CrossMark

Uncertainty in Hiring Does Not Justify Affirmative Action

Thomas Mulligan[1] iD

Abstract Luc Bovens has recently advanced a novel argument for affirmative action, grounded in the plausible idea that it is hard for an employer to evaluate the qualifications of candidates from underrepresented groups. Bovens claims that this provides a profit-maximizing employer with reason to shortlist prima facie less-qualified candidates from underrepresented groups. In this paper, I illuminate three flaws in Bovens's argument. First, it suffers from model error: A rational employer does not incur costs to scrutinize candidates when it knows their qualifications with perfect certainty, nor does it refuse to hire better-qualified candidates just because they did not require extra scrutiny. Second, Bovens's core premise–that there is greater variance in the evaluation of underrepresented candidates than there is the evaluation of other candidates–hurts underrepresented candidates rather than helps them. Third, candidates who are not shortlisted for the reasons Bovens gives have a plausible complaint about unfairness in the hiring process.

It is commonly thought that employers ought to hire the best-qualified applicants for jobs. This charge is typically justified in one of two ways. First, it may be a demand of *justice*: When we hire a less-qualified candidate (on the basis of his race, or because he is a friend, *etc.*), we do an injustice to the best-qualified candidate, who deserves the job on the basis of her merit. Second, meritocratic hiring may be required for reasons of efficiency: The owner of a widget factory will, generally speaking, make the most money if he hires the candidate who is best at making widgets.

✉ Thomas Mulligan
thomas.mulligan@georgetown.edu

[1] McDonough School of Business, Georgetown University, 37th and O Streets NW, Washington, DC 20057, USA

Springer

Luc Bovens has recently (2016) given a novel argument for affirmative action which rests on the simple and plausible idea that it is harder to evaluate the credentials of applicants from underrepresented groups than it is to evaluate other applicants' credentials. Bovens advances a model of hiring which has, as a surprising consequence, that a self-interested employer may prefer to shortlist *prima facie* less-qualified candidates from underrepresented groups. By showing such preference, Bovens claims, the employer maximizes the expected qualifications of its future hire. This is said to be a new justification for affirmative action, grounded in the "libertarian ideal that the business of business is business" (p. 434).[1]

Bovens's strategy is commendable: Another argument for affirmative action grounded in the promotion of diversity or the need to repair past racial wrongs would convince nobody who is not already disposed to the policy. But to show that a profit-maximizing firm may well want to give preference to racial minorities, as Bovens endeavors to do, is to appeal to precisely those in need of convincing.

However, I do not believe that Bovens is successful in this goal. I have three objections which I advance in this paper. First, Bovens's argument suffers from model error: A profit-maximizing business faced with the information Bovens considers should not behave as he suggests. Second, his core premise—that our evaluation of underrepresented candidates displays greater variance than our evaluation of other candidates—is likely to hurt the cause of racial equality rather than help it. Third, *pace* Bovens, there are reasonable worries about the fairness of the procedure he recommends.

I will rely on Bovens's nomenclature, with one exception. Those offering jobs (businesses, academic hiring committees, *etc.*) are *selectors*. The people applying for these jobs are *candidates*. Candidates' abilities to do the jobs at issue are their *qualifications*, denoted by q. Candidates send signals of their qualifications to selectors in the form of a *CV*, which in reality often includes more than the document itself (also: letters of recommendation, interviews, *etc.*) I denote this signal by θ. But instead of speaking of "underrepresented" and "overrepresented" candidates, I shall, for concreteness' sake, speak of black candidates and white candidates, respectively.

1 Bovens's Model Is Unrealistic in Important Ways

The idea that variance of one sort or another might have ramifications for hiring is not a new one. I am reminded of Larry Summers's comments on the underrepresentation of women in the sciences, and the outrage that they provoked (outrage which, I believe, was unfair). Summers conjectured that even if there is no difference in scientific ability between the average man and the average woman, greater variance in the distribution of ability for men might explain the dominance of men in science, since universities don't

---

[1] Whether this is in fact a libertarian ideal is debatable, as many libertarians will be loath to put any restrictions on owners' liberty to manage their businesses however they desire (see, *e.g.*, Narveson 1993).

hire people clustered around the mean but rather those who are a few standard deviations above it.[2]

Bovens concludes the opposite: that variance might in fact aid members of groups that have historically suffered from discrimination. In arguing for this conclusion, Bovens advances a two-stage model of hiring. In the first stage, candidates are selected for inclusion on a shortlist of determinate size; in the second stage, the shortlisted candidate whom the selector believes to be the most-qualified gets the job.

I reiterate that Bovens assumes that the goal of hiring is to maximize the qualifications of the chosen candidate. This normative framework enables his distinctive contribution to the literature, which is to show that self-interested, profit-maximizing behavior does not always include shortlisting the candidates whom the selector believes to be best: "It makes perfect business sense to shortlist with an eye to procuring the highest expected qualifications of the prospective hire. It does not make business sense to be told that we need to shortlist the candidates who have the highest expected qualifications." (p. 425).

