AMERICAN JOURNAL
of LAW *and* EQUALITY

PATTERNED INEQUALITY, COMPOUNDING INJUSTICE, AND ALGORITHMIC PREDICTION

Benjamin Eidelson*

INTRODUCTION

Large datasets and novel statistical methods have given rise to a new wave of predictive algorithms that increasingly guide all manner of public and private decisions.[1] Many (though not all) of these new-age predictive tools are generated by yet other algorithms —that is, by automated processes that scour a dataset for patterns and thereby construct a function for predicting, as best as the data permits, what will transpire in future cases.[2] These predictive tools promise both vital information and refreshing objectivity: They avoid many of the recurring mistakes made by human predictors, and, of course, they hold

Author: *Benjamin Eidelson is an Assistant Professor of Law at Harvard Law School. His prior work on the moral foundations of discrimination law includes *Discrimination and Disrespect* (2015) and *Respect, Individualism, and Colorblindness*, 129 Yale L.J. 1600 (2020). This article is based on a November 2020 keynote presentation for the Workshop on Algorithmic Fairness at the University of Copenhagen. Thanks to Kasper Lippert-Rasmussen and Sune Holm for that invitation; to Hannah Hilligoss, Adam Hosein, Martha Minow, Gerald Neuman, Cass Sunstein, and the workshop participants for very helpful comments on earlier drafts; to Jared Lin for valuable research assistance; and especially to Deborah Hellman for her characteristically generous engagement with this project, both through her reply in this volume and through many fruitful conversations about the issues addressed here.

[1] Nearly all predictions are made with "algorithms" in one sense of the word, but I will reserve the term for formal, computational processes. *Cf.* Solon Barocas & Andrew D. Selbst, *Big Data's Disparate Impact*, 104 Calif. L. Rev. 671 (2016).

[2] *See*, *e.g.*, Jon Kleinberg et al., *Discrimination in the Age of Algorithms*, 10 J. Legal Analysis 113, 115 (2018) (describing the relationship between training and screening algorithms).

https://doi.org/10.1162/ajle_a_00017

no positive or negative attitudes toward any of the people whose fates they are asked to foretell.[3]

Yet a large and growing body of evidence shows that these predictive algorithms tend to predict bad outcomes—recidivism, falling behind on a loan, and more—far more often for members of socially disadvantaged groups than for others.[4] And although the algorithms' predictions may be equally accurate for members of different groups, the ways in which they err (when they do) differ: The algorithms tend more strongly toward mistaken pessimism when it comes to members of disadvantaged groups but more strongly toward mistaken optimism when it comes to members of advantaged groups.[5] These disparities have fueled both technical and legal literatures about how different modifications to the predictive algorithms (or to the upstream algorithms that produce them) might achieve what many now term "algorithmic fairness."[6]

This article takes up a more basic normative question that looms in the background of those debates: Why are the disparities that I have just described morally troubling at all? After all, if we know that access to resources and opportunities is stratified by race and gender (as we do), then we should *expect* to see the effects of that inequality manifest in different people's likelihoods of realizing more and less favorable outcomes.[7] Put differently, if an algorithm were *not* more pessimistic about the prospects of members of disadvantaged groups, we would have to conclude either that the algorithm was distorted or that

[3] *See*, *e.g.*, Cass R. Sunstein, *Algorithms, Correcting Biases*, 86 Soc. Rsch. 499, 500–04 (2019) (recounting evidence that algorithms outperform judges in predicting flight risk, including because they do not suffer from "current offense bias"); Kleinberg et al., *supra* note 2, at 154 (explaining how the use of algorithms can protect against "explicit and implicit biases" regarding particular groups). As Kleinberg et al. argue, algorithmic decisionmaking is also potentially transparent to investigation in ways that human decisionmaking is not. *See id.* at 116–18. *But cf.* Andrew D. Selbst & Solon Barocas, *The Intuitive Appeal of Explainable Machines*, 87 Fordham L. Rev. 1085, 1089 (2018) (discussing a variety of "difficult questions about how to observe, access, audit, or understand . . . algorithms").

[4] *See*, *e.g.*, Julia Angwin et al., *Machine Bias*, ProPublica (May 23, 2016), https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing; Talia B. Gillis, *The Input Fallacy*, Minn. L. Rev. (forthcoming 2022), https://doi.org/10.2139/ssrn.3571266.

[5] *See*, *e.g.*, Deborah Hellman, *Measuring Algorithmic Fairness*, 106 Va. L. Rev. 811, 815–16 (2019); Jon Kleinberg et al., *Inherent Trade-Offs in the Fair Determination of Risk Scores*, 67 LIPIcs 43:1, 43:5–8 (2017), https://drops.dagstuhl.de/opus/volltexte/2017/8156/pdf/LIPIcs-ITCS-2017-43.pdf.

[6] *See*, *e.g.*, Anupam Chander, *The Racist Algorithm?*, 115 Mich. L. Rev. 1023, 1039–45 (2017); Cynthia Dwork et al., *Fairness Through Awareness*, Cornell Univ., arXiv:1104.3913v2 [cs.CC] (2011); Moritz Hardt et al., *Equality of Opportunity in Supervised Learning*, Cornell Univ., arXiv:1610.02413v1 (2016); Hellman, *supra* note 5; Kleinberg, *supra* note 5, at 43:5; Crystal Yang & Will Dobbie, *Equal Protection Under Algorithms: A New Statistical and Legal Framework*, 119 Mich. L. Rev. 291, 346–50 (2020).

[7] As Glenn Loury put the same point: "Contemporary American society has inherited a racial hierarchy. . . . It can be no surprise in such a society that the web of interconnections among persons that facilitate access to opportunity and shape the outlooks of individuals would be raced, which is to say, that processes of human development would be systematically conditioned by race. Thus, racially disparate outcomes at the end of the twentieth century can be no surprise, either." Glenn C. Loury, The Anatomy of Racial Inequality 106 (2002).

the putative disadvantage was not really disadvantageous. And, importantly, much the same goes for intergroup differences in the ratios of different kinds of errors: Any algorithm that aims to get the right answer in as many cases as possible will produce just such disparities. For, just as members of a disadvantaged group will necessarily tend to have more bad outcomes, so too will they tend to have more of the characteristics that are associated with a bad outcome even when one does not actually materialize.[8]

The point is that much of the putative unfairness that results from algorithmic predictions—for which various interventions represent proposed correctives—probably cannot just be chalked up to a biased view of reality.[9] I hasten to add that some significant portion of the observed disparities in predictions surely *is* due to forces that bias the data against members of disadvantaged groups, making them appear riskier than they actually are.[10] But it should be uncontroversial that another significant portion of the disparities is likely due to forces that bias the *relevant facts* against members of those same groups in a manner that the data may then faithfully describe. Again, in an unjust society, it could hardly be otherwise. And insofar as predictive algorithms yield disparate predictions for the latter reason—that is, because they are making the most reliable predictions that the relevant facts permit—how should we understand the moral concerns that the stark disparities nonetheless elicit? Does the fact of those disparities tell in favor of discounting the algorithms' predictions and, in effect, refusing to treat alike cases that present the same likelihood of a given outcome?[11]

[8] *See*, *e.g.*, Hellman, *supra* note 5, at 839–40 ("[I]f the information we have indicates that, collectively, one group is more likely to recidivate than the other, more people in that group will be scored as high risk (both correctly and incorrectly). . . . There are more false positives for blacks in the COMPAS data because the data shows that blacks commit more crime and so the algorithm will predict more black crime and will do so imperfectly.").

[9] Put differently, much of the putative unfairness cannot be understood as violating "the principle that any two individuals who are similar with respect to a particular task should be classified similarly." Dwork, *supra* note 6, at 1.

[10] For example, if a Black person is more likely to be arrested than a white person, conditional on committing the same offense, and if a model uses data about past arrests to predict who will commit future offenses, then the attributes that are disproportionately held by Black people will appear to be predictive of a future offense even if they are not. *See*, *e.g.*, Sandra G. Mayson, *Bias In, Bias Out*, 128 Yale L.J. 2218, 2251–56 (2019). A related, but distinct, issue concerns the choice of what outcome to treat as decisionally relevant in the first place. For instance, a hospital that predicts patients' expected future health costs, and triages its investments in their care on that basis, will thereby prioritize the disproportionately white patients who are likely to use more healthcare over others who are equally sick. *See* Ziad Obermeyer et al., *Dissecting Racial Bias in an Algorithm Used to Manage the Health of Populations*, 366 Sci. 447 (2019).

