California Law Review Online

Vol. 7 March 2017

Incomprehensible Discrimination

James Grimmelmann* & Daniel Westreich**

The following (fictional) opinion of the (fictional) Zootopia Supreme Court of the (fictional) State of Zootopia is designed to highlight one particularly interesting issue raised by Solon Barocas and Andrew Selbst in Big Data's Disparate Impact.[1] *Their article discusses many ways in which data-intensive algorithmic methods can go wrong when they are used to make employment and other sensitive decisions. Our vignette deals with one in particular: the use of algorithmically derived models that are both predictive of a legitimate goal and have a disparate impact on some individuals. Like Barocas and Selbst, we think it raises fundamental questions about how anti-discrimination law works and about what it ought to do. But we are perhaps slightly more optimistic than they are that the law already has the doctrinal tools it needs to deal appropriately with cases of this sort.*

DOI: https://dx.doi.org/10.15779/Z38707WN47

* Professor of Law, Cornell Tech and Cornell Law School. We thank Solon Barocas and Andrew Selbst, Kristen Bertch, Kiel Brennan-Marquez, Pauline Kim, Frank Pasquale, David Robinson, and Aaron Rieke for useful suggestions.

** Associate Professor of Epidemiology, University of North Carolina Gillings School of Global Public Health.

1. Solon Barocas & Andrew Selbst, *Big Data's Disparate Impact*, 104 CALIF. L. REV. 101 (2016).

After the statement of facts and procedural history, you will be given a chance to pause and reflect on how the case ought to be decided under existing United States law. Zootopia is south of East Dakota and north of West Carolina. It is a generic law-school hypothetical state, where federal statutes and caselaw apply, but without distracting state-specific variations. The citations to articles, statutes, regulations, and cases are real; RDL v. ZPD *and* Hopps v. Lionheart *are not.*[2] *Otherwise, life in Zootopia is much like life here, with one exception:*

It is populated entirely by animals.

RUMINANT DEFENSE LEAGUE *v.* ZOOTOPIA POLICE DEPARTMENT

CHIEF JUSTICE UPDOG delivered the opinion of the Court.

Respondent Zootopia Police Department (the ZPD) uses a mathematical model to predict which applicants will be successful police officers. Four facts about this model are undisputed. First, its scores are significantly correlated with a reasonable measure of job performance. Second, the model does not explicitly consider applicants' species. Third, it nonetheless systematically favors carnivorous applicants over herbivores. Fourth, no one has explained how and why the model works at predicting job performance or how and why it disadvantages herbivores. The question presented is whether the ZPD's use of such a model in making hiring decisions constitutes discrimination "on the basis of . . . race" prohibited by Title VII of the Civil Rights Act of 1964.[3]

I.

The ZPD is the largest police department in the world and the tenth-largest employer in Zootopia. Its name is synonymous with urban policing—and also, unfortunately, with employment discrimination. The landmark *Hopps v. Lionheart* litigation exposed a long tradition of intentional discrimination by ZPD leadership in promoting carnivores over herbivores. It documented numerous cases in which promotion decisions were based on blatant stereotypes, such as that herbivores, accustomed to being prey, would be too "meek" and "cowering" to serve effectively in leadership positions. Formal and informal

2. Readers may already have recognized "Zootopia" from the recent animated film of the same name. ZOOTOPIA (Walt Disney Animation Studios 2016). In *Zootopia*, animals of all shapes and sizes live side-by-side in imperfect harmony. Judy Hopps, the first rabbit officer to join the Zootopia Police Department, is given forty-eight hours to solve a the case of a kidnapped otter. She enlists the reluctant assistance of a small-time con-artist fox named Nick Wilde. Together, the two of them find that the otter and thirteen other animals have been drugged with a serum that makes them revert to a feral state, then track the scheme back to an unexpected source. We have borrowed *Zootopia*'s setting, a few of the names, and its theme of irrational prejudice in employment discrimination—but that is as far as the connection goes, and nothing in this essay should be regarded as reflecting any sponsorship, approval, or endorsement on the filmmakers' part.

