful action. Given the previously drawn distinction between reparative and restitutive claims, it seems warranted to think that the Supersession Thesis might be best applied to restitutive claims rather than reparative claims. To see why, let us return to our cases. Suppose we are in Case 2. The *Bs* wrongfully appropriated the resources of the *As* at t_1 , and (at least prior to t_2) the *As* seem entitled to redress as a result of the wrongful appropriation. This redress might take any number of forms, but it seems reasonable to expect that it would be in accord with the demands of justice that the *Bs* return the land in full (as restitution) and that they transfer resources to the *As* to redress them for whatever harms have befallen the *As* as a result of the wrongful appropriation (as reparation). Suppose, however, that before either restitution or reparation is paid, circumstances change such that we find ourselves at t_2 (the disaster has struck and the *Bs* need some of *As*' resources to survive). Circumstances at t_2 preclude restitution; restitution requires that the originally appropriated resource be returned in full to the legitimate rights holder, but there is no way to return the land to the *As* without threatening the very survival of the *Bs* in the process. The Supersession Thesis seems to hold with respect to restitution.²⁸

27 A further difference between our accounts is that the account I offer in this essay does not depend on a particular conception of justice (e.g., as divided between distribution and the rectification of maldistributions). My account should be compatible with theoretical frameworks that are concerned with, for instance, restorative justice, transformative justice, or transitional justice. See, e.g., Murphy, *The Conceptual Foundations of Transitional Justice*; Daly, "Transformative Justice"; and Walker, *Moral Repair*.

28 Cara Nine notes that there are important distinctions between property rights and territorial rights, and the relation of each to land. I have elided those distinctions here for the sake of simplicity, but they do call for brief comment. If we think of the relevant resources

It is not clear, however, that the circumstances at t_2 preclude reparation. Entitlement to claims of reparation requires that a person be injured as a result of a transgressor's wrongful action and that the transgressor owe the injured party some form of redress intended to restore the injured, either materially or socially (or both). There is no reason, given the details of Case 2 at t_2 , to think that the circumstances are such that all possible forms of financial or social restoration are incompatible with the present demands of justice. The only circumstances in which this would hold would be the circumstances in which any sort of transfer of resources whatever would threaten the survival of the *Bs*. Short of such dire circumstances, there are any number of ways that the *Bs* might redress the *As*: priority positions in the government could be reserved for *As*, special assistance programs could be instated to ensure that all *As* have access to the limited resources available, or a fund could be set up for the *As* to which the *Bs* would be responsible for contributing. In short, there is a lot that might be done to move toward full reparation—at the very least, there seems to be space for reparative action sufficient to push back against Waldron's claim that circumstances in Case 2 at t_2 are such that reparative claims are superseded by changes in circumstance.

One might object that Case 2 at t_2 is precisely a case wherein the circumstantial difference is so dire that reparations are precluded. After all, the disaster moved both groups to hand-to-mouth living—surely there is no room in such a lifestyle for any transfer of resources from *Bs* to *As*. Even if we grant that the circumstances of Case 2 at t_2 preclude a transfer of resources, reparative claims, as noted before, need not only concern a claim to the transfer of resources from

in the case as being tied to particular lands and we think of those lands in terms of property rights, where thinking of lands in terms of property rights centers our thought on claims to exclusive use or possession (e.g., Waldron, "Superseding Historic Injustice," 20), then supersession will be a live concern. If the satisfaction of a restitutive claim to land (conceived of in terms of property rights) requires that the land be transferred, for exclusive use and ownership, from, for example, the *Bs* to the *As*, then it may be the case that demands of justice would preclude such a transfer due to the threat it would pose to the survival of the *Bs*.

If we think of the relevant resources in the case as being tied to particular lands but we think of those lands in terms of territorial rights, where thinking of lands in terms of territorial rights centers our thought on claims to territorial sovereignty—or the exclusive right to make, adjudicate, and enforce laws within a region without interference from outside forces—then supersession is less likely to be a live concern. If the satisfaction of a restitutive claim to land (conceived of as territorial rights) requires that the right to territorial sovereignty be transferred from, for example, the *Bs* to the *As*, then it is unlikely that demands of justice will preclude such a transfer because such a transfer would pose no threat to the survival of the *Bs*. For more on the distinction between property rights and territorial rights, and their relevance to potential cases of supersession, see Nine, "Superseding Historic Injustice and Territorial Rights" and "Ecological Refugees, States Borders, and the Lockean Proviso."

the transgressor to the injured party. Reparative claims can also concern such socially restorative actions as sincere public apologies, and such actions are *not* precluded, even at the limit represented by Case 2 at t_2 . Reparative claims are robust to changes in circumstance, and if they are not superseded here at the limit, it is hard to imagine in which circumstances they are superseded.²⁹

Returning to Waldron's actual discussion, it is important to recall why Waldron was concerned, first and foremost, with making sense of how changes in circumstance affect the reparative claims generated by historic injustice. Waldron was particularly interested in the wrongful appropriation of aboriginal lands by white Europeans in the seventeenth to nineteenth centuries. In New Zealand in particular, several generations have passed since the wrongful appropriation of Maori lands by white settlers, and Waldron now wonders whether circumstances have changed in such a way as to preclude reparative claims by aboriginal peoples. He correctly notes that while returning land to the Maori immediately after its appropriation would have been relatively benign in effect, giving "exclusive rights [today] would mean many people going hungry who might otherwise be fed and many people living in poverty who might otherwise have an opportunity to make a decent life."³⁰ It is certainly reasonable to conclude that the injustice that would result from kicking thousands of people off of the land that they require to live an adequate life would outweigh the justice that could be satisfied by returning wrongfully appropriated land to the descendants of the injured. Still, this injustice seems only to preclude restitution, yet Waldron claims that "it has priority over reparation" (27). Drawing tighter the parallel between the cases examined above and Waldron's real-world example of wrongfully appropriated Maori lands, it seems that precluding restitution leaves wide open the possibility of satisfying reparative claims. Priority positions in the

29 One interesting line of thought that I do not have the space to explore here has to do with the nature of the reparations that would satisfy some reparative claim. Presumably, when we speak of "reparations" we speak of some set of policies that, together with their effects, will satisfy the injured party's reparative claim. Suppose the set of policies includes the following: (1) a sincere public apology, (2) structural reform, and (3) a transfer of resources. All three are necessary to satisfy this particular reparative claim, and the set of the three is sufficient to do so. Note that there is nothing requiring this set of policies to be implemented at once. Even in Case 2 at t_2 , we can imagine 1 and 2 being implemented, with 3 set for implementation once circumstances allow for it. In fact, the implementation of 1 and 2, before 3 becomes possible, might increase the likelihood that 3 actually comes about when it becomes feasible. See, e.g., the dialogue between Rodney Roberts, Laurence Thomas, and Bernard Boxill in the following essays: Roberts, "Why Have the Injustices Perpetrated against Blacks in America Not Been Rectified?" and "Toward a Moral Psychology of Rectification"; Thomas, "Morality, Consistency, and the Self"; and Boxill, "Power and Persuasion."

30 Waldron, "Superseding Historic Injustice," 26, hereafter cited parenthetically.

government could be reserved for people of Maori descent, or special assistance programs could be instated (and funded by taxes on those without Maori ancestry, or implemented as tax credits to those of Maori descent) to ensure that all Maoris have access to resources sufficient for financial or social restoration. If it is land specifically that is of concern, rather than just the harms that resulted from having land wrongfully appropriated from the Maori people, reparations might entail a program wherein Maori groups receive government assistance in purchasing any property within their initial land holdings that comes up for sale in the future. The details are not without difficulty, of course, but I hope that I have made my point clear: the Supersession Thesis may preclude restitution in the case of wrongfully appropriated Maori lands, but it seems to have little to say about reparations.

Given that there is a clear distinction between restitutive claims and reparative claims, we might wonder whether Waldron really did run the two together. The main evidence supporting the conclusion that he did can be found in his original 1992 article on the Supersession Thesis. He begins the article by explicitly stating, “The topic of this article is reparation,” before explaining that his understanding of reparation is one that recognizes that reparation has symbolic importance in addition to monetary implication (6). He goes on to note that he is considering arguments for “full and not merely symbolic reparation—a demand not just for remembrance but for substantial transfers of land, wealth, and resources in an effort to actually rectify past wrongs” (7). After referring almost exclusively to reparation throughout his article, he concludes by noting that the main claim of the Supersession Thesis only “has priority over reparation which might carry us in a direction contrary to that which is indicated by a proscriptive theory of justice” (27). Only once does the idea of restitution arise, and then in reference to the confiscation of property that has a questionable transactional history. Waldron notes that we often organize our lives and expectations around our possessions in a meaningful way and that “upsetting these expectations in the name of *restitutive* justice is bound to be costly and disruptive” (16, emphasis added). This statement is intended to support his conclusion that changes in circumstance—in this case, changes in a particular possession’s role in the lives of potential owners—can supersede what might be demanded by reparative justice. That he uses restitutive justice as an example of a reparative claim that can be superseded seems to strongly support the contention that he runs the two together.

Supposing I am right, one might reasonably concede that Waldron’s proposal was ultimately just that *restitution*—not reparations—can in some circumstances be superseded by changes in circumstance and wonder whether my essay

has any further contribution to make to the discussion. While I would be content with merely this concession, I do think my argument implicates a broader concern with assumptions underlying the Supersession Thesis. By limiting the scope of his argument to “full and not merely symbolic reparation,” or efforts “to actually rectify past wrongs,” Waldron attempts to ward off bad-faith actions that purport to address past wrongs: false apologies, insultingly low payoffs, or other conduct that actors can proffer as reparative acts that supposedly justify moving on from the historic injustices. Of course, such conduct *should* be written off as being in bad faith. But without explicitly distinguishing between reparations and restitution, as I do in this essay, I worry that the Supersession Thesis is liable to fall prey to a scenario whereby any effort to redress past wrongs that falls short of restitution would be dismissed as a bad-faith effort, and any effort that survives the bad-faith dismissal would trigger supersession.

For instance, suppose that “full and not merely symbolic reparation” required a formal apology by the state, a truly massive transfer of land and wealth, and the creation of new local and national public entities responsible for the promotion of the rights and well-being of the historically injured group. Suppose further that these could not be promoted simultaneously: no transfer of land or wealth at the appropriate scale could take place without a nationwide acknowledgment that the historic injury needed to be redressed, and no such acknowledgment could take place without a radical change in social attitudes toward the historically injured group, and no such attitudinal change could take place without a long-term education and consciousness-raising campaign at local and national levels. The Supersession Thesis suggests that only the entire bundle counts as “full and not merely symbolic reparation,” and so a change in circumstances affecting any constituent part of the bundle may trigger supersession, undermining the implementation of the whole bundle of policies. If a change in circumstances were to make a massive land transfer inconsistent with the demands of justice *right now*, then the demands of justice now would supersede any claim to land transfer, even if those claims could be warranted after implementing policies that altered present landowners’ attitudes toward such a transfer.

On my account, we can explain why land transfer now is superseded by present claims to land, while holding open the possibility that such a transfer could be a component of a broader, diachronic policy bundle. Namely, present circumstances supersede the *restitutive* claims to land, and so presently preclude such a land transfer. But present circumstances do not supersede *reparative* claims, and so a reparative package that includes the possibility of future land transfer is still a viable response to historic injustice. True, we will not know whether the reparative policy bundle that includes a transfer of land will satisfy the demands of the

present reparative claims at the moment of implementation, because it is only once the land transfer fails to transpire that we can know for sure that the call to delay land transfer was made in bad faith. But the fact that we cannot know at the moment of implementation whether a land transfer—a necessary component of a sufficient reparative policy bundle—is being delayed in bad faith does not mean that any sufficient transfer of land is necessarily superseded by the demands of justice in the present, nor does it mean that any conduct short of the transfer of land is a bad-faith act merely purporting to be reparative. Even if Waldron's Supersession Thesis is, ultimately, just a thesis about the effect of present circumstances on restitutive claims, it is still worthwhile to adopt the distinction between restitution and reparations that I offer in this essay so as to understand just how much room for reparations remains available in circumstances where restitution seems to be off the table.