[11] This discounting of the algorithms' predictions could take various forms. *See* sources cited *supra* note 6. But, in general, a rational decisionmaker who gives some weight to the demographic makeup of a class of people (such as those detained, hired, or the like) "should not change the choice of estimator" to accommodate that preference, but rather should simply "change how the estimated prediction function is used (such as setting a different [risk]

In the first part of this article, I suggest that the answer to this question is "yes" and sketch an explanation of why. The thrust and spirit of the argument will be familiar to students of influential "anti-subordination" accounts of anti-discrimination law, although I mean to draw out one thread of such accounts and disentangle it from others.[12] Simply stated, the explanation that I propose has two steps. First, the very fact that inequalities in status, resources, and opportunities are discernibly patterned in terms of certain socially salient identities brings about a great deal of human misery and social injustice. Second, practices that accurately screen for various forms of "merit" predictably reproduce and aggravate patterns of this kind.[13]

Those two premises together ground a powerful moral objection to the unbridled use of the relevant selection practices. Importantly, however, it is an objection that has little to do with the accuracy of any judgments about people. It would have force even if an omniscient decisionmaker could somehow select the *actual* top candidates for some position—those whose future job performance (or the like) would actually be best—as long as the selected group would be disproportionately white and male. And so, according to this line of thought, there is at least one serious problem with the use of algorithmic predictions to make allocative decisions that is not really about algorithms or even about predictions. It is just a general problem with allocating opportunities in ways that reproduce patterned inequality.

After sketching an argument along these lines, I introduce in Part II another moral objection leveled against the same selection practices—namely, that individuals are sometimes wronged by otherwise-fair decisions because of the causal history that underlies those individuals' relevant characteristics. Over a series of important papers, Deborah Hellman has developed an objection of that kind to explain both some traditional precepts

threshold for different groups)." Jon Kleinberg et al., *Algorithmic Fairness*, 108 AEA Papers & Procs. 22, 22–23 (2018).

[12] *See*, *e.g.*, Samuel R. Bagenstos, *Rational Discrimination, Accommodation, and the Politics of (Disability) Civil Rights*, 89 Va. L. Rev. 825, 847 (2003) (arguing that anti-discrimination law is best seen as "attacking practices that entrench the systemic subordination of particular groups"); Owen M. Fiss, *Groups and the Equal Protection Clause*, 5 Phil. & Pub. Affairs 107, 157 (1976) (suggesting that equal protection analysis should focus on whether "the state law or practice aggravates (or perpetuates?) the subordinate position of a specially disadvantaged group"); *see also* Richard Delgado & Jean Stefancic, Critical Race Theory: An Introduction 27–28 (2d ed. 2012) (arguing that if new laws "would not relieve the distress of the poorest group—or, worse, if they compound it—we should reject them").

[13] Following T.M. Scanlon, I will use "merit" in an "institution-dependent" sense; so understood, it refers not to any form of intrinsic value, but to the attributes that serve the justification for having some inequality in the first place. *See* T.M. Scanlon, Why Does Inequality Matter? 40–46 (2018); *see also* Elizabeth Anderson, The Imperative of Integration 163 (2010) (positing that "'merit' is any characteristic of individuals whereby they advance an institution's proper mission through their performance in an institutional role"). Although I appreciate that some will find all talk of merit problematic (at least without skeptical quotation marks), I think using the term helpfully underscores that the argument advanced here does not depend on any such skepticism.

of anti-discrimination law and the moral concerns raised by some decisions predicated on algorithmic predictions.[14] The core idea is that one wrongfully *compounds* a prior wrong that a person has suffered by taking that wrong, or its effects, as a reason for doing something to the victim that makes her still worse off. According to Hellman, legal and moral norms about indirect discrimination,[15] and related concerns about algorithmic prediction, take aim at this kind of wrongful complicity in others' earlier wrongdoing.

Laying these two accounts of anti-discrimination norms side-by-side facilitates the intramural debate between them that I frame in Part III. The debate is intramural because both concerns look beyond "meritocratic" worries about misjudging people and point instead to the moral significance of facts about how merit is distributed in the first place. But the debate is a debate because, at bottom, the normative visions are quite different: The objection from patterned inequality rests on the prospective effects, for the overall justice and goodness of the society, of decisions about how present opportunities will be allocated, whereas Hellman's "compounding injustice" objection identifies a personal wrong done to a disfavored individual and turns on the etiology of that individual's present disadvantage. I argue that the objection from patterned inequality is more plausible as a matter of moral theory and better fits and justifies the ambitious conception of anti-discrimination norms that I take to animate both views. And I suggest that this conclusion has important consequences for how we think about wrongful discrimination, in the context of algorithmic decisionmaking and beyond.

I. PATTERNED INEQUALITY, DISCRIMINATION, AND ALGORITHMS

A. The Problem of Patterned Inequality

All inequalities in resources, opportunities, and status raise questions of justice. But in addition to the problems posed by inequality generally, there are special problems posed when a society's inequalities are patterned in terms of race, gender, and other socially salient axes of identity. That is not just because a stark pattern can be evidence of unfairness

[14] See Deborah Hellman, *Indirect Discrimination and the Duty to Avoid Compounding Injustice*, *in* FOUNDATIONS OF INDIRECT DISCRIMINATION LAW 105–121 (H. Collins & T. Khaitan eds., 2018) [hereinafter "*Indirect Discrimination*"]; Deborah Hellman, *Sex, Causation, and Algorithms: How Equal Protection Prohibits Compounding Prior Injustice*, 98 WASH. U. L. REV. 481 (2020) [hereinafter "*Sex, Causation*"]; Hellman, *supra* note 5, at 841–42; Deborah Hellman, *Big Data and Compounding Injustice*, 18 J. MORAL PHIL. (forthcoming 2021).

[15] Indirect discrimination norms come in weaker and stronger forms. In their weaker form, they require covered actors to show that practices that have disproportionate effects on disadvantaged groups actually best serve the actor's ordinary goals (such as economic efficiency); apart from the burden of establishing such a justification, no compromise of those goals is required for the sake of avoiding the disproportionate effects. In their stronger form, indirect discrimination norms require that any asserted benefit in efficiency (or the like) be large enough to justify the presumed harm of causing the disproportionate effects. *See*, *e.g.*, Hugh Collins & Tarunabh Khaitan, *Indirect Discrimination Law: Controversies and Critical Questions*, *in* FOUNDATIONS OF INDIRECT DISCRIMINATION LAW, *supra* note 14, at 1, 16–17.

in the processes that yielded the pattern (although it certainly can be that). It is also because the very existence of patterned inequalities gives rise to grave social ills and undermines human flourishing in ways that other inequalities do not.

Race in the United States is an obvious paradigm case. The fact that Black Americans enjoy dramatically worse opportunities than white Americans, on average, has severely negative consequences—consequences that would not follow from an equally unequal distribution that lacked the correlation with race. That is so for several reasons well explained by others; I only summarize a few of the relevant dynamics here.

First, what I am calling patterned inequality (and only that kind of inequality) gives rise to the problem of *self-confirming stereotypes*.[16] Once a visible trait, meaningless in itself, is correlated with less visible attributes of interest, decisionmakers will tend rationally to use the visible trait as a proxy for the traits of interest.[17] And once they do that, individuals who hold the relevant visible trait will tend to adapt their own behavior in expectation of that sorting process, often in ways that sustain or amplify the original correlation. As Samuel Bagenstos explains, "[t]he result may be a vicious cycle of exclusion, in which members of subordinated groups rationally respond to exclusion by failing to develop their human capital, and employers, rationally believing that members of those groups are less likely to have developed their human capital, discriminate even more."[18] Although Bagenstos focuses on employment in particular, the same dynamic plays out across every sphere of social life in which "[i]nformation-hungry human agents" draw inferences about each other and modify their own aspirations, attitudes, and behavior in response to the inferences they have learned to expect that others will draw about them.[19]

Second, this dynamic is compounded when those on the losing end of inequalities in resources or status are *socially isolated*. Because residential patterns and social and economic networks in the United States are themselves overwhelmingly patterned by race, the fates of Black Americans are often linked together in ways that magnify the consequences of bad outcomes suffered by each individual.[20] The concentrated pockets of disadvantage that result are far worse, in terms of their impact on individual well-being

[16] *See* Loury, *supra* note 7, at 26-35; Anderson, *supra* note 13, at 55–57; Bagenstos, *supra* note 12, at 842–43; Kenneth J. Arrow, *The Theory of Discrimination*, in Discrimination and Labor Markets 3, 26–27 (Orley Ashenfelter & Albert Rees eds., 1973).