3. 42 U.S.C. § 2000e–2(b) (2016).

policies alike relegated herbivores to lower ranks and to lower-status assignments, like parking enforcement. There was also abundant evidence of anti-herbivore animus among the ZPD hierarchy, with herbivorous officers being referred to dismissively by epithets such as "vegetable-huggers," "lazy grazers," and "breakfast." The plaintiffs argued that the ZPD's policies constituted forbidden discrimination "on the basis of . . . race."

In *Hopps*, we held that Title VII applies to sentient talking anthropomorphic animals as well as to humans, and that the term "race" as used in Title VII includes an animal's species. Although neither carnivores nor herbivores constitute a "species" as such, we further held that discrimination on the basis of an immutable species characteristic—such as diet—is discrimination "on the basis of . . . race."

Hopps resulted in a consent decree under which the ZPD agreed to make promotion decisions strictly in accordance with a new composite measure of job performance. This measure, the Bellwether Index (or BI), incorporates case closure rates, numerical evaluations by supervisors, civilian complaints, and performance on a standardized examination. Those elements were selected as being broadly representative of the various responsibilities of ZPD officers while also fairly recognizing the achievements of officers of all species. Although that portion of the consent decree was lifted more than a decade ago, the ZPD continues to rely substantially on BIs in making promotion decisions. That practice is not at issue today.

Instead, the present suit challenges the ZPD's decision-making process in hiring new officers, where for obvious reasons BIs are not available. Two years ago, in an effort to reduce the attrition rate from the famously rigorous ZPD Academy and to improve the quality of its officer force, the ZPD used data-mining techniques to rework its hiring processes.[4] As described in the declaration of Officer Benjamin Clawhauser, who oversaw the project, it comprised four stages:

> *First*, Clawhauser assembled an extensive set of *training data* consisting of the ZPD's nearly complete personnel files on all of its past and present employees. These files contain thousands of items of information, or *factors*, on each employee, including for example their educational history, snack food preferences, credit score, best time in the 100-meter dash, favorite song by the pop singer Gazelle, and home address. We say "nearly complete" because Clawhauser stated that he scrubbed all protected demographic characteristics—race, sex and gender identity, sexual orientation, religion, country of origin, and species —from the personnel files before using them as training data.
>
> *Second*, Clawhauser selected a measurable *target variable* to serve as a proxy for overall job performance. He chose the Bellwether Index because experience under the consent decree and after had shown that it

4. *See generally* Barocas & Selbst, *supra* note 1 (describing data mining).

> was clearly defined, straightforward to extract from the ZPD's personnel files, and representative of ZPD officers' various responsibilities.
>
> *Third*, Clawhauser used data-mining methods to extract algorithmically a *model* of job performance based on the training data. The model is a function whose inputs are the values of the factors for a particular applicant, and whose output is a *score* (scaled from 0 to 250) predicting that applicant's Bellwether Index after five years of employment if he or she were to be hired. The function is remarkably complex and the significance of many of the factors it employs is obscure. For example, the model predicts that given two applicants whose factors otherwise match Clawhauser's own, the applicant whose favorite Gazelle song is "Try Everything" will have a BI that is 1.8 points higher than the BI of the applicant who prefers another song.
>
> *Fourth*, the ZPD began using the model's scores to *classify* applicants as "hire" or "do not hire." The threshold has varied slightly, but initially, applicants with a score of 120 or higher were hired; those with lower scores were not.

In the five years immediately preceding the ZPD's adoption of the new process it had hired representatively diverse cohorts of new officers: the fraction of herbivores among all officers hired was statistically close to the fraction of herbivores among all applicants, and both were statistically close to the fraction of herbivores among all inhabitants of Zootopia. Since the ZPD switched to model-based hiring, it has hired almost exclusively carnivorous applicants.

Two years later, petitioner Ruminant Defense League (the "League"), on behalf of its members, filed suit in the District Court for the District of Zootopia, claiming that the ZPD's use of the model to make hiring decisions violated Title VII. In discovery, the ZPD produced the training data from which the model was developed. The parties stipulated that the following subset is representative of the entire set of training data. The "species" and "carnivore" columns are in italics because the model does not explicitly take species into account. We include it to illustrate the effect that use of the model has on carnivorous and herbivorous applicants.