4. CONCLUDING REMARKS

The Supersession Thesis has wide-ranging applicability. Waldron focuses his discussion on historic, wrongful land appropriation, but the general claim of the Supersession Thesis can be applied to any case where circumstances have changed so dramatically over time that we would be wise to pause and consider how those changing circumstances affect the shape that justice might take. One case that takes this form is the case of Black reparations in the United States. The arguments made for Black reparations come in a variety of forms, though some of the most prominent in the philosophical literature today are the counterfactual and the inheritance arguments.³¹ The cases for Black reparations that depend on demonstrating ongoing injury to Blacks in the present are less relevant to the Supersession Thesis.³² After all, the circumstances that injure Black Americans are the very circumstances in which we find ourselves; there is no sense to be made of circumstances superseding themselves. The cases for Black reparations that depend on demonstrating that Black Americans today are entitled to claim redress for harms done to their ancestors, however, seem to be highly relevant to the Supersession Thesis.³³

31 See Boxill, "Black Reparations," for more detail.

32 Here I refer to arguments for reparations on the basis of ongoing injuries caused by Jim Crow policies, racist redlining policies backed by both local and federal governments, and racist carceral practices, among others. Ta-Nehisi Coates's 2014 article "The Case for Reparations" in the *Atlantic* is a prominent example of such an argument for reparations on the basis of ongoing injury.

33 Andrew Cohen and Janna Thompson each give different arguments for how we might understand the entitlement to reparative claims held by present-day Black Americans as being

If we apply the Supersession Thesis as understood by Waldron, then it seems right to conclude that circumstances have changed such that reparative claims are superseded by the demands of justice in the present. Confiscating the land formerly belonging to slaveholders and distributing it to emancipated slaves ought to have been done post-emancipation, and it is plausible to think that present-day descendants of slaves have inherited the right to that land. Confiscating the land today, and distributing the land to the descendants of slaves, however, is almost certainly precluded by the demands of justice in the present. Millions of (otherwise) innocent beneficiaries of the historic injustice of slavery would be forced from the very land that is necessary for their well-being, and it seems implausible that such an action would be compatible with justice given the circumstances of the present. As understood by Waldron, the Supersession Thesis would preclude reparations of this form. Once we distinguish between reparations and restitution, however, we can see that the Supersession Thesis at most says that it is the restitutive claim to land that is precluded by the demands of justice in the present. Assuming that there exists a plausible argument legitimately linking present-day conditions of Black Americans to entitlements to reparations held by their slave ancestors, then it seems that any form of reparation short of restitutive land redistribution would not satisfy the conditions required by supersession and would therefore be compatible with the demands of justice in the present day.³⁴

The Supersession Thesis is an important reminder that justice depends on the circumstance, and different circumstances will lead to different results when weighing competing claims of justice. Given the relevance of circumstance to the determinations of justice then, it is of the utmost importance that we make crystal clear the nature of the competing claims involved. We must recognize the distinction between restitutive claims and reparative claims, and Waldron's version of the Supersession Thesis does not make this distinction. Once the dis-

grounded in the legitimate entitlement to reparative claims held by their ancestors. Cohen's argument relies on claims about parental duties, while Thompson's argument relies on claims about inheritance rights. See Cohen, "Compensation for Historic Injustices"; and Thompson, "Historical Injustice and Reparation."

34 The assumption that such an argument exists is one that Waldron himself attacks—he thinks that counterfactual arguments for reparations fall short, and he offers up a version of property rights that he thinks cuts the inheritance argument off at the knees ("Superseding Historic Injustice," sec. 2, esp. pp. 14–18). However, given that reparative claims are distinct from restitutive claims and that reparative claims need not be linked to any specific property, even if his version of property rights is correct, it would still only support the supersession of restitutive claims.

inction is made, however, it seems clear that though the Supersession Thesis might have the final say regarding restitution, it is silent regarding reparations.³⁵

calebharrison@icloud.com

REFERENCES

- Boxill, Bernard. "Black Reparations." *Stanford Encyclopedia of Philosophy* (Spring 2016). <https://plato.stanford.edu/archives/sum2016/entries/black-reparations>.
- . "A Lockean Argument for Black Reparations." *Journal of Ethics* 7, no. 1 (March 2003): 63–91.
- . "The Morality of Reparation." *Social Theory and Practice* 2, no. 1 (Spring 1972): 113–23.
- . "Power and Persuasion." *Journal of Social Philosophy* 32, no. 3 (Fall 2001): 382–85.
- Bruenig, Matt. *Social Wealth Fund for America*. The People's Policy Project, 2018. <https://www.peoplespolicyproject.org/wp-content/uploads/2018/07/SocialWealthFund.pdf>.
- Coates, Ta-Nehisi. "The Case for Reparations." *Atlantic*, June 2014. <https://www.theatlantic.com/magazine/archive/2014/06/the-case-for-reparations/361631/>.
- Cohen, Andrew I. "Compensation for Historic Injustices: Completing the Boxill and Sher Argument." *Philosophy and Public Affairs* 37, no. 1 (Winter 2009): 81–102.
- Daly, Erin. "Transformative Justice: Charting a Path to Reconciliation." *International Legal Perspectives* 12, nos. 1–2 (2002): 74–183.
- Locke, John. *Two Treatises of Government*. Edited by Ian Shapiro. New Haven: Yale University Press, 2003.
- Murphy, Colleen. *The Conceptual Foundations of Transitional Justice*. Cambridge: Cambridge University Press, 2017.
- Nine, Cara. "Ecological Refugees, States Borders, and the Lockean Proviso." *Journal of Applied Philosophy* 27, no. 4 (November 2010): 359–75.

35 For helping me to sharpen the arguments on offer in this essay, I would like to thank Andrew Cohen, Cara Nine, Helen Frowe, and my life-giving cohort. For inspiring this essay and introducing me to the issues of reparations in the first place, I would like to thank Bernie Boxill.

- . “Superseding Historic Injustice and Territorial Rights.” *Critical Review of International Social and Political Philosophy* 11, no. 1 (2008): 79–87.
- Roberts, Rodney C. “Another Look at a Moral Statute of Limitations on Injustice.” *Journal of Ethics* 11, no. 2 (June 2007): 177–92.
- . “Criminalization and Compensation.” *Legal Theory* 11, no. 2 (June 2005): 143–62.
- . “Toward a Moral Psychology of Rectification: A Reply to Thomas and Boxill.” *Journal of Social Philosophy* 33, no. 2 (Summer 2002): 339–43.
- . “Why Have the Injustices Perpetrated against Blacks in America Not Been Rectified?” *Journal of Social Philosophy* 32, no. 3 (Fall 2001): 357–73.
- Thomas, Laurence. “Morality, Consistency, and the Self: A Lesson from Rectification.” *Journal of Social Philosophy* 32, no. 3 (Fall 2001): 374–81.
- Thompson, Janna. “Historical Injustice and Reparation: Justifying Claims of Descendants.” *Ethics* 112, no. 1 (October 2001): 114–35.
- Waldron, Jeremy. “Redressing Historic Injustice.” In “Liberal Democracy and Tribal Peoples: Group Rights in Aotearoa/New Zealand.” Special issue, *University of Toronto Law Journal* 52, no. 1 (Winter 2002): 135–60.
- . “Settlement, Return, and the Supersession Thesis.” *Theoretical Inquiries in Law* 5, no. 2 (2004): 237–68.
- . “Superseding Historic Injustice.” *Ethics* 103, no. 2 (October 1992): 4–28.
- Walker, Margaret Urban. *Moral Repair: Reconstructing Moral Relations after Wrongdoing*. Cambridge: Cambridge University Press, 2006.

RESISTING WRONGFUL EXPLANATIONS

Arianne Shahvisi

IN THE 2017 series of the UK reality television show *The Apprentice*, a group of women discussed a sales strategy for maximizing the revenue of their burger stand in London's financial sector. Celebrity businessperson Karren Brady eavesdropped on their conversation. One contestant remarked that since the financial sector is male dominated, they should make sure that the team members chosen to sell the food are "attractive." Here, Brady cut in: "What do you mean about attractive?" The contestant, now more tentatively, responded that the salesperson must be "good at selling and ... they have to be good to sell to men, if you see what I'm saying." Brady pressed her: "No, I don't know what you're saying. What are you saying?"¹ The team of hopefuls fell silent, Brady's feigned misunderstanding of a widely understood and commonly used sexist sales strategy hanging in the now charged air. They all knew what the contestant was saying, and they knew Brady understood and was pretending not to. To explain would be to bring the sexism into the open, to commit to its assumptions, and to admit to having suggested that those assumptions be capitalized on and thereby entrenched. The contestant, precisely because Brady refused to understand, was made to confront the fact that her comment was ethically dubious.

In this paper, I develop and endorse a generalized version of the tactic of epistemic resistance that Brady deployed to expose and disarm the contestant's sexism. In doing so, I draw on the work of Gaile Pohlhaus Jr., who shows that imploring marginalized people to understand marginalizing practices amounts to a request that they legitimize their own oppression.² I expand on Pohlhaus's analysis in two novel ways. First, I rehearse what it is to understand by exploring its association with explanation. Using Van Fraassen's and Achinstein's pragmatic theories of explanation, I describe explanations as answers to why-questions and as speech acts whose success depends on the explainee revising her background

- 1 Scott Bryan, "Karren Brady Shut Down a Sexist Comment on 'The Apprentice' and It's Great," *BuzzFeed*, October 5, 2017. <https://www.buzzfeed.com/scottybryan/none-of-you-strike-me-as-shy>.
- 2 Pohlhaus, "Wrongful Requests and Strategic Refusals to Understand."

assumptions as directed by the explainer.³ The revision to the explainee's background assumptions sometimes requires the acceptance of generalizations that are ethically and epistemically troubling. In those cases, the explanation should be blocked. I advocate a variety of explanatory resistance in which the explainee feigns misunderstanding to corner the explainer into exposing or retreating from the false, harmful assumptions upon which their explanation depends. I call this strategy "disunderstanding."

Second, I situate this strategy within Fricker's epistemic injustice schema as a response to what I call "explanatory injustice," emphasizing the fact that marginalized people are not able to participate fully in the construction of explanations and are liable to be harmed by wrongful explanations.⁴ I conclude that we should be more cognizant of the way power and marginalization delimit the epistemic terrain, and be prepared to undertake resistance in order to uncloak the ensuing ethical and epistemic shortcomings.

1. STRATEGIC REFUSALS TO UNDERSTAND

Conventional wisdom has it that attempting to understand others and follow their reasoning is ethically and epistemically virtuous. Consider the Principle of Charity, whose observance is often taken to be a cornerstone of courteous, productive dialogue: "We make maximum sense of the words and thoughts of others when we interpret in a way that optimizes agreement."⁵ Pohlhaus shows that, in certain contexts, refusing to understand can be ethically and epistemically preferable. She focuses on situations in which members of oppressed groups are asked to follow the reasoning of those in privileged positions as they attempt to justify their oppressive actions. Pohlhaus shows that in such cases a listener refusing to understand can be a form of resistance, an invitation to a more productive interaction, and a way of bringing oppressive beliefs "out of the background and to the fore."⁶ This is ethically productive since it combats oppressive ideologies, and is epistemically productive because it demands the rejection or revision of false or misleading assumptions.