[17] *Cf.* Anderson, *supra* note 13, at 161 ("[R]ace is typically less costly to detect than the underlying relevant criteria. This can make statistical discrimination on the basis of [race] instrumentally rational, even if the correlations are not due to character races but to racialization.").

[18] Bagenstos, *supra* note 12, at 843.

[19] Loury, *supra* note 7, at 17; *see id.* at 23–33 (offering illustrations).

[20] *See id.* at 103 ("Because access to developmental resources is mediated through race-segregated social networks, an individual's opportunities to acquire skills depend on present and past skill attainments by others in the same racial group."); Bagenstos, *supra* note 12, at 834; Kasper Lippert-Rasmussen, *Private Discrimination: A Prioritarian, Desert-Accommodating Account*, 43 San Diego L. Rev. 817, 835–36 (2006).

and opportunity, than the sum of their individual parts.[21] At the extreme, as Christopher Lewis explains, "[w]here opportunities for regular, legal work are sparse, people have stronger incentives to steal, rob, con, deceive, sell harmful and addictive drugs, [and] exploit others."[22] More generally, a social group whose members are disproportionately on the losing side of inequalities will disproportionately lack the social and cultural capital needed to access future opportunities, or to provide those opportunities for their children, as well.[23] Indeed, the very structures of the institutions affording those opportunities will be warped so as to accommodate the needs of their usual occupants and not the needs of others.[24] And the worse things go for members of the group—the more starkly various measures of well-being or achievement are racially patterned—the more strongly group membership is apt to be associated with negative qualities by decisionmakers and treated as a basis for further discrimination, fueling the same vicious cycle traced above.

Third, when the trait that correlates with a society's inequalities has deep cultural significance—and particularly when it carries a *stigma* in Erving Goffman's sense—that cultural significance will interact with the processes just described in ways that make them still more destructive.[25] For one thing, the stigma will color observers' interpretations of otherwise-ambiguous data in ways that fuel both epistemically rational and epistemically irrational stereotyping.[26] And for another, it will shape the larger society's reaction to the plight of those who find themselves on the losing side of inequality. Because Black Americans are stigmatized, for example, the disadvantages that many of them suffer are not met with the empathy and attendant social mobilization that would be triggered if the racial pattern were different.[27] Instead, the outcomes are tacitly accepted as ordinary and unremarkable or, indeed, treated as a basis for censure and punishment. The different societal responses to the "crack epidemic" in Black communities and the "opioid epidemic" in white ones offer a good example.[28]

[21] For an extended argument along these lines, see ANDERSON, *supra* note 13.

[22] Christopher Lewis, *Inequality, Incentives, Criminality, and Blame*, 22 LEGAL THEORY 153, 173–74 (2016).

[23] ANDERSON, *supra* note 13, at 33–38; *see also* LOURY, *supra* note 7, at 99–104.

[24] *See* SOPHIA MOREAU, FACES OF INEQUALITY 56–63 (2020).

[25] *See* ERVING GOFFMAN, STIGMA: NOTES ON THE MANAGEMENT OF SPOILED IDENTITY (1963). This point is emphasized by LOURY, *supra* note 7, at 55–108, and by Bagenstos, *supra* note 12, at 841–44, among others.

[26] *See*, *e.g.*, RANDALL KENNEDY, RACE, CRIME, AND THE LAW 154 (1997) (explaining how even well-meaning police officers "will unintentionally *exaggerate* the criminality or potential for criminality of African-Americans" in light of "age-old, derogatory images of the Negro as criminal"); ANDERSON, *supra* note 13, at 44–50, 55–57.

[27] *See* LOURY, *supra* note 7, at 104–07; *cf.* David A. Strauss, *Discriminatory Intent and the Taming of* Brown, 56 U. CHI. L. REV. 935, 971–75 (1989).

[28] *Cf.* LOURY, *supra* note 7, at 105–06.

All of this could naturally be taken as a thumbnail sketch of *group subordination* or *oppression*, and I have no quibble with those who would view it as such. I put the point in terms of "patterned inequality" instead simply because this formulation allows us to pick out a present state of affairs without any implicit reference to how it came about or to its moral significance.[29] "Subordination" and "oppression," after all, are nominalizations of verbs: They naturally refer to wrongful actions, or perhaps to states of affairs that are distinguished by being the effects of such actions. A pattern, by contrast, is not an action at all and need not have any causal connection to one. Moreover, a focus on the pattern nicely highlights that my argument here is individualistic in both its methodological and its normative commitments; a pattern is made up of individuals and has no existence apart from them.[30] In any event, my basic point is that when members of a socially salient group are disproportionately among the worse off in a society, the very existence of that pattern in the society's inequality poses grave problems for a range of widely endorsed values. It greatly undermines future equality of opportunity (measured at the level of individuals, not groups); it makes it less likely that people will relate to one another as equals; and, all issues of equality aside, it just means more overall misery and less overall flourishing.[31]

Finally, although I have focused here on race (and on Black–white inequality in particular), the story for some other socially salient identities would be only somewhat different. In the context of gender, for example, the lack of heritability and the comparative lack of social isolation undoubtedly change the relevant dynamics. Still, the pattern in which women tend to hold lower-status jobs than men is naturally analyzed in a fundamentally similar way. As with racial stratification, that pattern both causes and is caused by a blend of rational and irrational stereotyping, predictable and perhaps individually adaptive responses to stereotyping, and a background of social meanings and attitudes that shape individual and collective responses to the continuing inequality that results.[32] In what follows, I will sometimes bracket these differences and refer to race and gender in the same breath, or to race alone, for purposes of concision.

[29] The language of "patterned inequality" is also employed by Iris Marion Young in *Equality of Whom? Social Groups and Judgments of Injustice*, 9 J. Pol. Phil. 1 (2001). Young argues that warranted claims about "unjust structural inequalities" require "finding patterned inequality and being able to tell a plausible story about how the position in structures accounts for that inequality." *Id.* at 17.

[30] *Cf.* Tarunabh Khaitan, A Theory of Discrimination Law 129 (2015) ("[O]ur concern with relative group disadvantage is based on a concern for individuals. Groups are salient to discrimination law because group membership has a significant impact on the life-chances of a person.").

[31] Regarding the kind of substantive opportunity required to meet a complaint of inequality, see Scanlon, *supra* note 13, at 58–65; regarding the issue of status inequality, see *id.* at 26–39; and regarding the threat to well-being posed by the "abiding, pervasive, and substantial relative disadvantage faced by members of [certain] groups," see Khaitan, *supra* note 30, at 91–139.

[32] *See*, *e.g.*, Young, *supra* note 29, at 10–11.

B. Anti-Discrimination Norms and Patterned Inequality

Anti-discrimination norms serve multiple functions, but among the most important is that they undermine patterned inequality. When the law forbids employers to take account of race and gender in hiring, for example, it makes it less likely that the employers' hiring will reproduce the existing, racialized and gendered pattern of advantage.[33] Concretely, it makes it more likely that candidates other than white men will apply for desirable positions, and more likely that employers will hire those who do, than would otherwise have been the case.[34] And to whatever extent the law's intervention does that, it serves to erode the intra-generational and intergenerational patterns that, through the various mechanisms sketched above, limit many people's flourishing and undermine equality of status and opportunity.[35]

In many cases, to be sure, the same prohibitions also make the employer's decisions better track the candidates' actual qualifications. The rules do that insofar as they prohibit employers from bringing certain inaccurate beliefs or biased attitudes to bear.[36] But it is a mistake to think that anti-discrimination norms reduce patterned inequality solely *because* they reduce bias in this way—or, similarly, that the reduction in patterned inequality is just a happy by-product of the reduction in bias, rather than a central concern in its own right. In fact, a great many discriminatory decisions could not be criticized as expressions of the decisionmaker's irrational bias—both because, in a society such as ours, race and gender will often carry information about individual qualifications, and because integration along these dimensions sometimes comes with its own costs for employers.[37] Understanding anti-discrimination norms as an intervention against patterned inequality helps to explain why it is appropriate to prohibit discrimination in those cases (cases of "rational" discrimination) just as in others.[38] The harms of patterned inequality—and

[33] I will not take up the possible extension of the argument here to other socially salient axes of identity (or corresponding grounds of prohibited discrimination). To be clear, though, I do not presume that the justification for every protected ground must be the same or that each must involve a concern about patterned inequality. Likewise, the prohibited grounds of direct and indirect discrimination might justifiably differ. *See* Benjamin Eidelson, Discrimination and Disrespect 40 (2015). For an instructive and more general discussion of discrimination law's "protectorate," see Khaitan, *supra* note 30, at 49–58.