	BI	Model Score	*Species*	*Carnivore?*
Amy	50	100	*Antelope*	*No*
Bert	100	150	*Bear (Polar)*	*Yes*
Charles	150	100	*Capybara*	*No*
Denise	200	150	*Dingo*	*Yes*

Both parties moved for summary judgment. The District Court granted the ZPD's motion. First, it held that because the model did not explicitly take species into account, use of the model could not constitute disparate treatment of

herbivores. Second, it held that even if use of the model had a disparate impact on herbivores, the model's predictive power made it legal as a business necessity. It rejected the League's argument that the model inherently discriminated against herbivores because they consistently scored lower than carnivores.

Only these two holdings are before us, but we pause to note that in its thorough opinion the District Court also rejected a number of other arguments raised by the League, in each case on factual rather than legal grounds. In particular, the ZPD offered insufficient evidence to create a material issue of fact that the ZPD's choice of BI as the target variable was intended to disadvantage herbivores; that the ZPD's choice of a threshold of 120 for its hire/no-hire decisions had a more significant effect on herbivores than other possible choices, or was intended to; that the training data in the ZPD's personnel files was in any way inaccurate; that the training data was insufficient to derive reliable predictions about applicants or classes of applicants; that the training data reflected any existing pattern of discrimination against or animus against herbivores by others, such as civilians filing disciplinary complaints; that herbivores were over- or under-represented in the training data; or that the choice of features included in the training data had a more significant effect on herbivores than other possible choices, or was intended to. These findings were not challenged on appeal and so we take them as true for purposes of our opinion today. The District Court expressed no opinion on whether such findings would constitute a violation of Title VII in a case in which they were present, and neither do we, other than to note that they involve cases in which the accuracy of data-mined predictions would be *undermined* by discrimination rather than apparently *bound up* with it. It is hard to see, for example, that an employer has any legitimate interest (other than expedience) in using an inadequately small training dataset. A better dataset would yield predictions that were both more accurate and fairer. As such, we believe that these forms of potential bias pose less fundamental challenges for anti-discrimination law than the present appeal does; here, the goals of accuracy and fairness appear to be in tension.

The League appealed, arguing that on the undisputed facts, the ZPD's use of a biased model constituted disparate impact and disparate treatment as a matter of law. The Court of Appeals affirmed. It first held that because the model does not explicitly consider species and because the ZPD did not select or employ the model with the purpose of discriminating against herbivores, use of it cannot constitute disparate treatment. Turning to the League's disparate-impact argument, the Court of Appeals held that the ZPD was justified in relying on the model because it was predictive of job performance, notwithstanding its disparate impact. More specifically, because the League had shown that use of the model had a disparate impact on a protected class, the burden shifted to the ZPD to show that use of the model qualified as a business necessity. The Court of Appeals then held that ZPD carried this burden by showing an "undisputed statistically and practically significant correlation between the model score and

an undisputed measure of job performance," shifting the burden back to the League to show that an alternative employment practice would have had the same predictive power with less discriminatory impact. In the Court of Appeals' view, the League had not carried this burden: "We are sympathetic to the League's argument that it is being asked to improve on an algorithm it did not create and does not understand. But without a concrete and less discriminatory alternative, the League cannot be heard to say that the ZPD's algorithm is *unnecessarily* biased." A dissenting judge would have held that use of a model that "predictably and substantially disadvantages herbivores" constitutes prohibited disparate impact, notwithstanding the model's correlation with job performance, and that even if the ZPD's initial motivations were benign, "knowingly continuing to use a biased model" constitutes prohibited disparate treatment.

In view of the important issues presented by the case, we granted certiorari.

STOP.