Bovens is right that hiring in the real world often proceeds in two stages. Selectors frequently issue an unrestricted call for applications and cull from these a handful that merit more careful scrutiny. Note that the marginal cost of an application at this first stage is low—often just the time it takes to toss it onto the discard pile.

At the second stage—the examination of shortlisted candidates—marginal costs are high. In the case of academic hiring, for example, multiple members of the hiring committee must read multiple examples of the candidate's work; money is spent to fly her out for interviews and a research talk; students are asked to sit in on a teaching demonstration; and so on.

Bovens illustrates the way in which black candidates might be aided by variance in the selector's credence function by way of a toy example (p. 424). It is a useful example, and I reproduce it here:

	Candidates A, B, C (white)	Candidates D, E, F (black)		
Score:	8	7	8	9
Credence:	1	0.20	0.70	0.10

In Bovens's model, the selector arranges his shortlist (in this case, of size three) to maximize the expected qualifications of the new hire.[3] Bovens correctly notes that, with that aim in mind, shortlisting A, B, and C is inferior to shortlisting D, E, and F—the expected qualification of the former is 8, and of the latter, 8.26.[4] (Indeed, shortlisting D, E, and F is optimal.)

---

[2] At least as far as mathematical ability is concerned, Summers presented the empirical evidence fairly: There is no mean difference in ability between men and women, but men do display greater variance than women (see, *e.g.*, Hedges and Nowell 1995 and Hyde *et al.* 2008). Whether this explains, even in part, the underrepresentation of women is unclear.

[3] An optimal shortlist size will in fact be a function of both the costs involved in shortlisting and the uncertainty over the qualifications of the candidates who are being considered for inclusion on the shortlist.

[4] The probability that the selector gets a score 7 candidate is the probability that D, E, and F are all 7 s. This is $0.20^3 = 0.008$. The probability that the selector gets a score 9 candidate is the probability that at least one of D, E, and F is a 9. This is $1 - 0.9^3 = 0.271$. Therefore, the probability that the selector gets a score 8 candidate is $1 - 0.008 - 0.271 = 0.721$. The expected qualification is thus $(7 \times 0.008) + (8 \times 0.721) + (9 \times 0.271) = 8.263$.

But this model suffers from several defects which cast doubt on its generality. First, no real-world firm would proceed as the selector in Bovens's example does. A rational selector reasons as follows: "We have decided to shortlist three candidates. The purpose of the shortlist is to better understand our candidates' qualifications.[5] We know, with perfect certainty, that A, B, and C are all 8 s. So it would be a mistake to shortlist any of them; there is no potential benefit there. But we don't know the qualifications of D, E, and F, and there is at least some chance (10% each) that they are 9 s. So we will shortlist D, E, and F. Then we will know all the candidates' qualifications with perfect certainty."

It would, in fact, be doubly irrational for a selector to behave as Bovens's does. First, the correct justification for choosing D, E, and F in this case is that there is no informational gap to be filled when it comes to any of the A-Cs. Second, it would plainly be a mistake for the selector to insist on hiring one of the D-Fs if they all turned out to be 7 s (or 8 s—depending on how we wanted to handle ties between black and white candidates). In such a case, a rational selector concludes: "Upon further investigation the black candidates have turned out to be 7s. Now we know all the qualifications of candidates with perfect certainty—we have three white candidates (A-C) with qualification level 8 and three black candidates (D-F) with qualification level 7. So we will hire one of the A-Cs."[6] Under that reasoning, the expected qualification of the hire is 8.27—superior to that produced by Bovens's selector.[7]

It is not uncommon for real-world selectors to reach back into the applicant pool for a new candidate when their shortlisted candidates turn out, after scrutiny, to be wanting. It would be unusual if this new prospect were hired without further examination; but this is because, in contrast to Bovens's example, selectors grapple with uncertainty about the qualifications of *all* candidates, black and white alike.

And this is the second defect of Bovens's model: There is no uncertainty about the qualifications of the white candidates, A-C, at any stage in the hiring process—or, at least this is true in the example given. This feature gives rise to the unrealistic shortlisting behavior just described, but it also turns out to work against the interests of black candidates for reasons I will explain in the following section. For now, it suffices to note that it is actually very rare—if indeed it has ever happened—that a selector knows, with perfect certainty, the qualifications of a candidate.

It is true that, for Bovens's argument to go through, it is not required that the selector be perfectly certain about the white candidates' qualifications (*e.g.* modify the example such that the selector has credence 0.01 that their qualification level is 7; credence 0.98 that their qualification level is 8; and credence 0.01 that their qualification level is 9). But it is required that there be *more* uncertainty in the evaluation of black candidates than there is in the evaluation of white candidates—as Bovens himself says. As I will explain, that fact, along with the reasons for the difference in uncertainty between the two camps that Bovens gives, suffices to imperil his argument.