Pohlhaus's argument is elucidated via an example. She draws on the work of Patricia Williams, who describes the use of buzzer systems by shop owners

3 Van Fraassen, *The Scientific Image*; Achinstein, *The Nature of Explanation*.

4 Fricker, *Epistemic Injustice*.

5 Davidson, "On the Very Idea of a Conceptual Scheme," 197. The principle of charity is a content-based ideal, while "civility" is primarily a tone-based ideal for productive discussion. See Shahvisi, "Privilege, Platforms, and Power," for a critique of the expectation of civility.

6 Pohlhaus, "Wrongful Requests and Strategic Refusals to Understand," 238.

in New York City in the 1980s to screen customers and refuse entry to those who were deemed to look “undesirable,” where undesirability was primarily determined by race.⁷ The buzzer system was widely discussed and protested, but soon became standard practice in many small shops. Williams recounts the way in which the public debate was characterized by Black people being asked to *understand* the decisions of white shopkeepers and retail assistants. She refers to a letter to the *New York Times* whose white authors ask Black readers to admit that they too would exclude themselves. She refers to the “repeated public urging that blacks put themselves in the shoes of white store owners, and that, in effect, blacks look into the mirror of frightened whites [*sic*] faces to the reality of their undesirability; and that then blacks would ‘just as surely conclude that [they] would not let [themselves] in under similar circumstances.’”⁸

This case raises serious ethical and epistemic issues. Williams, and other Black people, are asked to join racist shopkeepers in rejecting themselves and accepting lines of reasoning that position them as violent and threatening in order to present racial profiling as justifiable. Williams *qua* Black person is urged to understand herself as a person who should be excluded as dangerous. Her understanding may be taken as an admission: *I can reasonably be categorized as that sort of person; your response is appropriate*. As such, she is being asked to cooperate in perpetuating a falsehood. That is the epistemic wrong.

The ethical wrong that is committed in requesting her understanding consists in the demand that she cooperate in the suppression of her subjectivity by limiting her range of action and by foreclosing the option of calling out the harm perpetuated by the understanding. If Williams agrees to understand the debate as it is presented, she must concede that she looks like a dangerous person who induces such fear in others that they cannot reasonably be asked to share enclosed public spaces with her and that she is automatically such a person by virtue of being Black. To accede to this is to limit her own epistemic possibilities. For if she agrees that it is acceptable to stereotype her and exclude her, she and other Black people are thereby hindered in challenging racism in this case and others. She is being asked to renounce a position from which she can criticize any subsequent harms, since understanding the racist assumptions would also imply an understanding of any actions premised on them. Further, consenting to be interpreted as an instance of a stereotype would mean relinquishing something of the individual agency that is required to participate meaningfully in the epistemic community. As Pohlhaus says:

7 Williams, “Spirit-Murdering the Messenger.”

8 Williams, “Spirit-Murdering the Messenger,” 129, brackets in original.

Persons are being called to understand something that only makes sense from within patterns and practices that hold oppressive power relations firmly in place and that actively prevent those asked to understand from calling attention to this fact. . . . In these cases, demonstrating the harm that the requested understanding does can only be done from worlds that actively resist the sense of the world one has been implicitly asked to inhabit.⁹

These examples demonstrate that the request that one understand can be constitutive of the marginalization one is being asked to understand. In these cases, the person who requests that the listener understand thereby wrongs the listener. And a marginalized person may not have the luxury of refusal; refusing to signal assent could escalate into more immediate harms. Yet agreeing to understand entails complicity, since the hearer is asked to join the speaker in recognizing the acceptability or inevitability of the oppressive claims, thereby entrenching their acceptability.¹⁰ One might describe this as a request that the listener *internalize* the oppression by affirming a negative self-perception of automatic wrongdoing, leading to reduced agency.¹¹ There is a double bind.

Pohlhaus's examples are not exceptional; they belong to a broader trend of requests for understanding that entreat the listener to accept oppressive assumptions. These requests need not be made of members of oppressed groups in order to be wrongful, though they are clearly *more* wrongful, and wrongful in a way that is more liable to cause harm, where the request to understand a form of oppression is made of a person whose oppression takes that form. Any request that oppressive assumptions be accepted presents ethical and epistemic concerns, and this paper considers listeners of all positionalities, since the responsibility for justice falls to all of us.

Wrongful requests for understanding are not rare. In conversations in which marginalizing comments are made, speakers typically deploy common expressions that enjoin the listener to see the sense and obviousness of what is being said and to offer comprehension (e.g., "put yourself in the shoes of"; "you have to understand that"; "surely you can see that"; "it goes without saying that"; "you know what I mean"). Consider this excerpt from a 2009 television appearance of US Fox News presenter Brian Kilmeade:

9 Pohlhaus, "Wrongful Requests and Strategic Refusals to Understand," 231–32.

10 This complicity is best understood as attributability, rather than accountability, in the sense described by Zheng in "Attributability, Accountability, and Implicit Bias."

11 Liebow, "Internalized Oppression and Its Varied Moral Harms"; Bartky, *Femininity and Domination*.

I asked [a Muslim] one time . . . “How do you feel about the extra scrutiny, clearly, you’re getting at the airports?” And he said, “I’m all for it, because I want to get home to my family, too.” And that’s really got to be the attitude. So, if you’re Islamic, or you’re Muslim and you’re in the military, *you have to understand* . . . and that’s just the fact right now in the war that was declared on us.¹²

Kilmeade requires that Muslims *understand* their racial profiling and, further, that they accordingly have the “right” *attitude* toward it—that is, one of acceptance and empathy with the assumptions underwriting the practice. Having the “right attitude” is a request for affective labor as well as understanding. Not only must oppressed people understand the oppressive practice, but they must also signal approval and strive to ensure that others do not feel bad about their (support for) oppressive behavior. Racial profiling is a common occasion for wrongful requests for understanding, where “safety” and “security” are taken to be concerns whose primacy one cannot reject without seeming unreasonable and reckless, even though the benefits and burdens are clearly unevenly distributed. Consider that more than half of British people support the racial profiling of (those who appear to be) Muslims or Arabs for “security” reasons.¹³ Those who are targeted are expected to prioritize this abstract notion of security even though doing so imperils their own more tangible personal security and comfort and that of other people of color.

In a similar way, *Daily Mail* columnist Max Pemberton invokes understandability in his appraisal of a man who murdered his own wife and daughter:

Of course, such men are often motivated by anger and a desire to punish the spouse.

But while killing their partner as an act of revenge *may be understandable*, for a man to kill his children (who are innocent bystanders in a marital breakdown) is a very different matter.

I believe it is often a twisted act of love, as the man crassly believes that the crisis in their lives is so great that the children would be better off dead.¹⁴

As feminist writer Laura Bates notes, Pemberton not only suggests that the murder of a wife may be “understandable” but also goes on to normalize the

12 Millican, Schwen, and Berrier, “What Does Brian Kilmeade Have to Say to Get Fired?” italics added.

13 YouGov and Arab News, “UK Attitudes toward the Arab World.”

14 Quoted in Bates, “A Cycle of Violence,” emphasis added.

act by distancing it from the murder of the daughter, which is a “very different matter” since she is, by contrast, “innocent.”¹⁵ Even so, the killing of the child is described as an “act of love,” if a “twisted” one, that is hypothesized to protect children from “crisis in their lives.” Pemberton urges us to see that both killings are understandable, provided one is charitable in considering the point of view of the killer.

In cases such as these, in which one is faced with a wrongful request for understanding, Pohlhaus argues that it is ethically and epistemically productive to strategically withhold comprehension. Such refusals are already in operation, as evidenced by Brady’s feigned misunderstanding in the opening example. My aim in the rest of this paper is to further analyze and systematize these forms of resistance. I begin by examining what is meant by “understanding” through a study of its relationship to explanation.

2. EXPLANATION, UNDERSTANDING, AND INJUSTICE

Wrongful requests for understanding are more easily identified and blocked if we have a clearer sense of what it means to understand. In this section, I explore pragmatic conceptions of explanation, which draw on the relationship between explanation and understanding, and then consider the ways in which explanations can be unjust. This lays the groundwork for the account of strategic refusals to understand that I introduce in section 3.

2.1. *Understanding as an Effect of Explanation*

Successful explanations impart understanding, and a person who does not understand is a person who lacks an (adequate) explanation.¹⁶ A wrongful request for understanding is therefore a request that a wrongful explanation be accepted. But what is an explanation? Since I am addressing a problem relating to the everyday use of explanations and understanding in particular social contexts, a *pragmatic* account of explanation is most apt. Pragmatic, or “contextual,” accounts of explanation are concerned with the *use* of explanations and the role of contextual factors, such as background knowledge and interests, in determining explanatory success. I explore a synthesis of two prominent pragmatic theories

15 Bates, “A Cycle of Violence.” Note that family annihilators—men who murder their families immediately prior to killing themselves—are chillingly common. In the US, these incidents occur as often as once a week (Manne, *Down Girl*).

16 Khalifa, “Inaugurating Understanding or Repackaging Explanation?”; Strevens, “No Understanding without Explanation.” Though it will not be important for the purposes of our inquiry here, note that there is a lively literature on the relationship between scientific explanations and understanding (De Regt, “Understanding and Explanation”).

of explanation: that of Van Fraassen and that of Achinstein.¹⁷ Examining these theories offers insights as to how to conceive of, and disrupt, understanding.

One of the most well-known pragmatic accounts of explanation is given by Van Fraassen, who describes a demand for explanation as a particular kind of question, and an explanation as an adequate answer to that question, which dispenses with the explainee's original explanatory demand by delivering the understanding that was sought. Though they are not always immediately framed as such, *explananda* can be read as contrastive "why-questions." That is, a request for explanation can be expressed in the following form: Why event P_k (the *explanandum*) rather than any of the alternative events in its contrast class X (P_1, P_2, \dots, P_n)? The contrast class enumerates all the other possible events that could have obtained instead of P_k . Context is critical. There are many different ways of forming the contrast class, depending on what precisely is being asked. If I see some children hitting another child in the street and say "Hey, what's going on here?" this is a demand for explanation that contains the why-question "Why are you hitting that child?" and the contrast class "rather than playing with her, leaving her alone, etc." The children must assess the context of the situation in order to ascertain that this is the contrast class I intend, and answer appropriately. They might get the contrast class wrong and say "What, you think we should kick her instead?"

Context also features in the kind of explanation that is requested. Explainers must take note of the "relevance relation" R , which encompasses other contextual factors that are relevant to providing an adequate answer. In the example above, the children must work out what kind of explanation I am asking for. They will likely assume that I am referring to the hitting, that I think that (unjustified) hitting is wrong and surprising, and that they ought therefore to attempt to *justify* the hitting, with reference to some causally relevant event ("She said something racist"), or to stop doing it in an attempt to evade the explanatory demand. If they get the relevance relation wrong, by making incorrect assumptions about my background knowledge, they might offer an explanation that is unsatisfactory—for example, "It was our turn" or "It's the first day of the month."¹⁸

One asks "Why P_k ?" when P_k is surprising, because one expected another possible state of affairs instead. That surprise is typically the result of the explainee having incomplete, inadequate, or false information relating to the occurrence of P_k . Accordingly, the explanation furnishes the explainee with information that renders P_k unsurprising, or even expectable. Therefore, to explain an event

17 Van Fraassen, *The Scientific Image*; Achinstein, *The Nature of Explanation*.

18 A British folk ritual involves starting the month by declaring "a pinch and a punch for the first day of the month!" while pinching and punching someone.

is to show that “given the particular circumstances and the laws in question, the occurrence of the phenomenon *was to be expected*; and it is in this sense that the explanation enables us to *understand why* the phenomenon occurred.”¹⁹

Making an *explanandum* expectable is the central function of an explanation. An explanation is requested when a particular occurrence seems unusual or surprising relative to an extant set of background assumptions; explaining eliminates that surprise by providing additional information that causes the explainee to revise her assumptions, making the occurrence seem ordinary. Therefore, “a (good) explanation raises or makes high its *explanandum’s* probability, *p*; and the more it does so (*ceteris paribus*) the better it is.”²⁰ Explanations therefore convert surprising facts into unsurprising facts by modifying the auxiliary assumptions of the explainee.