[34] In Owen Fiss's early and frank formulation, "the antidiscrimination prohibition is a strategy for conferring benefits on a racial class—blacks." Owen M. Fiss, *A Theory of Fair Employment Laws*, 38 U. Chi. L. Rev. 235, 313 (1971).

[35] On the intragenerational aspect in particular, see Joseph Fishkin, Bottlenecks: A Theory of Equal Opportunity 70 (2014).

[36] *See, e.g.*, Frederick Schauer, Profiles, Probabilities, and Stereotypes 151 (2003) (suggesting that "the moral, legal, and constitutional prohibitions on sex discrimination . . . are best understood as the mandated *under*use of gender-based generalizations to compensate for the likelihood of their exaggeration and the likelihood of their overuse").

[37] *See* Anderson, *supra* note 13, at 161; Bagenstos, *supra* note 12, at 849–51.

[38] *Cf.* Patrick Shin, *Is There a Unitary Concept of Discrimination?*, *in* Philosophical Foundations of Discrimination Law 163, 175–76 (Deborah Hellman & Sophia Moreau eds., 2013).

hence the compelling reasons for seeking to reduce it—do not depend on whether any particular person has or has not suffered any unfair bias in how she is rewarded for her decision-relevant characteristics (i.e., loosely speaking, for her "merit").[39]

Indeed, from this point of view, there is a basic unity to the traditional prohibition on direct discrimination, the somewhat more controversial prohibition on indirect (or "disparate impact") discrimination, and the voluntary use of affirmative action policies. All three can be understood partly as efforts to produce workforces, student bodies, and the like that include more members of certain groups than selection processes unconstrained by any concern for the effect on patterned inequality would yield.[40] By ensuring that more members of particular groups hold these positions and can access their various benefits, these interventions have the potential to disrupt the pattern-sensitive cultural and economic processes that otherwise inflict grave harm.[41] In the race context, for example, they can facilitate interracial social networks, reduce the racial wealth gap, and challenge through experience the stereotypic assumptions held by members of both dominant and subordinated racial groups.[42] Simply put, if the pattern itself is a problem, then a change in the pattern is itself a solution.[43]

To be clear, I claim here only that this normative vision offers a persuasive account of one of the important values served by anti-discrimination norms—not that it captures the vision reflected in the modern case law of the U.S. Supreme Court. Some Justices (and perhaps a majority) would balk at the notion of a concerted state effort to alter the racial pattern in social inequalities, even if that effort were undertaken in service of values defined without reference to race.[44] They view anti-discrimination norms as efforts to align

[39] Regarding my usage of "merit," see *supra* note 13.

[40] I offered an account of indirect discrimination law along these general lines in Eidelson, *supra* note 33, at ch. 2. As noted above, these laws generally impose a heightened standard of justification for practices that disproportionately exclude members of particular groups. *See supra* note 15.

[41] *Cf.* Bagenstos, *supra* note 12, at 843–44 ("Antidiscrimination law responds to the[] harms of social inequality by promoting the integration of workplaces and other important areas of civic life. Such integration helps to remove the stigmatic injury that results from exclusion in a number of ways. . . . ").

[42] *See*, *e.g.*, Anderson, *supra* note 13, at 149–50; Randall Kennedy, For Discrimination: Race, Affirmative Action, and the Law 85–86, 106–08, 133–34 (2013). I do not claim that this is *all* that these prohibitions and policies serve to do, or that the mechanisms are the same for the three different interventions. And I do not claim that integration inevitably or perfectly accomplishes these objectives. For criticism along those lines, see Tommie Shelby, *Integration, Inequality, and Imperatives of Justice: A Review Essay*, 42 Phil. & Pub. Affs. 253, 273–79 (2014).

[43] Or, as Lily Hu nicely puts it: "[T]o riff on the words of Chief Justice John Roberts[,] . . . sometimes 'the way to fix inequalities in the category of race is to fix inequalities in the category of race.'" Lily Hu, *Disparate Causes, Pt. II*, Phenomenal World (Oct. 17, 2019), https://phenomenalworld.org/analysis/disparate-causes-pt-ii.

[44] *See*, *e.g.*, Parents Involved in Cmty. Sch. v. Seattle Sch. Dist. No. 1, 551 U.S. 701, 730–46 (2007) (plurality opinion); Texas Dep't of Hous. & Cmty. Affs. v. Inclusive Cmtys. Project, Inc., 576 U.S. 519, 553–55 (2015) (Thomas, J., dissenting).

allocative decisions with individual qualifications and nothing more. And even reasoning in terms of racial groups—as one must in order to appreciate the harms and mechanisms of racially patterned inequality—might strike these Justices as inconsistent with due respect for the standing of each person as an individual. I think this vision is profoundly mistaken, for reasons that I (and many others) have elsewhere explained, and for present purposes I will simply set it aside.[45]

Before turning to what the outlook that I have sketched suggests about the use of predictive algorithms, let me highlight a few of its salient features. First, the concern I have described is not centered on any *personal wrong* to an individual who is not chosen by some selective process.[46] As a matter of interpersonal morality, I do believe that some common forms of discrimination involve serious moral wrongs of that kind; in past work, I have focused in particular on the failure to show respect for the equality or autonomy of those who are disfavored.[47] But the particular normative concern that I have described here—which is the one I take to provide much of the justification for anti-discrimination laws—is very different. If we understand anti-discrimination norms as interventions to reduce patterned inequality, then particular individuals are granted rights under those norms for the simple reason that their fates are bound up with the larger social problem under attack. In other words, it happens that getting them into more favorable positions would be an improvement in the pattern.[48] There are undoubtedly some other individuals who are just as badly off and yet receive no similar aid, and there are surely others who are morally wronged by the decisions made about them, in the same domains of decision-making, and yet are given no legal recourse. Excluding those others from the protection of anti-discrimination laws is justified, from this point of view, not because they are less deserving of help or have not suffered as serious a wrong, but simply because helping them would not have the social benefit of undermining patterned inequality.

Second, the imperative to reduce patterned inequality is also not centrally concerned with any history of injustice toward the group that finds itself on the losing side of the

[45] For my explanation of why invocations of dignity and respect in support of "colorblindness" are misguided (and, indeed, perverse), see Benjamin Eidelson, *Respect, Individualism, and Colorblindness*, 129 Yale L.J. 1600 (2020). And for a sampling of other critiques of the conservative arguments for purging attention to race from the law, see *id.* at 1605 n.16.

[46] *Cf.* R. Jay Wallace, The Moral Nexus 163 (2019) (positing that "a [moral] duty is directed to another party only if the considerations that go into establishing the duty center around that party" and that "it is personal interests of the putative claimholder that will be prominent within such a person-involving justification").

[47] *See* Eidelson, *supra* note 33, at chs. 3–5.