---

[5] "This is precisely why we are shortlisting: We need to get a better view of the candidates in order to reduce the variance in our assessment." (Bovens 2016: 424).

[6] I set aside the possibility of having to enter into battle with Human Resources, which might object—as is its wont—to the selection of a candidate not on the shortlist.

[7] Recall from n. 4 that the probability that the selector gets a score 9 candidate is 0.271. This time, if the selector doesn't get a score 9 candidate, it is guaranteed to get a score 8 candidate. So the expectation is $(8 \times 0.729) + (9 \times 0.271) = 8.271$.

The third defect of the model is that the second stage of hiring removes all uncertainty about all candidates' qualifications. This time, Bovens does concede that the feature is unrealistic; it is "unlikely to be the case but we make this assumption for modelling purposes" (p. 424). Again, this is a non-trivial assumption.

When it is made, there is no ambiguity about which candidate the selector ought to hire—it's the candidate with the highest q. But when we admit that *some* uncertainty will persist about candidates' qualifications, things are no longer so simple. Now the quantity of interest is not a candidate's q *simpliciter*, but rather the expected value of q given that candidate's CV. Again, the ramifications for this distinction will be fully explained in §2, but note here that the more difficult it is for a selector to evaluate a candidate's qualifications, the greater the incentive for the selector to rely on exogenous signals of candidate quality. And these signals could include facts about group qualifications which frequently work against the interests of members of groups who suffer from wrongful discrimination.

I close this section with a few comments about the empirical evidence that Bovens cites. I am sympathetic to the idea that it is harder to evaluate the qualifications of black candidates than white candidates, but it is essential for Bovens that this be true; for if not, then his argument, if it is sound, leads to an even less representative workforce.

In attempting to establish that there is more variance in the assessment of black candidates than in the assessment of white candidates, Bovens appeals to the empirical research on stereotype threat—"a kind of anxiety that is responsive to the social expectation that one will perform poorly" (p. 426). Bovens cites Steel and Aronson's (1995) study in support of this possibility, but serious worries have been raised about that study's conclusion by Sackett, Hardison, and Cullen (2004). And a recent meta-analysis (*viz.* Flore and Wicherts 2015) concludes that the literature on stereotype threat may be plagued by publication bias (only papers that find a statistically significant result get published; those that don't, don't).[8]

Bovens also raises the possibility that selectors suffer from *implicit bias* against underrepresented candidates—selectors may have formed injurious attitudes toward some groups and as a result might unconsciously discriminate against candidates from these groups during the hiring process. Here the evidence is more robust, but by no means settled.[9] Care is called for. There is a tendency to overlook the great diversity in hiring contexts and accept that evidence for bias in one is evidence for bias in them all. But in some contexts—such as academic hiring—progress has undoubtedly been made, and it is not at all clear that bias persists.

Bovens cites Steinpreis *et al.* 1999 in arguing for the existence of implicit bias against women in academic hiring, and I too have cited this study approvingly in the past. But it is dated, and the current data suggest that if one gender has an advantage on the academic job market, it is women. Williams and Ceci (2015) find that women are preferred over men in a ratio of 2:1, making it "a propitious time for women launching careers in academic science" (p. 5360). The authors subsequently found (Ceci and Williams 2015) that this preference does not violate meritocratic norms—gender is only

---

[8] See also Finnigan and Corker 2016 and Ganley *et al.* 2013.
[9] See, *e.g.*, Arkes and Tetlock 2004, Blanton *et al.* 2009, Forscher *et al.* 2017, Manuscript, and Oswald *et al.* 2013.

used as a "tie-breaker" between equally-qualified candidates, and thus would be permissible under the UK Equality Act, which Bovens criticizes.

As an aside, these results suggest that (1) hiring committees frequently find themselves deciding between finalist candidates whom they consider about equal in quality, and (2) a profession might significantly alter its gender demographics without giving up on a commitment to merit.

Along similar lines, Connolly, Lee, and Savoy (2015) find no significant differences between the genders when it comes to both initial hiring and promotion into the tenured ranks. Within philosophy, Dicey Jennings *et al.* (2015) find that women obtain permanent academic jobs in greater proportion than men do, and more quickly after completing their PhDs. And Allen-Hermanson's (2017) analysis of the philosophy job market concludes that "market outcomes starting in 2014 and going back 10 years offer no evidence women are at a disadvantage in tenure-track competitions.. . . The hypothesis that unconscious bias works against women in hiring and early career publishing is not well supported." (p. 8).

2 The Lesson of Statistical Discrimination

It is well-known that a rational selector may, when making employment decisions, give weight to *prima facie* irrelevant features of candidates, such as race. This is true even if the selector's owners, employees, and customers are in no way afflicted by racial animus. The reason is that the signals that a selector receives of candidates' qualifications are noisy, and race is both easily observable and, in some cases, correlated with qualifications. This is the core idea of *statistical discrimination*, first explored by Arrow (1973) and Phelps (1972).