A second pragmatic theory of explanation is Achinstein’s description of explanations as “illocutionary acts.”²¹ This term was coined by Austin, who described speech acts as utterances that not only provide information but also *do* something. Austin distinguishes three kinds of speech acts. A “locutionary” act is simply the action of making a meaningful utterance (e.g., “You upset me”). An “illocutionary” act is the action that results from the utterance, which in the case of “You upset me” is the action by which the speaker *informs* the hearer of the effects of their behavior. A “perlocutionary” act is the effect of the utterance, which in this case might be to induce feelings of guilt or regret. In Austin’s words:

Saying something will often, or even normally, produce certain consequential effects upon the feelings, thoughts, or actions of the audience, or of the speaker, or of other persons: and it may be done with the design, intention, or purpose of producing them.... We shall call the performance of an act of this kind the performance of a perlocutionary act or perlocution.²²

While illocutionary acts focus on the function of the utterance, perlocutionary acts describe its effects. Asking “Is anyone else cold?” is on the face of it a statement about a person’s own temperature and a question about the temperature of those in the room, but it can also perform the illocutionary act of requesting permission to raise the temperature of the thermostat and might have the perlocutionary effect of someone feeling obliged to offer you their jumper.

According to Achinstein’s theory, to explain is to *do* something. Explanations

19 Hempel, *Aspects of Scientific Explanation*, 337, italics in original.

20 Mellor, “Probable Explanation,” 232.

21 Achinstein, *The Nature of Explanation*.

22 Austin, *How to Do Things with Words*, 101.

are therefore illocutionary acts: explainers set out to do something when they offer explanations.²³ They intend to make something understandable, to answer a question, to make the *explanandum* less surprising. Yet this analysis can be taken further than Achinstein does, since explanations are also *perlocutionary* acts. Recall that on Van Fraassen's theory explanations answer why-questions. They do so by reconfiguring the background assumptions of the explainee in order to make the *explanandum* expectable. But, crucially, as I will explore later, that process is one over which the explainee exercises influence.

In a similar vein to Van Fraassen's contrast class and relevance relation, which summarize the contextual information that is necessary to providing a successful explanation, Achinstein distinguishes the "goodness" of explanations from their "correctness." A correct explanation is one whose propositional content is true; a good explanation is one that is appropriate given the background knowledge and interests of the explainee. Consider that "because of the initial conditions of the universe plus the fundamental laws of nature" might seem like a reasonable causal explanation of *any* event, and it is most likely also a correct explanation, but it is (almost) never a good one. A correct explanation need not be good, and that determination depends on the explainee. If I explain to a layperson why grass is green by reference to the electron configuration around the molecular structure of chlorophyll, that is unlikely to be a good explanation even though it is correct. Equally, a good explanation may not be a correct one. If a child asks why the mince pie is gone and I say "Santa ate it," they are likely to take this to be a good explanation even though it is not a correct one. We must strive for goodness and correctness in our explanations.

The idea of a good explanation mirrors Austin's "felicity conditions," which must be met in order for a speech act to succeed.²⁴ Among other things, success requires that the listener receive the utterance in the way in which the speaker intended it. This requires the speaker to pay attention to the background knowledge of the listener, but it also requires the listener to play her role in cooperatively granting uptake. Jennifer Hornsby describes "successful illocutionary acts" as those characterized by reciprocity, where interlocutors "recognize one another's speech as it is meant to be taken: An audience who participates reciprocally does not merely understand the speaker's words but also, *in taking the words as they are meant to be taken*, satisfies a condition for the speaker's having done the communicative thing that she intended."²⁵ In the context of explanations,

23 By contrast, Hempel was concerned with the locutionary aspects of explanation ("Studies in the Logic of Explanation"; *Aspects of Scientific Explanation*).

24 Austin, *How to Do Things with Words*.

25 Hornsby, "Disempowered Speech," 134, emphasis added.

the idea of “taking the words as they are meant to be taken”—or of indicating to the person explaining that their explanation was a “good” one or that your why-question has been answered—points to the important role of the listener in ensuring both that understanding goes through and that this uptake is communicated, for example, by saying “Ah, I see” or “That makes sense now,” or by nodding and asking no further questions.

Taking all this together, we can see that a refusal to understand in Pohlhaus’s sense amounts to a refusal to accept an explanation, which can be achieved by responding to the explanation in such a way as to block the perlocutionary act of having one’s auxiliary assumptions revised in the way the explainer intends. Note that the explanation may well have made sense in some limited sense, and the explainee may well understand in some limited sense (as Brady clearly did in the *Apprentice* case), but the point here is that the explainee has spotted that something is ethically and epistemically amiss, and is accordingly *performing* the refusal as a way of encouraging the explainer, and any onlooker, to change *their* auxiliary assumptions. Whatever surprise motivated the explanatory demand then persists, and the explainer is obliged to switch course, hopefully giving more careful thought to their explanation.

Returning to Williams’s experience of racial profiling, consider that customers wishing to enter shops are normally allowed to do so. Being refused entry, or learning one might be refused entry, is surprising and requires explanation. The explanation provides some additional information to Williams: *you, by virtue of your race, are threatening, and your presence in the shop will make people uncomfortable*. Williams is supposed to respond by revising her auxiliary assumptions; while she previously thought of herself as a nonthreatening person attempting to buy a gift for her mother, she must now mitigate her surprise at being denied entry to the shop by agreeing to see herself, and other Black people, as threatening. She is supposed to reason that if a person made her fearful, she too would want them to be denied admittance to a shop she was browsing or working in. The “error” that led to her not understanding and requiring an explanation was her belief that she is *not* a scary person, that she is an ordinary, nonthreatening person. If the explanation goes through, she stands corrected. But this revision to her auxiliary assumptions must be resisted. It cannot be right to ask a person to replace their surprise at an injustice with an acceptance that they are a person who automatically deserves to be treated unjustly. To do so is to ask them to concede that what would normally count as an injustice is not so in relation to them, on the grounds that they belong to a social group for whom that harmful treatment is deemed to be apt. We must ask what happens if Williams, or an

onlooker, responds with “I’m sorry, I don’t understand” or “What is it about me/her that warrants this treatment?” I return to this question in section 3.

2.2. *Explanatory Injustice*

As we have seen, explanations act on and shape our beliefs, and they are context dependent and interest relative. Values unavoidably affect the explanations that are and are not requested, offered, and accepted. It is therefore important to briefly consider the operation of power in relation to explanations: who gets to explain and who is generally required to accept explanations offered by others. In this section, I describe how explanations, as a core knowledge-production activity, are related to “epistemic injustices.”

Epistemic injustice occurs when people are wronged specifically in their capacities as knowers.²⁶ To be wronged as a knower is to be wronged as a member of a community of people who generate knowledge by interpreting the world and sharing their interpretations with others. Epistemic injustice, as characterized by Fricker, comes in two varieties: testimonial injustice, which limits a person’s ability to share knowledge, and hermeneutical injustice, which limits a person’s ability to generate knowledge.

A testimonial injustice occurs when a person, due to identity prejudices held by listeners, has her credibility as a testifier systematically misjudged. In assessing the quality of the testimony of others, we take shortcuts based on widespread stereotypes about the social groups we believe them to belong to. Members of marginalized groups are often subject to credibility deficits and are liable not to be believed or taken seriously even when they are authorities on the topic under discussion. Conversely, members of privileged groups are often subject to credibility excesses, where they are granted authority even in relation to topics on which they have no expertise. Women experience credibility deficits relative to men, people of color relative to white people, working-class people relative to middle-class people, and nonnative language speakers relative to native speakers.

A hermeneutical injustice obtains when a group of people, due to structural prejudice in the collective interpretational resources, has some substantial part of its social experience obscured from collective understanding. They are prevented from articulating their situation by a paucity in the shared inventory of available vocabulary, conceptual frameworks, and causal models. Again, members of marginalized groups are most likely to be subject to this kind of injustice, and these lacunae can result in the inability to successfully communicate injustices that affect them particularly or overwhelmingly, which may cause distress, alienation, and cognitive dissonance, and can obstruct meaningful change. Her-

26 Fricker, *Epistemic Injustice*.

hermeneutical injustice therefore limits the agency of those it affects. Consider that before the introduction of the term “sexual harassment” in the 1970s, women were unable to effectively communicate their experiences in the workplace and struggled to pursue justice against their harassers—there was simply no widely accepted concept or terminology for the wrongs committed against them. Hermeneutical injustices occur because marginalized people are also marginalized within the processes of developing concepts and terms for understanding the social world—they are rarely the toolmakers in our knowledge economy, or at least, their tools are seldom adopted by others. And while in the long term we are all epistemically impoverished by hermeneutical injustices, since understanding the social world around us is a collective good, the actual short-term burdens are not equally shared: hermeneutical injustices are “like holes in the ozone—it’s the people who live under them that get burned.”²⁷

As we have seen, explanations act on the world, emphasizing, obfuscating, and reconfiguring our communal and individual background assumptions. They are key elements of knowledge exchange and are essential to testimony and interpretation. Yet, as with other kinds of knowledge exchange, knowers are variably situated with respect to the product and receipt of explanations. Some knowers hold a monopoly on the production of explanations, via both their perceived credibility and their access to platforms, while others do not have the credibility to be influential or successful explainers, may be more liable to be harmed by wrongful explanations, and may particularly suffer for the lack of communal conceptual resources that are required to explain their situation to others. Members of privileged social groups typically dominate roles in which one teaches, instructs, or advises others, and are vastly overrepresented within disciplines that permit them to influence public discourse: academia, law, science, journalism, and politics. These are the people that society accepts as most authoritative. They have greater control over the production and reproduction of mainstream explanations that determine the way in which we receive and understand generalizations about the world.

It is therefore instructive to introduce a specific form of epistemic injustice that I will call “explanatory injustice.”²⁸ Explanatory injustice combines ele-

27 Fricker, *Epistemic Injustice*, 161.

28 One species of explanatory injustice that has received considerable attention in recent years is “mansplaining,” which refers to instances in which a man explains something to a woman, where: (a) he uses a condescending tone, and (b) she already knows about, or is positioned to know more about, the *explanandum* in question. Mansplaining exhibits testimonial injustice: it requires a man to have estimated his own credibility on the topic in question to be greater than his woman interlocutor’s. Likewise for whitesplaining, in which a white person condescendingly explains racism to people of color. Whitesplaining and mansplaining

ments of both testimonial and hermeneutical injustice. The act of explaining is affected by testimonial injustice, since those who are granted the platforms necessary for an explanation to be heard, and the credibility for an explanation to be believed, are generally those from privileged social groups. This tends to give members of privileged groups a monopoly on explaining. They are liable to produce and disseminate explanations that serve their own interests and agendas, or at least, they are unlikely to be able or willing to generate explanations that serve the explanatory needs of those from marginalized groups in challenging their marginalization. Accordingly, members of marginalized groups lack the conceptual resources to explain their experiences or dispute explanations that relate to them, which is a form of hermeneutical injustice. The explanations most widely and weightily circulated in our explanation economy are therefore those curated by more privileged people. The use of the word “economy” is significant: explanations compete with one another, and the victors are invariably those explanations proffered by the most dominant explainers that best fit with our extant background assumptions, as shaped by dominant explainers.