[48] *Cf.* David B. Wilkins, *A Systematic Response to Systematic Disadvantage: A Response to Sander*, 57 Stan. L. Rev. 1915, 1939–40 (2005) (emphasizing how "the presence of black lawyers in the nation's legal, business, and government elites," due in part to affirmative action policies, "confers benefits to the black community as a whole —and to our nation and to the world").

pattern. It is true (of course) that the racialized pattern of inequality in the United States is a consequence of centuries of race-based oppression. But the moral case for reducing racially patterned inequality does not depend on that backstory; it just depends on the pattern of present facts that this history brought about. If the same facts—the same social meanings, the same patterns of material advantage and disadvantage, and so on—had somehow come about innocently, the case for intervening to reduce the patterned inequality, so as to thwart its ill effects, would not be any weaker. The case for reducing racially patterned inequality that I have outlined thus could not be faulted for resorting to any notion of "either a creditor or a debtor race."[49]

Third, the justification for anti-discrimination norms that I have sketched here depends on a premise that enforcing those norms will, in fact, reduce patterned inequality and its baleful consequences.[50] The justification could thus be defeated by showing that this premise is wrong, either in general or in some particular case. Suppose, for example, that a bank's criteria for making mortgage loans—criteria based purely on conservative, data-driven predictions about applicants' ability to repay the loan—will result in making loans to very few Black applicants. But suppose also that, if the criteria were relaxed so as to produce more loans to less advantaged applicants, many of those newly added debtors would go on to default and face foreclosure or bankruptcy—leaving them worse off than if they had never obtained the loans. It is possible that this would amount to a net setback to the effort to redress patterned inequality. And if so, the concern that I am highlighting here would favor the original lending criteria over the proposed revision, even though the original criteria yielded greater racial disparity in the bank's lending decisions.[51]

C. Algorithmic Predictions and Patterned Inequality

If we understand patterned inequality as a serious problem for the kinds of reasons that I have outlined—and as a problem to which anti-discrimination norms represent a partial response—then it is easy to see why allocating goods and opportunities based on algorithmic predictions is troubling in much the same way as other forms of discrimination. After all, the very point of the algorithms produced by machine-learning processes is to identify patterns in the distribution of some form of merit and thereby to predict

[49] Adarand Const. v. Peña, 515 U.S. 200, 239 (1995) (Scalia, J., concurring); *cf.* Anderson, *supra* note 13, at 153 (explaining that "[t]he integrative model [of affirmative action] takes a proactive stance toward [racial] injustices: its aim is to *dismantle* the continuing causes of racial injustice").

[50] *Cf.* Shin, *supra* note 38, at 171 (noting a similar qualification).

[51] The same would go, in principle, for other contexts in which concerns about "mismatch" that disserves the beneficiaries of an intervention are raised. *Cf.* Kennedy, *supra* note 42, at 145 (condemning "any initiative that knowingly or negligently over-promotes beneficiaries, placing them in settings in which they are conspicuously less prepared than nonpreferred peers, a situation rife with risks of demoralization and the creation or reinforcement of racist stereotypes").

accurately individuals' likely performance. And given a baseline of patterned inequality, succeeding in that endeavor is a surefire way to sustain or compound the existing pattern in the relevant domain.

Once again, it is important to appreciate that the concern here is not that the algorithm might be "biased"—at least not in the most familiar sense of the word.[52] To the contrary, the concern that I have outlined would have significant force even if we could somehow allocate opportunities based on the *actual* facts about who would perform best on some relevant metric. Disadvantaged groups will usually be underrepresented in that class—because they (and, in the context of intergenerational disadvantages, their parents) will have had inferior opportunities, on average, in the past. Disadvantage, in other words, is disadvantageous. And that means that allocating opportunities to those who would *in fact* perform best will extend, and likely exacerbate, patterns that give rise to grave harm and injustice.

Nonetheless, the imperfection of all actual predictions makes the problem significantly worse in practice than it would be even in that idealized hypothetical, because members of disadvantaged groups will usually be even more starkly underrepresented in the set of *predictable* best performers than in the set of actual ones. On average, that is, even the members of disadvantaged groups who would be among the top *n* candidates ex post will not look as promising as the others in that class ex ante. Their future high performance would be judged more surprising, because they will tend to share less in common with the other highest performers than the other highest performers share with each other, and they will tend to share more in common with lower performers than the other highest performers do. A process that picks the "best bets" will thus tend even more strongly toward picking the comparatively privileged, thereby affording them (and those whose fates are linked to theirs) still more opportunities, and at the same time denying yet more opportunities to members of worse-off groups.

The bottom line is very simple: Basing allocative decisions on even the most sophisticated predictive algorithms will tend to reproduce existing patterns in inequality and cement the matrix of stereotypes and social meanings that both cause and result from those patterns. And that, I am suggesting, is a central reason why allocating goods and opportunities in this way is cause for moral concern—not because it necessarily treats any individual unfairly, but because it cuts directly against the urgent project of scrambling existing patterns in societal inequalities. Indeed, from this point of view, the literature's

[52] Regarding the distinct concern that the data might overstate the riskiness of members of disadvantaged groups—which does amount to a form of "bias" in the ordinary sense—see *supra* note 10 and accompanying text. (And although I focus here on the concern that even a maximally accurate algorithm will contribute to patterned inequality, I do not discount the possibility that biases of various kinds are importantly connected to patterned inequality as well.)

emphasis on so-called "algorithmic fairness" is unfortunate. For one benefit of extending our thinking about discrimination to this new context should be an enhanced recognition of how little the normative case for anti-discrimination norms ever depended on fairness—in the intuitive sense of assessing individual merits without favor or prejudice—at all.[53]

II. COMPOUNDING INJUSTICE: HELLMAN'S ACCOUNT

In a series of illuminating papers, Deborah Hellman has developed an importantly different explanation of the justifying rationale for some anti-discrimination norms and, likewise, for concerns about the use of predictive algorithms to allocate goods and opportunities. According to Hellman, it is a pro tanto moral wrong to compound a prior injustice. And an actor compounds a prior injustice, as Hellman understands it, when the actor both "amplif[ies] the harm" that the injustice inflicted on someone and "take[s] the fact of [that person's] victimisation or its effects as her reason for acting."[54] I focus on Hellman's account here both because of its own influence and because I take it as the most developed form of a more general idea about the moral logic of anti-discrimination norms.[55]

Hellman's lead examples of "compounding injustice" involve the use of someone's prior victimization as a basis for adverse predictions about future outcomes. In one hypothetical case, a state decides to grant early release to prisoners with a low risk of recidivism. One "highly predictive" indicator of recidivism, Hellman posits, is a history of suffering child abuse.[56] Nonetheless, she suggests, the state has "a strong reason" not to include this variable in its predictive model: If the state denies someone early release because he suffered

[53] *See*, *e.g.*, Scanlon, *supra* note 13, at 42–43 (explicating a sense of "procedural fairness" according to which "decisions [must] be made on grounds that are 'rationally related' to the justification for [the] positions," and giving nepotism and cronyism as examples); *cf.* Loury, *supra* note 7, at 98 (noting "the tendency . . . to say that individuals are being treated unfairly and not being given their due" when they are subjected to "reward bias," i.e., not being rewarded equally for their qualifications). Of course, I do not deny that "fairness" can also be used in many other, increasingly extended senses; the literature on "algorithmic fairness" amply demonstrates that it can be. *See*, *e.g.*, Arvind Narayanan, *Translation Tutorial: 21 Fairness Definitions and Their Politics* (Mar. 1, 2018), https://youtu.be/jIXIuYdnyyk (surveying the many criteria that computer scientists have treated as possible definitions of "fairness"); *see also* sources cited *supra* note 6.

[54] Hellman, *Indirect Discrimination*, *supra* note 14, at 114.

[55] For example, Sophia Moreau identifies Hellman's approach as one of two possible pathways for those seeking to explain the wrongness of indirect discrimination in a significant class of cases (those in which "'recognition'-based accounts" do not apply). Sophia Moreau, *Equality and Discrimination*, *in* The Cambridge Companion to Philosophy of Law 171, 180 (2016). The second of the two is along the lines that I have previously defended in the context of indirect discrimination, *see id.*; *see also supra* note 40, and that I flesh out in more general terms in this article.

[56] Hellman, *Indirect Discrimination*, *supra* note 14, at 108–09.

child abuse, it will be adding to the harms caused by that earlier wrong—and it will be doing so not just incidentally, but by taking the fact of that wrong, or at least the fact of its anticipated effects, *as a reason* for treating the victim worse than others. The same analysis applies, Hellman suggests, to a life insurer's decision whether to use a woman's history of suffering domestic abuse as a factor in setting her premiums.[57] Although doing so might be rational in the sense that past abuse predicts a lower life expectancy, charging a woman more for that reason would wrongfully compound the prior injustice of the abuse itself.