In Bovens's normative framework, the selector seeks to hire the candidate with the highest q. But the selector cannot observe q directly. Again, all that the selector can observe is (1) a noisy signal θ (*i.e.* her CV), and (2) her race. Note first that θ is equal to q plus an error term ε that has a mean of zero and constant variance[10]:

$$\theta = q + \varepsilon. \tag{1}$$

We suppose that q and ε are normally distributed. Denote the mean of q by μ. The quantity of interest to Bovens's selectors is the expected value of q given θ, which is:

$$E[q|\theta] = (1-\gamma)\mu + \gamma\theta, \tag{2}$$

where γ is a measure of the reliability of θ,

$$\gamma = \frac{Var(q)}{Var(q) + Var(\varepsilon)}. \tag{3}$$

Let's look at the reliability measure, γ, and its effect on the quantity of interest, $E[q|\theta]$. Consider first the limiting case in which the CV is a perfectly reliable signal of

---

[10] Here I shall follow the models of Phelps and Aigner and Cain (1977).

qualifications; that is, $Var(\varepsilon) = 0$. Then $\gamma = 1$, and the first term in the right-hand-side of (2) vanishes. The selector puts no weight on the proxy of race. That makes sense. Then, as $Var(\varepsilon)$ increases, γ decreases (*ceteris paribus*), and the proxy is given more and more evaluative weight. This is intuitive: The noisier a signal is, the less useful it is to us.

Second, let's look at how $Var(q)$ affects γ. Again consider the limiting case of $Var(q) = 0$. Here, $\gamma = 0$, the second term in the right-hand-side of (2) vanishes, and $E[q \mid \theta] = \mu$. If there is no variance in the distribution of q, then the mean of that distribution is perfectly predictive. As $Var(q)$ increases, γ increases (*ceteris paribus*), and we come to rely more on the signal and less on the proxy. Why? Because the more variance there is in the distribution of q, the worse the mean becomes as a predictor of qualifications.

Bovens maintains that black candidates display greater variance than white candidates do, but this might mean two different things. First, it might be that q itself is distributed more broadly among blacks than it is among whites; that is, $Var(q_B) > Var(q_W)$. (The subscript "B" refers to blacks and the "W" to whites.) Call this "Type 1" variance. Larry Summers appealed to Type 1 variance in his comments about women in science.

Second, there may a difference in the quality of the signals sent by black candidates and white candidates; that is, there may be a difference in the variance of the error term ε: $Var(\varepsilon_B) > Var(\varepsilon_W)$. Call this "Type 2" variance.

We consider Bovens's scenario: The selector is evaluating black candidates and white candidates. We shall not assume that blacks are, on average, less-qualified than whites; to the contrary, we shall assume that $\mu_B = \mu_W$. The case of unequal means will be discussed later. We will assume that our candidates' qualifications fall above the mean, which is reasonable since selectors seek to hire the best candidates that they can—not the most average ones, nor the worst.[11]

Here is the problem for Bovens's argument: In the face of variance, it follows that θ is no longer an equally accurate predictor of q for blacks and for whites. In the face of Type 1 variance, the black candidate will have an advantage over the white candidate: If the two have equally strong CVs, the selector is justified in hiring the black candidate because her expected qualifications are higher. This follows from eqs. (2) and (3): $Var(q_B) > Var(q_W)$ means that γ will be larger for the black candidate than for the white candidate. That means that, relative to the white candidate, greater weight is placed on the black candidate's CV and less on the mean. Therefore, since we have assumed that $\theta > \mu$, Type 1 variance aids the cause of minority representation.

For Type 2 variance, the opposite is true; here the *prima facie* equally-qualified black candidate loses out to the white candidate. $Var(\varepsilon_B) > Var(\varepsilon_W)$ means that γ will be smaller for the black candidate than for the white candidate. And, by the argument just given, this harms black candidates.

Thus, in talking about variance in apparent qualifications between races, it is critical to identify the source of the variance. To the extent to which Bovens adduces Type 1

---

[11] Although this is a weak assumption, obviously satisfied in the real world, it is an important one, since everything gets reversed for candidates below the mean: $Var(\varepsilon_B) > Var(\varepsilon_W)$ helps rather than hurts blacks, and $Var(q_B) > Var(q_W)$ hurts rather than helps them. With equation (2) in mind, think of it this way: If you're above the mean, you want the signal to be as accurate as possible. If you're below the mean, you're aided by noise in the signal, since selectors will place greater weight on the mean, which exceeds your actual qualifications.

variance, his conclusion is strengthened. To the extent to which he adduces Type 2 variance, it is weakened.

Bovens appeals to four sources of variance, and three of these are clearly Type 2 variance. I have already mentioned (i) *anxiety*: the possibility that blacks might suffer from stereotype threat and do worse than whites during evaluation processes as a result; and (ii) *implicit bias*: the possibility that selectors do not accurately interpret CVs owing to unconscious prejudice. These are certainly generators of Type 2 variance; they are unrelated to the way in which qualifications are actually distributed within the candidate population.