If particular groups have greater power over the creation and distribution of explanations, then explanations that favorably represent their interests are likely to gain traction and explanations that represent the interests of other groups are liable to be quashed where they contradict more popular explanations. The background assumptions that are the substrate of explanations are therefore liable to be epistemically faulty (i.e., false, superficial, or misleading) where they refer to the properties or experiences of marginalized groups. In the next subsection, I describe how the epistemic and ethical status of explanations is also undermined by the nature of the generalizations upon which they rely.

2.3. *When Explanations Go Wrong: The Trouble with Generalizations*

I have shown (in section 2.1) that explanations act on the background assumptions of the explainee. The explainee is given new information, or alerted to information she knew but did not consider relevant, in light of which her why-question is answered and she understands something she did not understand before. Explanations that relate to the social world draw on background assumptions that cite social patterns and generalizations. But things can easily go wrong, both because of the identities of those most likely to generate and distribute generalizations (section 2.2) and because of the nature of the background assumptions themselves. In this subsection, I discuss the latter.

Generalizations are the staple of the background assumptions whose re-

provoke ire for marginalized people because they emphasize the monopoly that privileged people have over explanation and because they tend to silence and abase the listener.

sion explanation consists in. As Langton says, if we wish to understand patterns and objects in the social world, we must look “for the regularities that reveal them in normal circumstances.”²⁹ In other words, we must look for the patterns that the object of interest is implicated in and make generalizations accordingly. We may then base our explanations on those generalizations. Yet Langton warns that things may not always be so simple, since in “abnormal circumstances things may be distorted, and the regularities we see may not reveal their natures.”³⁰

It seems that wrongful requests for understanding occur precisely in those “abnormal circumstances.” In many cases, the regularities we observe do not reflect the nature of things, rather the “world ‘arranges itself’—at least in part—to fit what the powerful believe.”³¹ Explanations can cite superficial regularities and still function as good and even correct explanations. And if no further explanation is requested or offered, sometimes the superficially true regularity that is referenced is taken to be descriptive of the *nature* of the social objects under study.

Recall that explanatory demands can be framed as why-questions. And the answer to a why-question can also be cast as a why-question, producing a regress, where each new answer precedes the last along a causal chain. Where on that chain an explanation stops depends on the knowledge and interests of the explainer and the explainee, and not all stopping points are equal. If I ask why women, rather than men, tend to wear makeup, and I am told that women are generally more concerned about their appearance and being regarded as beautiful, that is not an *incorrect* answer in terms of its veracity with respect to our social world (it could be argued to be correct *and* good in Achinstein’s sense), but it is a *truncated* one. That truncation is epistemically and ethically troubling because it *implies* something false and damaging about the nature of women. I could follow up by asking “Why are women generally more concerned about their appearance and being regarded as beautiful?” At some point, a patient, well-informed interlocutor would have to describe the objectification and sexualization of women, but we may never get there if I take the first explanation at face value. Therefore, the explainee plays a critical role in *what sort* of explanation is given and in whether that explanation itself needs explaining.

These harmful explanatory truncations often rely on nonquantified generalizations, also known as generics. Generics, such as “women are superficial” or “Muslims are terrorists,” do not specify how many instantiations are necessary for their truth. They are often accepted even at low prevalence rates, especially

29 Langton, “Feminism in Epistemology,” 142.

30 Langton, “Feminism in Epistemology,” 142.

31 Langton, “Feminism in Epistemology,” 140.

in the case of “striking property generics,” which pick out a property that is dangerous or considered to be particularly characteristic of members of that group.³² Generics play a central role in our communication and understanding of patterns in the world, yet we struggle to formalize their truth conditions, which is critical to their ability to mislead despite seeming correct.³³ We have a tendency to erroneously interpret and evaluate quantified statements as generics (e.g., it is widely noted that people tend to accept “all ducks lay eggs” despite knowing that drakes do not, implying it has instead been taken to mean “ducks lay eggs”), and studies show that people are willing to accept generics based on prevalence levels as low as 10 percent, yet when presented with a generic, infer prevalence estimates as high as 100 percent.³⁴ Generics are frequently used in common parlance, as they offer succinct, memorable heuristics for navigating the social world, while quantified statements can be clumsy and require greater sophistication. Indeed, infants as young as thirty months have already acquired certain generics, and generics are frequently used to teach children about the world through straightforward patterns and associations (e.g., “dogs go woof”; “boys do not cry”).³⁵

Of the many errors in reasoning associated with generics, perhaps the most worrying is that where they attribute properties to members of a particular social group, listeners tend to essentialize that group—that is, to assume that members of that group have those properties *by nature*.³⁶ Generics have false implicatures of naturalness, which Haslanger recommends be blocked via metalinguistic negation—for example, “It’s harmful and misleading to say Muslims are terrorists; there are extremists in every social group. Lots of Muslim-majority states have been destroyed by Western imperialism, which pushes people toward extremism, and Muslims have been portrayed particularly unfairly by Western politicians and the media.” Identifying and blocking erroneous or misleading statements in this way is important because otherwise:

Implicatures and presuppositions of this sort become part of the common ground, often in ways that are hard to notice and hard to combat,

32 Prasada et al., “Conceptual Distinction amongst Generics”; Leslie, “The Original Sin of Cognition.”

33 Saul, “Are Generics Especially Pernicious?”

34 Leslie, Khemlani, and Glucksberg, “Do All Ducks Lay Eggs?”; Leslie and Lerner, “Generic Generalizations”; Cimpian, Brandone, and Gelman, “Generic Statements Require Little Evidence for Acceptance but Have Powerful Implications.”

35 Graham, Nayer, and Gelman, “Two-Year-Olds Use the Generic/Nongeneric Distinction to Guide Their Inferences about Novel Kinds.”

36 Leslie, “The Original Sin of Cognition.”

and they become the background for our conversations and our practices. Once the assumption of, e.g., women's submissive nature has been inserted into the cultural common ground, it is extremely difficult and disruptive to dislodge it. . . .

It is not the case that women are submissive, even if most women are submissive, in fact, even if all women are submissive, because submission is no part of women's nature.³⁷

Most explanations are built on generalizations, some of which falsely imply the innate, immutable natures of particular social groups. In particular, when racism or sexism are questioned, explainers are liable to use generalizations that cite the purported properties of particular groups (e.g., "Black people are dangerous"; "women are better at child-rearing"). Generics are not the only troubling generalizations, and quantified generalizations can be just as concerning (e.g., "most men are bad at cleaning"). Generalizations of all kinds have the persuasive advantage of referring to what is immediate and simple rather than distant and complex, and they therefore benefit from the tyranny of the face value. Consider how Ockham's razor—which is informally encouraged as a heuristic in the practice of science and medicine, and is sometimes referred to in everyday conversation—directs us toward the explanation that draws on the simplest hypothesis, which usually means the one that is most parsimonious in its assumptions. One might crudely apply this principle to erroneously conclude that the reason a disproportionate number of Black people have died from coronavirus is because they have some genetic susceptibility, rather than because of the complex, interlocking determinants of health in a racist society. Our proclivity for simplicity must be closely scrutinized. Indeed, it has been demonstrated empirically that explainers prefer explanations that draw on relationships that are stable across changing circumstances, which might be taken to be a preference for explanations that rest on stereotypes.³⁸ Moreover, generics and other generalizations are often not, strictly speaking, false, even if their implicatures are. This leads to the worrying realization that explanations can go wrong even without the construction of outright falsehoods or any violation of the rules of explaining, so that correct explanations can nonetheless be misleading and can lead to the entrenchment of *incorrect* assumptions. The explanatory process is itself warped by the substrate of assumptions of the explainer and explainee, either deliberately or incidentally.

Strategic refusals to understand must be attentive to these errors of reasoning and cognizant of the opportunities presented by the collaborative nature of ex-

37 Haslanger, "Ideology, Generics, and Common Ground," 198–99.

38 Lombrozo, "Causal-Explanatory Pluralism."

planations in exposing and disarming harmful assumptions. In the next section, I suggest one such strategy.

3. EXPLANATORY RESISTANCE: THE CASE FOR DISUNDERSTANDING

According to the pragmatic theories of explanation described in section 2.1, an *explanans* that does not explain is not an explanation. Importantly, the explainee is the primary adjudicator of whether the explanation succeeds. The process of explanation is therefore collaborative and dynamic, and the explainee plays a key role in the negotiation and its results. As such, the explainee might subvert this role as a way of resisting explanations that are morally troubling. Rejecting the *explanans* preserves the explanatory burden on the explainer, forcing her to elaborate or provide another *explanans*.

What I am describing does not amount to *misunderstanding* the explainer. A misunderstanding is a *genuine* failure to understand. In cases of interest, the listener is capable of understanding by accepting the explanation at face value with respect to the social norms with which it is offered, but refuses to do so, since the request for understanding and the agreement to understand are wrongful. The explanation may be a good one in Achinstein's sense (and it may even be correct in some superficial sense), but it is not a good one in an ethical or epistemic sense. Consider that when a racist or sexist joke is made, the joke is usually well understood even by those who find it morally troubling—indeed, they may not be able to see why it is morally troubling *unless* they understand it—to the extent that they may see why it is funny or might find it funny in spite of their awareness of its harms.³⁹ As Bergmann says, “Being aware of a [racist or sexist] belief is not the same as holding it.”⁴⁰ A person can understand, but wish they did not. Misunderstanding therefore does not adequately capture the *deliberate* maneuver that I am recommending; the explainee understands, but strategically pretends not to. One must understand in order to pretend not to, so that the tactic of strategic refusal I am describing is only available to those who see the sense of the explanation relative to the social world we live in but want to reconfigure that world so that it has greater epistemic and ethical integrity.

Let us instead refer to the form of resistance under study as “disunderstanding” (a portmanteau of “deliberate” and “misunderstanding”). An explainee misunderstands when she feigns ignorance by pretending that a good explanation in Achinstein's sense is a bad explanation, thereby preserving the explanatory demand, or by producing a novel feigned explanatory demand in relation to a

39 Consider, though, Anderson, “Racist Humor.”

40 Bergmann, “How Many Feminists Does It Take to Make a Joke?” 74.

troubling action or utterance. Disunderstanding demands an interrogation of explanations that are ethically and epistemically troubling. It urges the explainer to expose the dubious assumptions underwriting their explanation. The rejection can take various forms: an outright refusal to accept the revision to background assumptions (“No, I don’t get what you mean by ‘America for Americans.’”); the conversion of the explanation into a new *explanandum* (“But why do you think that searching Muslims will make you safer?”); an articulation of unexpected surprise in response to an utterance or action that is normally taken as unsurprising (“Why did you expect your wife to adopt your surname?”); or an interrogation of the wrongful surprise of others (“Why would you assume I [a person of color] like spicy food?”).

Explanations for marginalizing actions or perspectives often rely on vague, euphemistic phrases (e.g., “we need to take back control”; “we’re losing our family values”). These phrases can act, without the need to say anything overtly troubling, as dog whistles that resonate strongly with those receptive to particular ideologies. Disunderstanding tries to force explainers to meet the explanatory demand by asking them to explicitly state what lies beneath their explanation, thereby exposing those ideologies, which leaves their harmful assumptions vulnerable to direct critical attention.⁴¹ As Saul says in relation to challenging generics, we need to “press people to spell out their evidence for their generic claims and to reflect on what that evidence really does or doesn’t warrant.”⁴²

Let us return to Patricia Williams’s experience of being refused entry to a shop on account of being Black in order to see how this might work in practice. Imagine I have observed this refusal and decide to challenge the shopkeeper:

Why didn’t you let her in?

We have a buzzer system so we can keep our customers safe.

What made you think she was a threat to our safety?

Well, we’ve been told to not let certain kinds of people in.

What kind of people?

People who look like they might cause trouble.

You only got a quick glance at her. What was it about her that made you think she would cause trouble?