As Hellman emphasizes, the "key element" in her account is that the actor must not only aggravate the harms caused by a prior wrong, but also "interact with the injustice or the effects of the injustice" in a way that makes the actor "implicate[d] in" it.[58] This "mixing element" is satisfied when, as I have just noted, the actor takes the prior wrong or its effects as a reason for action. To mark the contrast, Hellman offers the example of an entrepreneur who opens a business down the street from a struggling competitor. If the competitor's difficulty is due to a recent robbery, then the entrepreneur's market entry will amplify the harm of that prior wrong. But the entrepreneur does not *compound* that wrong, in Hellman's sense, unless the entrepreneur takes the robbery or the competitor's resulting difficulties as a reason for opening the new business now.[59]

Hellman posits that a moral prohibition on compounding injustice justifies significant strands of our body of anti-discrimination norms. First, she suggests that prohibitions on indirect discrimination (or "disparate impact") are justified in this way. Such norms impose a heightened standard of justification for selection criteria that disproportionately exclude members of particular groups. Thus, for example, an employer will need a good reason for using a standardized test on which white people tend to score higher than Black people. Such a demand for special justification, Hellman suggests, can be understood as frowning upon actions that compound injustice. For if a Black person scores poorly because of prior injustice that he or she has suffered (for instance, in being denied fair educational opportunities), then an employer who held the test score against him or her would be compounding that prior injustice. Of course, not *all* Black people who score poorly will do so because of prior injustice. But the stronger the correlation is between a person's race and his or her test score, the likelier it is that any given Black person who is being excluded by the test is also being subjected to the wrong of compounding injustice.[60]

[57] *Id.* at 110.

[58] *Id.* at 109, 112.

[59] *Id.* at 112.

[60] Importantly, and as Hellman acknowledges, "not all instances of indirect discrimination compound injustice" in this way. *Id.* at 114–15. For example, in *Dothard v. Rawlinson*, 433 U.S. 321 (1977), the U.S. Supreme Court recognized a claim of indirect sex discrimination based on minimum height requirements for prison guards. Skepticism of that requirement makes sense from the point of view of intervening against patterned inequality, but not as a way of avoiding compounding injustice in Hellman's sense.

Hellman offers a parallel account of the line of cases in the U.S. Supreme Court that frown upon some sex-based generalizations but allow others.[61] In the classic case of *Reed v. Reed,* for example, the Court struck down a state law that preferred men over women as administrators of estates. The generalization underlying the law may well have been sound: It stands to reason that, in Idaho in 1970, men had more of the relevant capabilities than women. But, as Frederick Schauer observes, "[i]f in 1970 men were on average better trained and more experienced in business and finance than women, the reason was surely that generations of women had been steered away from the world of business and finance and in the direction of occupations thought to be more suitable for their gender."[62]

Hellman thus argues that disfavoring a woman based on her (presumed) lack of business acumen would amount to compounding the prior injustice that she suffered in being denied the opportunity to develop that acumen in the first place.[63] Meanwhile, and in contrast to cases such as *Reed*, the Court has occasionally upheld sex-based distinctions that it understands to be predicated on "real"—roughly meaning biological—differences, such as the connection between sex and pregnancy. Hellman's view yields a natural explanation for that: If it were really true that the only causal mechanism of some difference between males and females is biology, then it would follow that the mechanism must not be a prior injustice suffered by anyone—and thus that there is no prior injustice for differential treatment predicated on that difference to compound.[64]

Finally, Hellman has recently extended the same line of thought to machine learning.[65] Suppose that some risk-assessment algorithm that informs lending or parole decisions predicts more bad outcomes for Black people than for white people. That will be because, as far as the data reflect, Black people more often possess characteristics that more often coincide with these outcomes. But if a given person possesses those characteristics as a result of having suffered an injustice—or if, as with the history of child abuse, the relevant predictive characteristic just *is* being the victim of a particular injustice—then holding those characteristics against the person amounts to compounding the prior injustice. People who are treated unfavorably on account of algorithmic predictions about them can thus be morally wronged—even though the predictions are not impeachable on grounds

[61] *See* Hellman, *Sex, Causation*, *supra* note 14.

[62] Schauer, *supra* note 36, at 140.

[63] Hellman, *Sex, Causation*, *supra* note 14, at 43.

[64] *Id.* at 33, 46; *see also supra* note 60 (explaining that Hellman's account does not extend to claims of indirect sex discrimination based on height requirements).

[65] *See* Hellman, *supra* note 8; Hellman, *Big Data and Compounding Injustice*, *supra* note 14.

of accuracy—in light of the causal history that explains their possession of the relevant risk factors.

III. IS THERE A DISTINCT WRONG OF COMPOUNDING INJUSTICE?

We now have two distinct ideas on the table: (1) the concern that certain allocative practices sustain or aggravate patterned inequality and (2) Hellman's argument that the same practices subject individuals to the wrong that she terms "compounding injustice." The ideas differ both in their temporal orientations and in the natures of the wrongs they identify. The argument from patterned inequality rests on the future consequences that, thanks to present patterns, a present selection practice will have—consequences that may ripple out to any number of other people injured by the pattern. In contrast, Hellman's argument identifies a personal wrong in disfavoring an individual, in the present, in light of how others treated her in the past. Despite these differences, the arguments reflect a common ambition to justify anti-discrimination norms on terms that apply even when a decision-maker does all that one could do to ensure that likes (in respect of a predicted attribute) are treated alike. I turn now to offer some grounds for doubt about Hellman's strategy for accomplishing that objective—and thus, indirectly, for centering the argument from patterned inequality in our understanding of why anti-discrimination norms have force even when there is no accuracy-based objection to be made.

A. The Role of Prioritarian Regard

In measuring these two accounts against one another, we should begin by identifying and extracting yet a third moral concern that is not uniquely attached to either of the two, but that figures importantly in Hellman's examples. Specifically, in each of Hellman's cases, the central actor makes things worse for some people—and not for just anyone, but for people who are already worse off than others through no fault of their own. Many would agree that this alone is a bad feature of an action from a moral point of view.[66] When our actions will affect other people, we should prefer choices that will not make things even worse for those who are already less well-off than others—or, at least, we should observe that preference if the existing disparity is not somehow deserved.[67] For concision, I will call this general moral concern *prioritarian regard*.

Hellman's examples highlight the relevance of this concern to the morality of many allocative decisions. Consider the life insurer that chooses to impose a higher premium on victims of abuse. The burdened class here is defined by an undeserved injury, so choosing

[66] *See*, *e.g.*, Derek Parfit, *Equality and Priority*, 10 RATIO 202, 212–14 (1997).

[67] This qualification makes for "desert-adjusted" or "desert-accommodating" prioritarianism. *See*, *e.g.*, KASPER LIPPERT-RASMUSSEN, BORN FREE AND EQUAL? 168–70 (2013).

to make the lots of those people even worse—for the benefit of others in the insurance pool or (more likely) for the benefit of the insurer itself—is a problematic choice from the point of view of prioritarian regard. And as Hellman's analogy underscores, something similar is happening in many paradigmatic cases of wrongful discrimination. One reason that it strikes us as morally problematic for an employer to use a standardized test that disproportionately excludes Black Americans, for instance, is that we know this decision means making individual Black candidates who are already undeservedly disadvantaged in myriad ways even worse off.

Nonetheless, the prioritarian regard to which I have just appealed is importantly different than Hellman's conception of compounding injustice—and so should not be mistaken as supporting it. For one thing, prioritarian regard includes no corollary to Hellman's "mixing" element; its force does not depend on what an actor takes as a reason for acting as he or she does. And even with respect to Hellman's "amplifying" element, prioritarian regard differs from Hellman's more specific concern in at least two ways. First, prioritarian regard is not sensitive (as Hellman's view is) to whether someone has suffered a prior *wrong* as opposed to some other misfortune.[68] And second, prioritarian regard does not depend (as Hellman's objection does) on whether the additional burdens that an already-disadvantaged person might face would be *causally due* to a prior disadvantage or not; rather, the force of the prioritarian concern just depends on how badly off that person already is. In other words, the prioritarian concern is about piling more burdens onto the same people, not pinning more effects to the same causal chains.[69]

Prioritarian regard adds something significant to the argument based on patterned inequality that I outlined in Part I; it furnishes a moral argument for anti-discrimination norms that is rooted in the interests of the particular individuals who are discriminated against and rests on the undeserved character of their existing disadvantage. Still, a concern of this kind can play only a modest, supporting role in justifying anti-discrimination norms. For one thing, it lacks any clear connection to socially salient groups. (Thus, the concern might apply with equal force to the use of criteria that correlate with educational

[68] Hellman thus asks whether "the fact that *injustice* produced [certain] effects constrains how others interact with these victims"—and, as explained above, answers yes. Hellman, *Sex, Causation*, *supra* note 14, at 509; *see also* Hellman, *Indirect Discrimination*, *supra* note 14, at 110 ("Should the insurance company take into account the fact that battered women are victims of wrongdoing when determining whether to include this relevant risk factor?").