The third source, *familiarity*, is also a source of Type 2 variance. If it is true that white candidates "typically come from the same cultural and educational background as the committee of selectors" and that this "leads to more uncertainty in assessing the qualifications of [black] candidates" (p. 426), then again we have a case in which there is noise in the signal of qualifications, θ, sent by candidates to selectors.

The fourth potential source of variance is *promise*: "[white] candidates have typically been exposed to an environment in which they can develop their talents and hone their skills, whereas many [black] candidates have not . . . hence we are more likely to be hiring [black] candidates on promise. Hiring on promise is more uncertain than hiring on actual achievements." (p. 426). I am not sure how to interpret this claim. Although it does appear to be a source of Type 1 variance, *promise* should, if anything, produce a *smaller* variance in blacks' qualifications, not a larger one. After all, if whites are, relative to blacks, raised in an environment in which they can more fully develop their human capital, this will lead to a broader, not a narrower, distribution of qualifications.[12] If you do not teach American children Chinese, they will all be equally bad at it; but if you do teach them, some will pick up the language more quickly than others— and so there will be variance.[13]

So *promise*, if it exists as Bovens believes, implies $Var(q_{\mathrm{B}}) < Var(q_{\mathrm{W}})$—not the other way around. And this, like *anxiety*, *implicit bias*, and *familiarity* (all of which imply $Var(\varepsilon_{\mathrm{B}}) > Var(\varepsilon_{\mathrm{W}})$), redounds to the benefit of whites, not blacks.

There is a second interpretation of *promise* worth considering.[14] For most positions, newly-hired employees undergo significant on-the-job training and socialization. And selectors want to know how effectively their candidates will grapple with these processes. There is Type 2 variance here.

But which way does it point? Well, for white candidates, who have already been "exposed to an environment in which they can develop their talents", there is relatively low uncertainty about how effectively they will grow into their role in the firm. Present qualifications predict future qualifications well. For black candidates, the opposite is true. Thus, under this, alternative interpretation of promise, $Var(\varepsilon_{\mathrm{B}}) > Var(\varepsilon_{\mathrm{W}})$—just like *anxiety*, *implicit bias*, and *familiarity*. Once again, black candidates suffer.

I stress that it is the *variance itself* that gives an advantage to whites in hiring; precisely those facts about the distribution of apparent qualifications that Bovens

---

[12] *Cf.* Aigner and Cain: "Blacks confront environmental restrictions on fulfilling their capacities, and this may lead to a smaller variance of [q_{B}]." (1977: 180 n. 13).

[13] I thank an anonymous referee for the evocative example.

[14] I thank an anonymous referee for suggesting this interpretation of *promise* to me.

appeals to in making his case for affirmative action in fact work *against* black candidates.

It is also worth pointing out that differences in mean ability will be taken into account by a rational employer, and these are likely, once again, to work against the interests of black candidates. Intuitively, and as eq. (2) makes clear, if there is any noise in the evaluation process, the selector will place weight on both the signal θ and the mean μ. And, in salient hiring contexts, if there is a difference between the races here it is that blacks have lower mean ability; that is, $\mu_B < \mu_W$. Indeed, this is a predictable consequence of what many proponents of affirmative action claim to be true: Namely, that (1) there is little genetic variation between the races, and (2) blacks face human capital-impeding discrimination and other forms of social disadvantage.

Thus, we have two models of hiring—Bovens's model and the statistical discrimination model—and they are incompatible, implying diametrically opposed selector behavior in the face of greater variance in the evaluation of black candidates. In §1 I pointed out some artificialities which characterize Bovens's model (and which are not present in the statistical discrimination model). And the fact that real world selectors do seem to care about profit maximization but yet don't shortlist high-risk, high-reward candidates suggests that Bovens's model is not an accurate one. The statistical discrimination model, on the other hand, may be applied at any stage in the hiring process; allows for variance (or not) in both the distribution of qualifications and in the error related to the CV; allows for mean differences (or not) in qualifications between candidate groups; and makes use of more general, continuously distributed variables. Finally, there is evidence that real world employers engage in statistical discrimination, and that black candidates suffer as a result.[15]

3 Is Bovens's Procedure Fair?

Bovens concludes his essay by arguing that the preference that he shows to black candidates does not violate the meritocratic "conception of *fairness* which requires that the best candidate get the job. What would be unfair on this conception is for a selector to knowingly appoint a less qualified person over a more qualified person" (p. 434). Bovens claims two things in support: First, "the selector did not *know* A, B or C to be better candidates" (p. 434), and, second, "the preference [for black candidates] was given during the selection procedure and not at the final hiring stage" (p. 435).

But as we have seen (§1), under Bovens's model, if D, E, and F turn out to be 7 s— the relevant case for considering questions of fairness—then at the second stage of hiring the selector does, indeed, know that A, B, and C are better candidates. And preference *is* shown at the final hiring stage, owing to the artificial requirement that a candidate cannot be hired unless he has undergone shortlist scrutiny (no matter the irrationality of such scrutiny).