41 It is similar to the strategy of “pedantry” that Elisabeth Camp recommends to disrupt troubling insinuations (“Insinuation, Common Ground, and the Conversational Record”).

42 Saul, “Are Generics Especially Pernicious?” 14.

And so the regress goes on, and the shopkeeper (assuming they are reasonably cooperative in this exchange and do not instead turn nasty or refuse to discuss the matter) is cornered toward an ever more uncomfortable position of feeling rumbled and forced to grapple with their own assumptions or those of their managers.

Disunderstanding already happens, and its results are mixed. In the ideal case, the explainer sees a couple of moves ahead, spots their troubling assumptions, and recants them, or falls silent, feeling checkmated. This is what happened in the *Apprentice* example that opened the paper. Often, people become defensive and withdraw from any meaningful discussion, or they become aggressive, as is so often the case in online discussions. Even then, an important intervention has often been made from which onlookers may benefit, and the troll may be deterred in the future. Some people do not spot the trap or their moral failings and are happy to follow their troubling assumptions all the way down. In those cases, one can only hope for a more astute witness to benefit from the intervention, but there can still be value in informing the explainer of their shortcomings.

One might wonder whether it is more effective to simply say “that is racist/sexist” rather than opting for this playful strategy. There are certainly situations in which that would be more appropriate. If a student were to make a blunt, run-of-the-mill racist comment, or use an obviously racist phrase, it would be more apt to cut them down with a comment like “that is racist and we do not tolerate racism in this classroom” (with the offer to talk them through this in private after class) rather than to painstakingly tease out their assumptions while other students are potentially harmed by the ambiguity in the instructor’s position and strategy. Disunderstanding should be reserved for cases in which the process of realization will be valuable to the explainer and any onlookers, and where taking that route is likely to result in a more robust and self-directed reevaluation of their position. If a student makes a more complex or obscure racist or sexist statement or insinuation, as is more commonly the case, it would be more educational to use disunderstanding to *show* them and their peers where they have gone wrong, rather than to merely *tell* them. If they spot their error before the instructor names it, they are less likely to feel humiliated or become defensive.

The same holds for cases outside the classroom: disunderstanding is generally the wrong tool to deal with direct, violent, intentionally hateful acts and the people who commit them, but it is a powerful way of helping people to spot a moral shortcoming (or cornering them into seeing one as such) and allowing them to follow the reasoning that will help them to call out the wrong in others. Ironically, disunderstanding, as opposed to merely condemning, might help people to *understand* their wrongdoing, which is a more robust way of en-

couraging change. Further, if the wrongful explanation proceeds via a generic, say, “women care about how they look,” it might seem opaque or irrelevant to counter it with “that is sexist” and it would appear *more false* to say “no, they do not,” because in many cases, they clearly do. In this sense, disunderstanding can function as a dialogic form of metalinguistic negation, in which the explainee’s prerogative to determine whether an explanation answers the why-question (as Van Fraassen demands) or is good (in Achinstein’s sense) can be put to use in subverting the dynamic and explaining to the explainer. The explainee could disunderstand by responding with “Sure, but why *do* women care so much about how they look?” which puts the conversation on course to reveal a more overtly sexist belief that may be apt for straightforward negation, or to arrive at the social origins of women’s anxieties.

Disunderstanding need not be verbal. It may consist of behaviors that enact a refusal to follow rules that are supported by a widely accepted explanation. Consider Claudette Colvin’s and Rosa Parks’s deliberate refusals to understand the rules concerning segregated seating on buses, forcing the racist policy into overt discussion. Lawmakers were forced either to attempt to explain their reasoning more forcefully or to change the law. Queering one’s performance of gender is also in some cases a deliberate refusal to understand gender essentialism or the sex-gender binary.

It is important to note that the method of resistance I am recommending could be, in a different social context, a harmful practice. Kristie Dotson describes the way in which marginalized speakers discussing their marginalization are often silenced by audiences precisely because successful communication requires “an audience willing and capable of hearing [them].”⁴³ If an audience refuses uptake to a speaker because of ignorance (intentional or otherwise) or because the credibility of the speaker has been deflated (i.e., testimonial injustice), an explanation will be blocked in ways that are epistemically and ethically troubling. It is therefore important to specify the scope of disunderstanding, which is intended as a purposeful tool for resisting marginalization, rather than as a way of reinscribing power. It is doubtful that disunderstanding could be successfully misused in this way, since oppressive utterances and actions have a habit of resting on various unstated, harmful assumptions in ways that anti-oppressive expla-

43 Dotson, “Tracking Epistemic Violence, Tracking Practices of Silencing,” 238. Similarly, Cull describes “dismissive incomprehension,” whereby a listener feigns incomprehension of another person’s speech in order to discredit them in the eyes of others. Gaslighting is an extreme form of dismissive incomprehension in which the speaker is made to doubt her own credibility (“Dismissive Incomprehension”). See also Kukla’s “discursive injustice” (“Performative Force, Convention, and Discursive Injustice”) and Langton’s “illocutionary silencing” and “perlocutionary silencing” (“Speech Acts and Unspeakable Acts”).

nations do not, but the analysis in this paper may shed light on the ways in which something akin to disunderstanding is attempted by those who intend harm.

Of course, a person's ability to resist an explanation, like her ability to resist *simpliciter*, is critically dependent on her positionality. Those who are most likely to identify the need to resist are also those for whom resistance might turn out to be costliest or most dangerous. Further, due to widespread epistemic prejudices, they are also most likely to be deemed to genuinely *misunderstand* and require additional instruction from a person who deems himself to have greater expertise.⁴⁴ There is an irony here: disunderstanding often requires that a person who is seen as less knowledgeable *perform* ignorance in order to force the person who is seen as more knowledgeable to face up to their *actual* ignorance. These points emphasize the importance of allyship: those who are relatively privileged are best placed to practice disunderstanding in order to erode harmful assumptions and reduce the likelihood that those who are directly affected will encounter them. This requires those with greater comparative privilege (particularly in an epistemic sense) to be continually attentive to the teachings of oppressed people, so that when the moment arises, they will be equipped to sense the underlying assumptions and be prepared to excavate them.

Accepting an explanation that is marginalizing requires that a person revise (or appear to revise) their background assumptions to accommodate content that produces, entrenches, or ignores injustices. Conversely, if a person outwardly refuses to understand, the speech act is thwarted, and the explanatory demand persists, forcing the explainer to more carefully scrutinize their own assumptions in the course of attempting to generate an alternative explanation. Disunderstanding blocks wrongful requests for understanding and can therefore contribute to destabilizing explanatory injustice.

4. CONCLUSION

In this paper, I have built on the work of Pohlhaus to show that understanding, which is conferred as a result of successful explanations, is not always ethically and epistemically virtuous. I have formalized a way of refusing to understand by disrupting the successful operation of an oppressive explanation. A successful explanation requires that the explainee's why-question be answered by the perlocutionary act of revision to her background assumptions and that she communicate that success. Importantly, this means that explanations are collaborative, and the explainee plays a critical role in the process of explaining; an explanation is not successful unless an explainee deems it to be so. By feigning misunder-

44 Consider mansplaining. See note 28.

standing in specific ways, explainees can therefore disrupt oppressive explanations that rely on problematic assumptions and generalizations with false implicatures. As such, there is epistemic and ethical merit in the explainee subverting her expected role as the cooperative recipient of an explanation, in strategically *disunderstanding* in order to force the conspicuousness and interrogation of marginalizing epistemologies.⁴⁵

Brighton and Sussex Medical School
a.shahvisi@bsms.ac.uk

REFERENCES

- Achinstein, Peter. *The Nature of Explanation*. Oxford: Oxford University Press, 1983.
- Anderson, Luveell. "Racist Humor." *Philosophy Compass* 10, no. 8 (August 2015): 501–9.
- Austin, J. L. *How to Do Things with Words*. 2nd ed. Edited by J. O. Urmson and Marina Sbisa. Oxford: Oxford University Press, 1975.
- Bartky, Sandra Lee. *Femininity and Domination: Studies in the Phenomenology of Oppression*. New York: Routledge, 2015.
- Bates, Laura. "A Cycle of Violence: When a Woman's Murder Is Called 'Understandable.'" *Guardian*, July 26, 2016.
<https://www.theguardian.com/lifeandstyle/womens-blog/2016/jul/26/womans-murder-called-understandable-lance-hart>.
- Bergmann, Merrie. "How Many Feminists Does It Take to Make a Joke? Sexist Humor and What's Wrong with It." *Hypatia* 1, no. 1 (Spring 1986): 63–82.
- Camp, Elisabeth. "Insinuation, Common Ground, and the Conversational Record." In *New Work on Speech Acts*, edited by Daniel Fogal, Daniel W. Harris, and Matt Moss, 40–66. Oxford: Oxford University Press, 2018.
- Cimpian, Andrei, Amanda C. Brandone, and Susan A. Gelman. "Generic Statements Require Little Evidence for Acceptance but Have Powerful Implications." *Cognitive Science* 34, no. 8 (November 2010): 1452–82.

45 I wish to thank audiences at the Five College Women's Studies Research Center and the Program for the Study of Women and Gender at Smith College, Boston University's Colloquium for the Philosophy of Science, the Decolonising Science Society at King's College London, the CRASH program at the University of Cambridge, and the 2019 Society of Women in Philosophy's annual conference for their feedback on earlier drafts. I would also like to thank Lisa Armstrong, Chanda Prescod-Weinstein, Kristie Dotson, Martina Rosola, and Neil Singh.

- Cull, Matthew J. "Dismissive Incomprehension: A Use of Purported Ignorance to Undermine Others." *Social Epistemology* 33, no. 3 (2019): 262–71.
- Davidson, Donald. "On the Very Idea of a Conceptual Scheme." In *Inquiries into Truth and Interpretation*, 183–98. Oxford: Oxford University Press, 2001.
- De Regt, Henk W. "Understanding and Explanation: Living Apart Together?" *Studies in History and Philosophy of Science Part A* 44, no. 3 (September 2013): 505–9.
- Dotson, Kristie. "Tracking Epistemic Violence, Tracking Practices of Silencing." *Hypatia* 26, no. 2 (Spring 2011): 236–57.
- Fricke, Miranda. *Epistemic Injustice: Power and the Ethics of Knowing*. Oxford: Oxford University Press, 2007.
- Graham, Susan A., Samantha L. Nayer, and Susan A. Gelman. "Two-Year-Olds Use the Generic/Nongeneric Distinction to Guide Their Inferences about Novel Kinds." *Child Development* 82, no. 2 (March/April 2011): 493–507.
- Haslanger, Sally. "Ideology, Generics, and Common Ground." In *Feminist Metaphysics: Explorations in the Ontology of Sex, Gender and the Self*, edited by Charlotte Witt, 179–207. Dordrecht: Springer, 2011.
- Hempel, Carl G. *Aspects of Scientific Explanation and Other Essays in the Philosophy of Science*. New York: Free Press, 1965.
- Hempel, Carl G., and Paul Oppenheim. "Studies in the Logic of Explanation." *Philosophy of Science* 15, no. 2 (April 1948): 135–75.
- Hornsby, Jennifer. "Disempowered Speech." *Philosophical Topics* 23, no. 2 (Fall 1995): 127–47.
- Khalifa, Kareem. "Inaugurating Understanding or Repackaging Explanation?" *Philosophy of Science* 79, no. 1 (January 2012): 15–37.
- Kukla, Rebecca. "Performative Force, Convention, and Discursive Injustice." *Hypatia* 29, no. 2 (Spring 2014): 440–57.
- Langton, Rae. "Feminism in Epistemology: Exclusion and Objectification." In *The Cambridge Companion to Feminism in Philosophy*, edited by Miranda Fricke and Jennifer Hornsby, 127–45. Cambridge: Cambridge University Press, 2000.
- . "Speech Acts and Unspeakable Acts." *Philosophy and Public Affairs* 22, no. 4 (Autumn 1993): 293–330.
- Leslie, Sarah-Jane. "The Original Sin of Cognition: Fear, Prejudice, and Generalization." *Journal of Philosophy* 114, no. 8 (August 2017): 393–421.
- Leslie, Sarah-Jane, and Adam Lerner. "Generic Generalizations." *Stanford Encyclopedia of Philosophy* (Winter 2016). <https://plato.stanford.edu/archives/win2016/entries/generics/>.
- Leslie, Sarah-Jane, Sangeet Khemlani, and Sam Glucksberg. "Do All Ducks Lay