[69] *Cf.* Kasper Lippert-Rasmussen, Algorithmic Discrimination and Compounding Injustice 14, 22 (n.d.) (unpublished manuscript) (on file with author) (drawing a like distinction between Hellman's proposed duty and an alternative "duty not to cause additional harms to the unjustly worse off").

attainment or, for that matter, with any facet of the so-called lottery of birth.[70]) Moreover, prioritarian regard itself can furnish no explanation of why the obligation to practice it should take the particular form of avoiding practices that disproportionately exclude disadvantaged individuals from the opportunities that employers (or others) have to allocate. For example, it could well be better, from a prioritarian point of view, for an employer to spare the costs it incurs for the sake of avoiding disparate impact and spend the savings on benefits for those who are truly worst off—people who generally will not have the borderline qualifications needed to benefit from disparate-impact prohibitions in the first place.[71] So although prioritarian regard does add something to the moral case for anti-discrimination norms—in that such norms presumably do make the world better from a prioritarian point of view—I doubt that this concern can play a central role in their justification.[72]

B. The Roles of Exploitation and Expression

If Hellman's account of "compounding injustice" is to earn its keep—and, in particular, if it is to play a central role in the justification of anti-discrimination norms—it has to identify a distinct wrong, going beyond the failure to afford due prioritarian regard, that occurs when an agent takes the fact or effects of a prior wrong as a reason for subjecting its victim to additional burdens. I will now argue that Hellman's arguments and examples do not warrant the conclusion that any wrong inheres in acts meeting that broad description.

It will help to begin with a related but narrower class of actions: those that *exploit* a prior wrong or unjust social arrangement for the actor's benefit. Patrick Shin's defense of rules prohibiting direct discrimination on the basis of race or sex—even when such discrimination is statistically rational—nicely captures the intuition that such acts are immoral.[73] As he puts it:

> What is wrong with using an enumerated factor as a statistical proxy for some employment-relevant deficiency (assuming a valid statistical relation does hold) is that it exploits the very circumstances of injustice that justify the enumerated

[70] *Cf.* Fishkin, *supra* note 35, at 37, 49–50. In *Reed*, for instance, a concern of this kind could not explain why sex is an unacceptable proxy for business acumen, but a high school degree—which will also separate the better-off from the worse off—would be a permissible one. Similarly, an employment practice with a disparate impact in terms of any axis of relative disadvantage—such as health or wealth—would be open to a comparable objection.

[71] *Cf.* Anderson, *supra* note 13, at 140 (noting that "[r]ough compensatory justice would be better served if," rather than employing affirmative action policies, "we distributed lump-sum cash reparations to every member of the disadvantaged group, or concentrated compensation on the least well-off within the group").

[72] Sophia Moreau voices similar doubts about prioritarianism as an account of anti-discrimination norms in Moreau, *supra* note 55, at 183–84.

[73] Shin, *supra* note 38, at 175–76.

> factor approach in the first place. A commitment to that approach … is based in part on the empirical judgment that the enumerated factors pick out categories of unacceptable inequality that exist or have existed in our society, such that we have reason to regard the use of those classifications as especially pernicious to the ends of justice. The use of an enumerated factor as a proxy for employment-relevant deficiencies is problematic because its rationale depends on the existence of, and then makes profitable use of, these same circumstances: it exploits correlations that arise out of the very conditions of injustice that the enumerated factor approach is intended to help eradicate.[74]

Supposing that we grant the intuitive force of Shin's argument, the question is *why* the employer's making profitable use of a prior injustice is wrong. Once we factor out both the effect of the employer's conduct on the maintenance of patterned inequality and the prioritarian regard discussed above, what is left in such an action for us to condemn?

I think the residual concern is that, in making use of unjust social conditions, one expresses a kind of comfort with or acquiescence in them.[75] In Hellman's life-insurance scenario, for example, the insurer seems somehow to treat a woman's victimization as tolerable by reducing it to just another actuarial variable. In reality, there is nothing inevitable about that inference about the insurer's attitudes: One can condemn an act as wrongful and also recognize its bearing on some question of independent relevance to one's decisions. Still, in the noisy register of what our actions are reasonably taken to express about our attitudes, a decisionmaker who accepts an unjust pattern as real and decisionally relevant—and adapts her own behavior so as to maximize her own welfare in light of it—might reasonably be heard to accept that injustice as, well, acceptable. This inference about the actor's attitudes will be especially reasonable against a backdrop of justifiable suspicion, on the part of victims, that those with power do not take the relevant wrong (such as domestic abuse) seriously.[76] Moreover, the very fact that the action could foreseeably be interpreted to convey an attitude of acquiescence or acceptance—and that the actor did it anyway—might say something additional about the meaning of his or her choice.[77]

[74] *Id.*

[75] *Cf. id.* (suggesting that "[i]f nothing else, one might say that the employer's action in such a case violates a general ethos" of the anti-discrimination project).

[76] *Cf.* Adam Omar Hosein, *Racial Profiling and a Reasonable Sense of Inferior Political Status*, 26 J. Pol. Phil. e1, e8 (2018) (arguing that the state acts unjustly by causing "members of a political community to have a reasonable sense of inferior political status," and that "whether it would be reasonable for someone to have this sense depends on the evidence available to her").

[77] For a fuller discussion of this point, see Eidelson, *supra* note 45, at 1619, 1622–23.

This expressive understanding of the wrong in exploiting injustice is buttressed by the fact that the wrong in such cases does not seem to depend on whether a person is actually subjected to any additional increment of harm on account of any prior injustice she suffered (as "compounding injustice" in Hellman's sense requires). Suppose, for instance, that a particular Black candidate for some opportunity has had the unusual good fortune to suffer little meaningful racial injustice in her life; perhaps she is a wealthy and recent African immigrant. Nonetheless, she is denied some benefit based on a generalization about the relative aptitudes or qualifications of Black candidates—a relationship traceable to existing racial injustice. The decision that is made about this candidate does not amplify a prior injustice that she suffered, but it seems to me to trigger the same unease that is at work in Shin's (and Hellman's) scenarios. I take that to be because—at least in light of present shared understandings about the meaning of actions that exploit injustice—the decision manifests the same problematic attitude of acceptance toward the underlying racial pattern.[78]

But if the moral concern triggered by Shin's and Hellman's examples is a concern about what a person expresses in making use of an injustice, that concern necessarily loses its force when the actor is not making use of an injustice at all. This is not a problem for Shin: His argument does not extend to the use of a "merit-based procedure" with disparate impact, precisely because the "injustice-exploiting dimension" is then absent.[79] For Hellman's argument about anti-discrimination norms, however, it is critical that the moral concern she identifies not be similarly limited. After all, the account is offered precisely to explain why "merit-based" selection processes do sometimes wrong those whom they disfavor. And yet I suspect that the account draws its appeal from the discomfort with exploitation that we have just identified and distinguished—perhaps joined with prioritarian regard and, in the background, the concern about perpetuating patterned inequality—and so does not reach many of the cases for which it is designed. In other words, when the actor cannot be said to be using any injustice against anyone—because the decisional criterion would be equally useful with or without the prior injustice—the expressive concern that powers Hellman's account fades away.

Suppose, for instance, that in Hellman's early-release case, the state administers a personality test that gauges an individual's propensity to commit violent crimes. And suppose that a history of suffering child abuse increases a person's likelihood of violence by means of affecting one of the broad aspects of personality that the test measures—an aspect of personality that would maintain its relevance even in a world without child abuse. Now imagine both that a prisoner is denied early release because of his test results, and also that

[78] As Brennan and Jaworski emphasize in another context, such meanings are often mutable. *See* Jason Brennan & Peter Martin Jaworski, *Markets Without Symbolic Limits*, 125 ETHICS 1053 (2015).

[79] Shin, *supra* note 38, at 176.

his test results would have been different if he had not suffered child abuse. Even if one is uneasy about the state's use of "history of child abuse" itself as a decisional criterion—because this uses the prior wrong or "mixes" the state's agency with it—I do not think there could be any like objection here. After all, the state cares only about crime-prone personality; it is completely indifferent to child abuse. What it owes to the victim of child abuse, I think, is precisely what it owes to others who are equally badly off, bear equal responsibility for their circumstances, and pose equal risks to others.