There are other reasons to worry about the fairness of the procedure. Bovens regards race not as a merit *per se* (unlike manual dexterity, being black doesn't make you better at making widgets), but rather as a feature of a person which can be exploited in the name of efficient hiring. However, this is precisely the sort of argument used by those

---

[15] See, *e.g.*, Ewens, Tomlin, and Choon Wang 2014 and List 2004.

who prefer not to hire blacks, or women, or gays, *et al.* for statistical discrimination-related reasons. This is precisely the sort of argument used by those who advocate racial profiling—we'll minimize the occurrence of violent crime, they say, if we target only members of the race that is known, statistically, to commit violent crimes in greater measure. This reasoning may be perfectly correct, and the goal of minimizing violent crime a good one—but to say that there is no plausible complaint about fairness here is too quick, as evidenced by the public outcry over such practices. And if it makes us uneasy to discriminate on the basis of facts about the group mean, why shouldn't it make us uneasy to discriminate on the basis of facts about group variance? It seems to me that it should, and that this is true no matter whether we are talking about variance in the distribution of qualifications themselves, or, as Bovens suggests, variance in the distribution of the error term ε.

So to say that Bovens's procedure is unfair, as I have, is not to say that the statistical discrimination approach is without its own problems. What makes statistical discrimination especially confounding is that it is *unavoidable*: A selector can *never* assess qualifications without error, and *everything* is a proxy. This is true even in those rare contexts in which there exist highly accurate assessment mechanisms. Think of a factory manager who wishes to hire a new widget-maker for the assembly line. The manager subjects job applicants to a widget-making test, thereby getting an excellent sense of their qualifications. But it is not a perfect sense; some candidates will, for whatever reason, do better during the test than they will on the line; and for others just the opposite will be true. And when these differences are inconsistent across races, genders, *etc.*—which they might well be for myriad unobjectionable reasons—then a rational manager will make use of them even if he does not have a racist or sexist (*etc.*) bone in his body. And this is true, *a fortiori*, in more typical contexts like academic hiring, in which highly accurate assessment methods are not available.

For these reasons, the morality of statistical discrimination is complicated. I explore it in detail elsewhere (Mulligan 2017). I'll just point out, here, one way in which it is often unjustly used. This is when a relatively inaccurate proxy is applied in an unjustifiably strong way, or, worse, used as a litmus test to screen out candidates. Academic hiring in fact provides a perfect example of this: Setting demographic advantages (§1) aside, a candidate's pedigree—the identity of her PhD-granting institution—is the most important, indeed the overwhelming, determinant of whether she'll get a permanent academic job.[16]

The justification for relying on pedigree goes like this: (1) It's hard to assess a candidate's scholarly productivity (there's a lot of noise); (2) pedigree is positively correlated with scholarly productivity; and therefore (3) a pedigreed candidate should be preferred to a non-pedigreed candidate. Now, this is a valid argument. Indeed, in my view it's probably sound—the mean pedigreed candidate *really is* better than the mean non-pedigreed candidate. But it simply does not follow that pedigree should be afforded the consideration that it currently is. Why? Well, for one thing, because we know that there are *better* predictors of scholarly productivity than pedigree. Indeed, "there is little evidence of any independent effect of the prestige of one's PhD-granting department upon productivity, either at the predoctoral stage or later in the academic

---

[16] See, *e.g.*, Baldi 1995, Burris 2004, Clauset, Arbesman, and Larremore 2015, Jacobs 1999 and 2004, Keith and Babchuk 1998, McGinnis and Long 1997, and Oprisko, Dobbs, and DiGrazia 2013.

career. The strongest predictors of postdoctoral productivity are predoctoral productivity (publications and/or citations) and how quickly candidates complete the doctorate" (Burris 2004: 241). Thus, if an academic hiring committee wishes to hire the best candidate that it can, and it insists on appealing to a proxy for screening purposes (as many apparently do—see, *e.g.*, Marinoff 2009), it should appeal to the most accurate one available. And that's the publication record—not pedigree.

Whatever else justice requires when it comes to the use of statistical discrimination, it requires that (1) if selectors engage in litmus test screening, they apply the most accurate proxy available, and (2) if selectors engage in more holistic assessment, they weight proxies appropriately, which is to say in accordance with their accuracy.

Consider an analogy: A police department decides to deploy a "stop and frisk" task force in an attempt to get guns off the street. There are a finite number of officers on the task force, each of whom can stop only a finite number of pedestrians. A maximally efficient task force might engage in statistical discrimination, disproportionately stopping members of a minority race which is known, empirically, to be more likely to be carrying illegal weapons (*i.e.* the task force might engage in "racial profiling").

Now consider what a task force interested in efficiency *and justice* would do if some new research showed that some other statistical feature—dress, say—were more strongly correlated with the possession of illegal weapons. Wouldn't it be wrong—from the points-of-view of both efficiency and justice—for the task force to continue using race as the proxy for criminality? Yes, because a more accurate proxy—namely, dress—had been discovered. Whatever else we might want to say about how this task force ought to comport itself, it should certainly replace discrimination-based-on-race with discrimination-based-on-dress.