- Eggs? The Generic Overgeneralization Effect." *Journal of Memory and Language* 65, no. 1 (July 2011): 15–31.
- Liebow, Nabina. "Internalized Oppression and Its Varied Moral Harms: Self-Perceptions of Reduced Agency and Criminality." *Hypatia* 31, no. 4 (Fall 2016): 713–29.
- Lombrozo, Tania. "Causal–Explanatory Pluralism: How Intentions, Functions, and Mechanisms Influence Causal Ascriptions." *Cognitive Psychology* 61, no. 4 (December 2010): 303–32.
- Manne, Kate. *Down Girl: The Logic of Misogyny*. Oxford: Oxford University Press, 2017.
- Millican, Julie, Christine Schwen, and Justin Berrier. "What Does Brian Kilmeade Have to Say to Get Fired?" Media Matters for America, October 15, 2010. <https://www.mediamatters.org/fox-friends/what-does-brian-kilmeade-have-say-get-fired>.
- Mellor, D. H. "Probable Explanation." *Australasian Journal of Philosophy* 54, no. 3 (December 1976): 231–41.
- Pohlhaus, Gaile, Jr. "Wrongful Requests and Strategic Refusals to Understand." In *Feminist Epistemology and Philosophy of Science: Power in Knowledge*, edited by Heidi E. Grasswick, 223–40. Netherlands: Springer, 2011.
- Prasada, Sandeep, Sangeet Khemlani, Sarah-Jane Leslie, and Sam Glucksberg. "Conceptual Distinctions amongst Generics." *Cognition* 126, no. 3 (March 2013): 405–22.
- Saul, Jennifer. "Are Generics Especially Pernicious?" *Inquiry* (2017): 1–18.
- Shahvisi, Arianne. "Privilege, Platforms, and Power: Uses and Abuses of Academic Freedom." In *Enforcing Silence: Academic Freedom, Palestine and the Criticism of Israel*, edited by David Landy, Ronit Lentin, and Conor McCarthy, 310–30. London: Zed Books, 2020.
- Strevens, Michael. "No Understanding without Explanation." *Studies in History and Philosophy of Science Part A* 44, no. 3 (September 2013): 510–15.
- Van Fraassen, Bas C. *The Scientific Image*. Oxford: Oxford University Press, 1980.
- Williams, Patricia. "Spirit-Murdering the Messenger: The Discourse of Fingerprinting As the Law's Response to Racism." *University of Miami Law Review* 42, no. 1 (1987): 127–57.
- YouGov and Arab News. "UK Attitudes toward the Arab World." 2017. https://www.arabnews.com/sites/default/files/pdf/topline_pdf_finalx.pdf.
- Zheng, Robin. "Attributability, Accountability, and Implicit Bias." In *Moral Responsibility, Structural Injustice, and Ethics*, edited by Michael Brownstein and Jennifer Saul, 62–89. Vol. 2 of *Implicit Bias and Philosophy*. Oxford: Oxford University Press, 2016.

POLITICAL DISAGREEMENT AND MINIMAL EPISTOCRACY

Adam F. Gibbons

DEPARTING FROM democratic ideals is heavily controversial among most contemporary western philosophers. Democracy, in its various forms, is widely seen as the all-things-considered best political arrangement. Still, recent work in political philosophy has challenged this orthodoxy. Central to these challenges lie worries about high levels of voter ignorance among modern democratic populations. Such ignorance, one might think, leads democracies to occasionally produce bad outcomes. If that is right, perhaps allocating comparatively more political power to voters who know more politically relevant facts will lead to better outcomes. Call political arrangements that make the possession of a certain amount of political knowledge a legal requirement for holding political power *epistocratic*.¹

In a recent paper, Julian Reiss articulates an important challenge to epistocracy.² At the core of any defense of epistocracy is the conviction that we can reliably identify a subset of voters who possess more politically relevant knowledge than others. But if we cannot identify such a subset of voters, the case for epistocracy falls at the first hurdle. We cannot allocate comparatively more political power to voters who know more politically relevant facts if we cannot even identify such voters.

Why think that we are unable to identify the appropriate subset of voters? Oversimplifying for the moment, it is natural to think that such voters should possess knowledge of various politically relevant social-scientific facts. Perhaps they should possess knowledge of basic economics, sociology, political science, and more. However, the social sciences are filled with controversy, and this controversy makes it exceedingly difficult to know which facts ought to be known by voters. Indeed, it makes it difficult to know the relevant facts *at all*. Reiss claims that since there are no uncontroversial social-scientific facts, we

1 There are several forms of epistocracy. See Brennan, *Against Democracy*, 204–30, for discussion.

2 Reiss, “Expertise, Agreement, and the Nature of Social Scientific Facts.”

cannot definitively say of some voters that they possess more politically relevant knowledge than others. The prevalence of disagreement about most issues in the social sciences precludes the possibility of identifying an uncontroversial body of knowledge against which to measure the putative competence of potential voters. Call this the *Argument from Political Disagreement*.³

In this paper, I respond to the Argument from Political Disagreement. After outlining the argument at length, I begin by arguing that there is a distinction between social-scientific knowledge and politically relevant knowledge. Not all politically relevant knowledge is social scientific, and there is much uncontroversial politically relevant knowledge. More specifically, there are *basic political facts*, and knowledge of these facts requires no acquaintance with the social sciences. I then establish the significance of knowledge of these basic political facts. While these basic political facts can seem more like unimportant political trivia than vital political information, knowledge of such facts is often central to voter decision-making. This body of knowledge paves the way for a *minimal epistocracy* wherein those who possess more of the relevant knowledge are allocated comparatively more political power.

1. THE ARGUMENT FROM POLITICAL DISAGREEMENT

Epistocrats think that we should allocate comparatively more political power to voters who possess more politically relevant knowledge. What constitutes politically relevant knowledge? Defenses of epistocracy emphasize the importance of *social-scientific knowledge*. Among other things, voters should have some knowledge of basic economics, sociology, political science, and the like.⁴ For example, one might think that the near consensus view among economists indicates that restrictive immigration policies harm the global economy, or that price controls are generally to be avoided.⁵ When voters are ignorant of such facts, they might vote for political candidates endorsing objectively harmful policies.⁶ Addition-

3 A challenge to epistocracy not considered in this paper questions the purported relationship between possessing more knowledge of politically relevant facts and competent political decision-making (Estlund, *Democratic Authority*, 206–22; Gaus, “Is the Public Competent?”). Even if we can reliably identify more knowledgeable voters, there is no guarantee that they will competently make political decisions. As outlined here, the Argument from Political Disagreement is different, concerning only the initial identification of more knowledgeable voters. Whether such voters are all-things-considered more competent than their less informed peers is an empirical question about which this paper is silent.

4 Brennan, *Against Democracy*, 212.

5 Caplan, *The Myth of the Rational Voter*, 85; Brennan, *Against Democracy*, 192.

6 For overviews of the relevant empirical literature on voter ignorance, see Oppenheimer and

ally, they might push self-interested politicians who are looking to pander to the electorate's preferences in the direction of such policies. If one wants to improve the overall quality of governance, those voters who know such facts ought to have comparatively more political power—or so epistocrats claim.

However, one might think that it is controversial whether the above really are facts. One might even think that for *any* putative social-scientific fact, there will be associated controversy.⁷ If such controversy abounds for any given putative social-scientific fact, the prospects for epistocrats seeking to delineate some uncontroversial body of political facts against which to measure the knowledge of voters seem dim. A crucial assumption underlying the case for epistocracy fails if there is no way to reliably identify voters who, in virtue of their greater levels of political knowledge, ought to be allocated comparatively more political power.

In support of the claim that the social sciences are mired in controversy, Reiss appeals to widespread *expert disagreement* among social scientists of various kinds.⁸ Consider the purported benefits of free trade. Such benefits are, at least in broad outline, agreed on by very many economists.⁹ But this agreement is not universal. Many economists, some lying outside of the mainstream, dissent. As Reiss puts it when discussing mainstream economic agreement on free trade and price controls: "The problem is that such agreement exists, if at all, at best among mainstream economists. When we look a little farther afield, for instance to heterodox economists, historians of economics, socio-economists and the like we are very unlikely to encounter agreement."¹⁰ Experts disagree about the benefits of free trade. It seems plausible, then, to grant that there is controversy on this issue. Of course, such controversy likely exists regarding virtually every other social-scientific issue of political importance. Social scientists often disagree with each other in a variety of domains, about the social problems that impact the most people, the underlying causes of various social problems, the appropriate policies to tackle such problems, the costliness of competing policies, and more.¹¹

Additional support comes from reflection on two important sources of expert disagreement. First, although controversy often arises because of disagreement about the relevant nonmoral facts, much disagreement in the social sci-

Edwards, *Democracy Despite Itself*, 9–38; Somin, *Democracy and Political Ignorance*, 17–37; Brennan, *Against Democracy*, 23–53; and Achen and Bartels, *Democracy for Realists*, 36–41.

7 Reiss, "Expertise, Agreement, and the Nature of Social Scientific Facts," 186–91.

8 Reiss, "Expertise, Agreement, and the Nature of Social Scientific Facts," 187.

9 Caplan, *The Myth of the Rational Voter*, 50–93.

10 Reiss, "Expertise, Agreement, and the Nature of Social Scientific Facts," 186.

11 Friedman, *Power without Knowledge*, 46–47.

ences has its roots in substantive *moral* disagreement. Claims about the benefits of free trade, about the economic harms wrought by restrictive immigration control, and about a wide array of other issues inevitably overlap with deeply controversial moral questions. Questions about the allocation of benefits and burdens from different economic policies are straightforwardly moral, as are questions about the appropriate way to aggregate different kinds of benefits or burdens. Questions about how we ought to make trade-offs between different values such as economic efficiency, equality, and freedom (among others) are paradigmatically moral questions, as are questions about the assignment of rights and responsibilities. Each of these questions is subject to vigorous dispute among professional social scientists and political theorists. To the extent that issues in the social sciences intersect with these questions, we should expect a certain degree of controversy in settling them.

Second, empirical generalizations in the social sciences are true (if they are true at all) only once certain contextual parameters are held fixed.¹² Empirical generalizations that are true in certain locations over certain timescales may not be true in other locations or over other timescales. With the introduction of such parameters, new loci of disagreement are thereby introduced, for the very choice of contextual parameter may be disputed. Naturally, there will also be straightforward nonmoral disagreement about the truth of certain empirical generalizations under transformations of the relevant parameters.¹³

Together, these sources of disagreement greatly limit the number of uncontroversial social-scientific facts against which a prospective epistocrat can measure the knowledge of voters. This, in turn, greatly decreases the feasibility of identifying some subset of voters who, because they possess the appropriate knowledge in greater proportions, ought to be allocated comparatively more political power. If this argument succeeds, a core assumption underlying the case for epistocracy is false.

To make the following discussion more precise, we can express the argument as follows:

1. There are no uncontroversial social-scientific facts.
2. If there are no uncontroversial social-scientific facts, then it is not possible to identify a subset of voters who possess more politically relevant knowledge than others.