Or take the example of the business that opens alongside a competitor struggling with the aftermath of a robbery. If the new entrant is somehow seizing on the very fact of the robbery, maybe its action is improperly exploiting another's wrong. But suppose instead that the new entrant has a standing practice of setting up shop whenever there is unmet demand for its products in a given neighborhood. The company certainly could decide to make an exception in order to allow the robbed business to recover and meet the same demand instead. But it seems to me that would be essentially an act of charity—and almost surely a misdirected one, given others' far greater need for assistance of the same economic value. It is not something to which the robbed business has a moral entitlement.

Hellman might reply that, although the actors in these last two scenarios are not using *the wrong* against the victim, they are still using the *effects* of the wrong against the victim.[80] There is certainly a sense in which that is true, but I think it is not a sense with moral force. As Hellman's "mixing" imagery nicely conveys, the force of the "using" charge comes from the sense that the actor is, although not committing the wrong, engaging with it—that is, *with the wrongful act*—in a way that involves some positive orientation toward it. That is what gives the action its potentially problematic meaning. But the actor who simply applies merit-based criteria—criteria that some fail to satisfy because of past wrongs they have suffered and others fail to satisfy for other reasons—does not thereby orient his or her agency to the prior wrongs in any way. Even if he knows about them (which he might or might not), they do not figure at all in the justifying reasons for his action. And in such a case, it is difficult to see how his incorporating what are in fact "effects of the injustice" in his decision—without incorporating them under any description *relating to* the injustice—would "implicate[] him in the [prior] wrongdoing" or "make[] it in part [his] own."[81] Put another way: Why is such a case any different than the ordinary case (which Hellman would not count as wrongful) in which an actor knowingly aggravates the harms to another of a prior injustice, but acts for reasons that are causally independent of that injustice?[82] The prior injustice figures in the actor's reasoning to no greater degree in one kind of case than in the other—and yet, as Hellman's "mixing" condition

[80] *Cf.* Hellman, *Indirect Discrimination*, *supra* note 14, at 112–13.

[81] *Id.*

[82] Regarding Hellman's treatment of such cases, see *supra* notes 58–59 and accompanying text.

reflects, the actor's reasoning is essential to the felt wrongness of the actions in her examples (at least once any lapse of prioritarian regard is factored out of them).[83]

Now, because I have interpreted the wrong in exploitation cases in expressive terms, I have to acknowledge the possibility that there could be a problematic meaning expressed in these other, nonexploitation cases as well. And because the relevant meanings are determined in part by our shared beliefs and conventions—which bear on what we know to expect others to make of our choices—it is rarely possible to offer an abstract argument that a given action does not bear, to any extent, a problematic meaning. Nonetheless, any suggestion that the actors in the hypothetical cases I have just described convey acceptance of or acquiescence in the upstream moral wrongs—the wrongs of child abuse and robbery—seems to me to misapprehend our existing conventions. And although I am less certain about whether such a convention would be a desirable one to erect (a change that would make Hellman's moral objection sound by a kind of bootstrapping), I doubt that, too. A hypothetical shared understanding that those who employ merit-based selection procedures thereby express a positive attitude toward upstream injustices might have the benefit of inducing valuable departures from those means of selection. But, at least at its onset, this novel understanding would still be both inaccurate and unfair.

The better meaning to read into the use of merit-based selection procedures would be one that is actually true—and this returns us to the point about patterned inequality. As we saw in Part I, there are urgent moral reasons to avoid perpetuating the racial pattern of which any given Black candidate's disadvantage is a part. These reasons justify departing from narrowly merit-based selection criteria, or at least second-guessing the employer's choice of such criteria, as indirect discrimination norms demand. And the force of these reasons, in turn, means that we often *can* take an employer's use of merit-based decision procedures, heedless of their disparate impact, as expressing something—not an attitude toward past injustices, but a troubling willingness to sustain a society characterized by patterned inequality and its devastating effects. Significantly, that problematic meaning extends (like the objection from patterned inequality itself) even to cases that do *not* involve

[83] In her gracious reply to this article, Hellman adheres to her view that there is a wrong inherent in the class of actions that her definition of compounding injustice picks out (even independent of what such actions might express) and suggests that this "may be simply a point at which our [respective] intuitions lead in different directions." Deborah Hellman, *Personal Responsibility in an Unjust World: A Reply to Eidelson*, 1 Am. J.L. & Equality 282 (2021). I certainly join in her invitation for "the reader to mine her own intuitions on these points." *Id.* I would simply emphasize (which is not to say that Hellman would not) the importance of extending this mining endeavor to "all our judgments, whatever their level of generality"—and thus of not only cataloging one's verdicts on various particular cases, but also probing the intrinsic plausibility of the principles and distinctions to which one would need to appeal in order to ground the case-specific judgments that one is inclined to make. John Rawls, Justice As Fairness: A Restatement 29–30 (Erin Kelly ed., 2001) (describing the method of reflective equilibrium).

compounding injustice in Hellman's sense: When employers in traditionally male fields unnecessarily employ height requirements that disproportionately exclude women, for example,[84] they may both wrongfully contribute to patterned inequality and express a morally defective attitude toward those who will suffer from the continuing pattern, even though they do not disfavor any woman on account of a prior wrong that she suffered. But, in any event, the more fundamental point is that once we have accounted for all of these expressive and non-expressive considerations, they seem to squeeze out any remaining sense that making an adverse decision about someone based on a fact about him or her is wrong simply because that fact is itself due to some prior wrong. And if that is right, we should not seek to justify anti-discrimination norms on the basis of such causal connections.

IV. CONCLUSION: ALGORITHMS AND COMPOUNDING INJUSTICE

I began this article by emphasizing that the problems posed by algorithmic prediction do not solely concern failures to identify some form of merit as accurately as possible, but also concern the ways in which merit-tracking decisionmaking can itself be morally problematic. My central aim has been to distinguish, elucidate, and evaluate two different explanations of what those ways are—both of which, in fact, could intuitively be described as concerns about "compounding injustice." One points to the fact that purely merit-based decisions will tend to perpetuate patterned inequality that grievously constrains some people's flourishing and makes a mockery of equality of opportunity. The other, championed by Hellman, charges that the same merit-based decisions are moral wrongs because they implicate the decisionmaker in the prior injustice to which many of the affected individuals have been subject. I have suggested that Hellman's analysis highlights the relevance of prioritarian regard to allocative decisions, and that it might identify a problematic form of exploitation in some cases, but that it does not convincingly account for anti-discrimination norms that turn on the disparate impact of merit-based procedures.[85]

[84] *E.g.*, Dothard v. Rawlinson, 433 U.S. 321 (1977); *see supra* note 60.

[85] This argument could leave untouched Hellman's parallel account of direct sex discrimination (*see supra* notes 61–63 and accompanying text), because such cases *do* involve the use of an unjust pattern. But in that context, too, the cases and principles that Hellman analyzes might better be explained by the concerns that I emphasized in Part I. As the Supreme Court once explained the constitutional rule: "Intentional discrimination on the basis of gender by state actors violates the Equal Protection Clause, particularly where, as here, the discrimination *serves to ratify and perpetuate* invidious, archaic, and overbroad stereotypes about the relative abilities of men and women. . . . [We] acknowledge[] that a shred of truth may be contained in some stereotypes, but require[] that state actors look beyond the surface before making judgments about people that are *likely to stigmatize as well as to perpetuate historical patterns of discrimination*." J.E.B. v. Alabama, 511 U.S. 127, 130–31 (1994) (emphasis added). So understood, the norm's logic is forward-looking and systemic; it does not turn on the value of avoiding "compounding" an injustice that a present victim of discrimination may have suffered in the past.

That upshot, in turn, means that the backward-looking, complicity-based account also cannot capture the serious moral issues posed by the allocative use of algorithmic predictions. If an objection such as Hellman's is compelling only when a decisionmaker uses a criterion whose value rests on prior injustice, then such an objection will have little to say about a large share of the predictive relationships that algorithms are apt to uncover and employ. Many of those relationships—such as between criminal history and recidivism, or employment history and job performance—would surely emerge even without any upstream injustice in how the predictors themselves are distributed. Nonetheless, allocative decisions that are based solely on such predictions will certainly contribute to sustaining patterned inequality. And thus my broader conclusion: To understand and confront the threats that algorithmic selection processes pose, just as to understand anti-discrimination norms more generally, we have to attend to the effects that certain decision procedures have on the larger social pattern of resources, opportunities, and status—not the causal history by which any particular person came to occupy his or her place in the pattern.