It is true that the task force might continue to stop and frisk, at disproportionately high rates, members of the minority race (if dress and race were correlated). But we would, I think, view it as a plain improvement to the status quo; after all, our task force, which has an obviously worthy goal—getting guns off the street—would both be doing a better job and treating pedestrians more justly. In a similar way, by bypassing pedigree to rely on more accurate proxies, like the publication record, academic selectors would treat applicants more justly while hiring better scholars. And if there is a correlation between scholarly quality and pedigree—as I suspect there is—then we will find that we have disproportionately hired pedigreed scholars. Indeed, this would be true even if pedigree played *no* role in candidate assessment.

I conclude by noting a subtle ambiguity related to the meritocratic hiring ideal: This is the question of whether meritocratic justice requires that (1) the best-qualified candidate be hired ("the more qualified person deserves to get the job" (Bovens 2016: 434)), or (2) the candidate with the greatest expected productivity be hired ("the business of business is business" (Bovens 2016: 434)). To illustrate how the two may come apart, suppose that Jones, who is black, is unambiguously the best-qualified candidate for a job at the widget store. The store's customers, however, are racists who will buy more widgets from a white widget salesman than a black widget salesman.[17] If justice is a matter of satisfying (1), as I believe (Mulligan 2017), then Jones will be hired. If justice is a matter of satisfying (2), he may not be. It is too quick to talk about meritocratic fairness *simpliciter*—these complexities must be considered.

---

[17] These are cases of *taste discrimination* (Becker 1957).

My own view—and perhaps Bovens would agree with me here—is that justice in hiring is unlikely to be achieved within the profit-maximizing normative framework (2). Within such a framework, statistical facts about group mean competence and variance in group competence will disadvantage precisely those candidates who have already suffered from wrongful disadvantage, like blacks. Meritorious men and women may be denied the jobs that they deserve because of the morally repugnant tastes of customers. And no room is left for allowing any facts about candidates unrelated to efficiency to enter into the calculus of justice—and yet these facts do sometimes seem to be relevant (consider, *e.g.*, that we might wish to give a job to a slightly less-qualified veteran who has suffered mightily for his country).[18]

I have argued that the larger variance in our evaluation of black candidates hurts rather than helps these candidates in hiring. But even if this were not the case—even if Bovens were right—I would still recommend against actually showing preference to black candidates in this way. Our meritocratic intuitions are so strong, and so widely held, that these tortuous justifications for preferring candidates whom we believe, using our best judgment, to be less-qualified will only produce resentment in the labor market, serve as a disincentive for human capital acquisition, and perpetuate nasty racial stereotypes. Racial justice cannot be achieved by showing preference to blacks at the hiring stage. It requires *ex ante* investment in education, healthcare, *etc.* for black children; that is, it requires that all citizens, regardless of their race, have an equal opportunity to develop their human capital. And this is, I believe, the goal we are striving towards: For all of us, black and white, an equal opportunity to develop our skills and then to be judged, not on the basis of our race, but on our merit alone.

References

Aigner, D. J., & Cain, G. C. (1977). Statistical theories of discrimination in labor markets. *Industrial and Labor Relations Review, 30*, 175–187.

Allen-Hermanson, S. (2017). Leaky pipeline myths: in search of gender effects on the job market and early career publishing in philosophy. *Frontiers in Psychology, 8*, 1–10.

Arkes, H., & Tetlock, P. E. (2004). Attributions of implicit prejudice, or "Would Jesse Jackson 'fail' the Implicit Association Test?". *Psychological Inquiry, 15*, 257–278.

Arrow, K. J. (1973). The theory of discrimination. In O. Ashenfelter & A. Rees (Eds.), *Discrimination in Labor Markets* (pp. 3–33). Princeton: Princeton University Press.

Baldi, S. (1995). Prestige determinants of first academic job for new sociology Ph.D.s 1985–1992. *Sociological Quarterly, 36*, 777–789.

Becker, G. (1957). *The Economics of Discrimination*. Chicago: University of Chicago Press.

Blanton, H., Jaccard, J., Klick, J., Mellers, B., Mitchell, G., & Tetlock, P. E. (2009). Strong claims and weak evidence: reassessing the predictive validity of the IAT. *Journal of Applied Psychology, 94*, 567–582.

Bovens, L. (2016). Selection under uncertainty: affirmative action at shortlisting stage. *Mind, 125*, 421–437.

Burris, V. (2004). The academic caste system: prestige hierarchies in PhD exchange networks. *American Sociological Review, 69*, 239–264.

Ceci, S. J., & Williams, W. M. (2015). Women have substantial advantage in STEM faculty hiring, except when competing against more-accomplished men. *Frontiers in Psychology, 6*, 1–10.

---

[18] I discuss these issues of justice in hiring in detail in my Mulligan 2017.