12 Reiss, "Expertise, Agreement, and the Nature of Social Scientific Facts," 188–89.

13 Such disagreement will sometimes be nonmoral disagreement, but other times it will be about the sort of moral issues mentioned earlier. The worry about empirical generalizations, then, can be seen as a special form of our earlier two worries. Still, Reiss treats it separately, and I follow his lead here.

3. If it is not possible to identify a subset of voters who possess more politically relevant knowledge than others, then epistocracy is not feasible.
4. Therefore, epistocracy is not feasible.

2. BASIC POLITICAL FACTS

The centrality accorded to knowledge of social-scientific facts in the Argument from Political Disagreement is understandable. Indeed, epistocrats themselves stress the importance of social-scientific knowledge. Still, this emphasis is something of a red herring. Not all politically relevant facts are social-scientific facts, and epistocratic proposals recognize this distinction. For instance, while discussing potential qualification exams for voters, Jason Brennan writes that “to keep the test objective and nonideological, we could limit it to *basic facts* and fundamental, largely uncontested social-scientific claims.”¹⁴ Clearly, then, epistocrats do not think that all politically relevant facts are social scientific. Among such facts, they also include basic political facts. The upshot of this is clear: since there are some politically relevant facts that are not social scientific, the Argument from Political Disagreement fails.¹⁵

What are these basic political facts? Generally speaking, there are seemingly uncontroversial facts about the structure and function of important political institutions, the policy proposals of different candidates for office, existing policy and legislation (at the local, state, federal, and constitutional levels), the past actions of political figures, current budgetary spending, and more. Call facts like these *basic political facts*. The Argument from Political Disagreement fails if such facts are politically relevant. Specifically, premise 2 is false. One could simply grant the claim that there are no uncontroversial social-scientific facts while denying the further claim that such controversy precludes the identification of some subset of voters who possess more politically relevant knowledge than others. Epistocrats could endorse the allocation of more political power to voters who know more of the basic political facts since, as the empirical literature on voter ignorance shows, many voters are indeed ignorant of the basic political facts.¹⁶

14 Brennan, *Against Democracy*, 212, emphasis added.

15 Of course, one could also question the claim that the social sciences are as controversial as Reiss maintains. If, as Brennan claims, there are fundamental and largely uncontested social-scientific facts, the Argument from Political Disagreement fails in yet another way. However, I set this issue aside in this paper. The Argument from Political Disagreement fails even if the social sciences are controversial through and through.

16 Somin, *Democracy and Political Ignorance*, 17–37; Brennan, *Against Democracy*, 23–53.

There are several ways in which a proponent of the Argument from Political Disagreement might push back on this appeal to the basic political facts. First, one might reject the claim that such basic political facts are politically relevant. Some such facts look more like political trivia than vitally important political information. For instance, knowing the identity of the twenty-fourth president of the United States is unimportant, even though it is a basic historical political fact. One might even question the significance of facts about things like the identity of one's political representatives, the details of national budgets, and so on.¹⁷ Knowledge of these independent facts does not obviously play a role in the decision-making processes of voters. If that is right, then Reiss's argument might succeed after all.

But the claim that all such basic political facts are unimportant trivia is deeply implausible. When voters do not know who has enacted certain policies, they can assign praise (or blame) inaccurately; when they do not know how much of the federal budget is apportioned to different areas, they can be misled into believing that spending should be cut (or increased) in these areas; when they do not know the policy proposals of candidates, they might vote in ways they would not otherwise; and so on.¹⁸ This last point is important: when voters are ignorant of the basic political facts, they can vote in ways that they would not have wanted to if they had known otherwise.¹⁹ Voters often go wrong by their own lights when they have false beliefs about the basic political facts. Whenever they do go wrong, they can end up with policies (and leaders) they do not want. The basic political facts are not only politically relevant but also often central to voter decision-making.

Rather than denying the political relevance of the basic political facts, a proponent of the Argument from Political Disagreement might instead reject the claim that they are uncontroversial. After all, many voters do not know them, and many voters disagree about them. Instead of the earlier Argument from Political Disagreement, we could have an amended version focusing on the basic political facts:

1. There are no uncontroversial basic political facts.
2. If there are no uncontroversial basic political facts, then it is not possible to identify a subset of voters who possess more politically relevant knowledge than others.

17 Fuller, *In Defense of Democracy*, 28–29.

18 Hochschild and Einstein, *Do Facts Matter?* 23–27.

19 Bartels, "Uninformed Votes."

3. If it is not possible to identify a subset of voters who possess more politically relevant knowledge than others, then epistocracy is not feasible.
4. Therefore, epistocracy is not feasible.

But the basic political facts are controversial only in a highly attenuated sense. First, it is important to note that the basic political facts bear none of the indicators of controversy outlined in the previous section regarding controversy in the social sciences. There is no expert disagreement about the basic political facts, there are no underlying substantive moral disputes lurking beneath the basic political facts, the truth of basic political facts is not hostage to contextual parameters about which there may be disagreement, and so on. Additionally, the basic political facts, unlike controversial putative facts in the social sciences, are easily confirmed. Given the hallmarks of controversy enumerated earlier in the paper, things like the purported benefits of free trade are understandably hard to confirm or disconfirm. But it is not hard to confirm, say, who your senators are, what their policies are (at least in broad outline), and the like.

A proponent of the Argument from Political Disagreement might insist that the presence of disagreement is by itself necessary and sufficient for there to be controversy. Since there is disagreement about some of the basic political facts, there is therefore controversy about them. But if *any* disagreement whatsoever constitutes controversy, then virtually nothing is uncontroversial. On this account, it is controversial whether the earth is flat, whether Dublin is the capital of Ireland, and more. I simply assume that such verdicts are misguided, and that a conception of controversy this expansive cannot bear the weight placed on it in the Argument from Political Disagreement. Presumably, proponents of this argument have something more demanding in mind. But more demanding conceptions of controversy, while much more plausible, will not count the basic political facts as controversial. For instance, if controversy requires expert disagreement, then the basic political facts are not controversial. If it requires epistemic peer disagreement, then the basic political facts are not controversial since the relevant disputes do not always involve epistemic peers.²⁰ It is plausible, then, to conclude that the basic political facts are not controversial in the right way for the Argument from Political Disagreement to succeed. Since the basic political facts are both politically relevant and uncontroversial, the Argument from Political Disagreement fails.

20 For helpful discussion of competing accounts of what it is to be an epistemic peer, see Gelfert, "Who Is an Epistemic Peer?"

3. MINIMAL EPISTOCRACY

If we set aside all putative social-scientific facts on the grounds that they are controversial, what sort of epistocracy are we left with? More demanding epistocratic proposals seeking to empower groups of political experts in virtue of their specialized social-scientific knowledge are ruled out. But a *minimal* form of epistocracy focused on those voters who possess more knowledge of the basic political facts is still viable. A minimal epistocracy might implement voter qualification exams pivoting around the relevant facts, with failure to pass the exams resulting in disenfranchisement. Alternatively, it could amplify the political power of more knowledgeable voters by allocating more votes to them in proportion to their knowledge. A different method still could be to simulate voter political preferences, relative to their demographic group, under simulated conditions of full knowledge of some set of the basic political facts.²¹

Nevertheless, one might still have reservations about minimal epistocracy. For instance, one might claim that the basic political facts, while neither politically irrelevant nor controversial in any meaningful sense, are such that allocating more political power to voters who know them will not gain us much. Epistocratic reforms, after all, are supposed to mitigate the harmful effects of voter ignorance. But perhaps epistocratic reforms pivoting around a minimum core of basic political facts will not mitigate such effects enough. Perhaps they would even *worsen* outcomes relative to the status quo.²²

However, this criticism is entirely distinct from the original argument with which we began. The Argument from Political Disagreement is *not* an argument to the effect that the overall costs of transitioning to epistocracy (and away from democracy) outweigh the benefits. Such concerns about the overall expected costs and benefits are perfectly general, applying to prospective epistocratic arrangements even if—*pace* the Argument from Political Disagreement—it were trivially easy to identify some subset of voters who possess much more politically relevant knowledge than others. Instead, the Argument from Political Disagreement is an attempt to show that epistocrats cannot identify some subset of

21 Those familiar with the literature on epistocracy will recognize these options as, respectively, restricted suffrage, plural voting, and what Brennan calls “rule by simulated oracle” (*Against Democracy*, 204–30).

22 For instance, epistemic democrats claim that collections of individually ill-informed agents can, under the right conditions, epistemically outperform numerically smaller collections of more knowledgeable agents. See Landemore, *Democratic Reason*; Schwartzberg, “Epistemic Democracy and Its Challenges”; and Goodin and Spiekermann, *An Epistemic Theory of Democracy*. Naturally, it is controversial whether actual democracies satisfy the relevant conditions. For some critical discussion, see Brennan, *Against Democracy*, 172–203.

voters who possess more knowledge of politically relevant facts. The existence of uncontroversial and politically relevant basic political facts shows that this is mistaken.

I conclude, then, that some voters do possess more politically relevant knowledge than others and that the Argument from Political Disagreement fails. At the very least, a minimal form of epistocracy is still feasible. If we are to reject epistocracy, we must do so on other grounds.²³

Rutgers University—New Brunswick
adam.gibbons@rutgers.edu

REFERENCES

- Achen, Christopher H., and Larry M. Bartels, *Democracy for Realists: Why Elections Do Not Produce Responsive Government*. Princeton, NJ: Princeton University Press, 2016.
- Bartels, Larry M. "Uninformed Votes: Information Effects in Presidential Elections." *American Journal of Political Science* 40, no. 1 (February 1996): 194–230.
- Brennan, Jason. *Against Democracy*. Princeton, NJ: Princeton University Press, 2016.
- Caplan, Bryan. *The Myth of the Rational Voter: Why Democracies Choose Bad Policies*. Princeton, NJ: Princeton University Press, 2007.
- Estlund, David M. *Democratic Authority: A Philosophical Framework*. Princeton, NJ: Princeton University Press, 2008.
- Friedman, Jeffrey. *Power without Knowledge: A Critique of Technocracy*. New York: Oxford University Press, 2019.
- Fuller, Roslyn. *In Defense of Democracy*. Medford, MA: Polity Press, 2019.
- Gaus, Gerald. "Is the Public Competent? Compared to Whom? About What?" *Critical Review* 20, no. 3 (2008): 291–311.
- Gelfert, Axel. "Who Is an Epistemic Peer?" *Logos and Episteme* 2, no. 4 (2011): 507–14.
- Goodin, Robert E., and Kai Spiekermann. *An Epistemic Theory of Democracy*. Oxford: Oxford University Press, 2018.
- Hochschild, Jennifer L., and Katherine Levine Einstein. *Do Facts Matter? Information and Misinformation in American Politics*. Norman: University of Oklahoma Press, 2015.

23 Thanks to Alex Guerrero, Oliver Traldi, and three anonymous reviewers for their feedback on this paper.

- Landemore, Hélène. *Democratic Reason: Politics, Collective Intelligence, and the Rule of the Many*. Princeton, NJ: Princeton University Press, 2013.
- Oppenheimer, Danny, and Mike Edwards. *Democracy Despite Itself: Why a System That Shouldn't Work at All Works So Well*. Cambridge, MA: MIT Press, 2012.
- Reiss, Julian. "Expertise, Agreement, and the Nature of Social Scientific Facts or: Against Epistocracy." *Social Epistemology* 33, no. 2 (2019): 183–92.
- Schwartzberg, Melissa. "Epistemic Democracy and Its Challenges." *Annual Review of Political Science* 18 (May 2015): 187–203.
- Somin, Ilya. *Democracy and Political Ignorance: Why Smaller Government Is Better*. Stanford, CA: Stanford University Press, 2013.