Clauset, A., Arbesman, S., & Larremore, D. B. (2015). Systematic inequality and hierarchy in faculty hiring networks. *Science Advances, 1.*

Connolly, M. R., Lee, Y.-G., and Savoy, J. N. 2015. Faculty hiring and tenure by sex and race: new evidence from a national survey. Paper presented at the Annual Meeting of the American Educational Research Association, 16-20 April 2015.

Dicey Jennings, C., Kyrilov, A., Cobb, P., Vlasits, J., Vinson, D. W., Montes, E., & Franco, C. (2015). Academic placement data and analysis: 2015 final report. *At* https://www.dropbox.com/s/b9p0dx817 qkcb3e/APDAFinalReport2015.pdf?dl=0.

Ewens, M., Tomlin, B., & Choon Wang, L. (2014). Statistical discrimination or prejudice? A large sample field experiment. *Review of Economics and Statistics, 96*, 119–134.

Finnigan, K. M., & Corker, K. S. (2016). Do performance avoidance goals moderate the effect of different types of stereotype threat on women's math performance? *Journal of Research in Personality, 63*, 36–43.

Flore, P. C., & Wicherts, J. M. (2015). Does stereotype threat influence performance of girls in stereotyped domain? A meta-analysis. *Journal of School Psychology, 53*, 25–44.

Forscher, P. S., Lai, C. K., Axt, J. R., Ebersole, C. R., Herman, M., Devine, P. G., Nosek, B. A. (2017). A meta-analysis of change in implicit bias. https://osf.io/preprints/psyarxiv/dv8tu/. Accessed 22 July 2017.

Ganley, C. M., Mingle, L. A., Ryan, A. M., Ryan, K., Vasilyeva, M., & Perry, M. (2013). An examination of stereotype effects on girls' mathematics performance. *Developmental Psychology, 49*, 1886–1897.

Hedges, L. V., & Nowell, A. (1995). Sex differences in mental test scores, variability, and numbers of high-scoring individuals. *Science, 269*, 41–45.

Hyde, J. S., Lindberg, S. M., Linn, M. C., Ellis, A. B., & Williams, C. C. (2008). Gender similarities characterize math performance. *Science, 321*, 494–495.

Jacobs, D. (1999). Ascription or productivity? The determinants of departmental success in the NRC quality rankings. *Social Science Research, 28*, 228–239.

Jacobs, D. (2004). Ascription and departmental rankings revisited: A correction and a reanalysis. *Social Science Research, 33*, 183–186.

Keith, B., & Babchuk, N. (1998). The quest for institutional recognition: A longitudinal analysis of scholarly productivity and academic prestige among sociology departments. *Social Forces, 76*, 1495–1533.

List, J. A. (2004). The nature and extent of discrimination in the marketplace: evidence from the field. *Quarterly Journal of Economics, 119*, 49–89.

Marinoff, L. 2009. Inside a search. *Inside Higher Ed,* <https://www.insidehighered.com/advice/2009/08/31 /marinoff>, retrieved 3 July 2017.

McGinnis, R., & Long, J. S. (1997). Entry into academia: Effects of stratification, geography and ecology. In M. J. Finkelstein & P. G. Altbach (Eds.), *The Academic Profession: The Professorate in Crisis* (pp. 342–366). New York: Routledge.

Mulligan, T. (2017). *Justice and the Meritocratic State*. New York: Routledge.

Narveson, J. (1993). *Moral Matters*. Peterborough, Canada: Broadview Press.

Oswald, F. L., Mitchell, G., Blanton, H., Jaccard, J., & Tetlock, P. E. (2013). Predicting ethnic and racial discrimination: a meta-analysis of IAT criterion studies. *Journal of Personality and Social Psychology, 105*, 171–192.

Oprisko, R. L., Dobbs, K. L., and DiGrazia, J. 2013. Honor, prestige, and the academy: A portrait of political science tenured and tenure-track faculty in Ph.D.-granting institutions (2012-2013). Paper presented at the 2013 annual meeting of the American Political Science Association.

Phelps, E. S. (1972). The statistical theory of racism and sexism. *American Economic Review, 62*, 659–661.

Sackett, P. R., Hardison, C. M., & Cullen, M. J. (2004). On interpreting stereotype threat as accounting for African American-White differences on cognitive tests. *American Psychologist, 59*, 7–13.

Steele, C. M., & Aronson, J. (1995). Stereotype threat and the intellectual test performance of African Americans. *Journal of Personality and Social Psychology, 69*, 797–811.

Steinpreis, R. E., Anders, K. A., & Ritzke, D. (1999). The impact of gender on the review of the curricula vitae of job applicants and tenure candidates: a national empirical study. *Sex Roles, 41*, 509–528.

Williams, W. M., & Ceci, S. J. (2015). National hiring experiments reveal 2:1 faculty preference for women on STEM tenure track. *Proceedings of the National Academy of Sciences, 112*, 5360–5365.