The facts above present what Barocas and Selbst call the problem of "proxies for proscribed criteria."[5] *Barocas and Selbst argue that current doctrine requires a result along the lines of the Court of Appeals'. While they are normatively sympathetic to the dissent, they believe it has little likelihood of becoming the law. In their view, correcting for the disparate impact of a model in which a proxy is correlated both with job performance and with a protected category would require the kind of conscious rebalancing that the Supreme Court has condemned as disparate treatment in* Ricci v. Destefano.[6] *"At some point, society will be forced to acknowledge that this is really a discussion about what constitutes a tolerable level of disparate impact in employment," they conclude. "Under the current constitutional order and in the political climate, it is tough to even imagine having such a conversation. But, until that happens, data mining will be permitted to exacerbate existing inequalities in difficult-to-counter ways."* [7] *Are they right? Or is it possible to make at least some progress in cases like* RDL v. ZPD*? How should the Zootopia Supreme Court decide the case?*

5. Barocas & Selbst, *supra* note 1, at 693. *See also* Pauline T. Kim, *Data-Driven Discrimination at Work*, WM. & MARY L. REV. (forthcoming 2017), *available at* http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2801251 (discussing problem of proxies).

6. 557 U.S. 557 (2009).

7. Barocas & Selbst, *supra* note 1, at 728.

II

In our view, the District Court and Court of Appeals failed to accord the proper significance to the fact that no one—neither the League, nor the ZPD, nor any of the capable judges who have heard this case—has been able to give a satisfactory explanation of *why* the ZPD's model both predicts job performance and discriminates against herbivores, or of the relationship between these two facts. The League is correct that the factors that the model identified correlate with species, and the ZPD is correct that these factors also correlate with job performance. The problem is that there is no explanation in the record as to which of these two correlations, if either, is causal. It may be that the factors directly measure applicant characteristics that determine success in the challenging and dangerous field of police work, and that those characteristics happen to be unequally distributed in our diverse society. It may also be that these factors are instead measuring applicants' species and that they measure likely job performance *only because they are identifying species* in an applicant pool where the relevant characteristics are unequally distributed. We believe that where a plaintiff has identified a disparate impact, the defendant's burden to show a business necessity requires it to show not just that its model's scores are not just *correlated* with job performance but *explain* it. The Court of Appeals improperly placed the consequences of ignorance about the model's operation on the League, instead of on the party best positioned to understand and control it, the ZPD.

A

We begin by emphasizing that this is not a case about the choice of target variable. The portion of our opinion in *Hopps* upholding the consent decree against an objection by the Patrolrodents' Benevolent Society stands for the proposition that the ZPD may rely on BIs in making promotion decisions. There, we held that the Bellwether Index is a facially neutral and generally accepted measure of job performance, so that its use in promotion decisions does not constitute disparate treatment.

As for disparate impact, the subset of training data listed above gives a good sense of why *Hopps* also upheld the use of BIs against a disparate impact challenge. Specifically, within the stipulated subset, the two carnivores have a mean BI of 150 (the average of 100 for Bert and 200 for Denise), while the two herbivores have a mean BI of 100 (the average of 50 for Amy and 150 for Charles). On average, an herbivore scores 50 points lower. To put it another way, species correlates with BI, so that knowing an officer's species is sufficient to predict some (though not all) of her BI. We can assume, as the League asserts, that this difference primarily reflects a legacy of pre-*Hopps* discrimination against herbivores. In many years, this disparity has little or no effect on actual promotions. A typical promotion threshold is 130, under which Charles (an herbivore with a BI of 150) and Denise (carnivore, 200) would be promoted

while Amy (herbivore, 50) and Bert (carnivore, 100) would not. Although BIs are not evenly distributed across the different populations of officers, there is a broad range of performance within each population, so the actual promotion decisions are representatively diverse: one carnivore and one herbivore. In some years, the disparity in BIs is reflected in disparate promotion rates. In a year where the promotion threshold is set at 170, the ZPD will promote one carnivore Denise (200) but no herbivores at all, not even Charles (150).

We held in *Hopps* that this disparate impact could be justified as a business necessity, given the comprehensive nature of the Bellwether Index as a measure of job performance and the lack of good and less discriminatory measures. While a lower threshold might promote Charles along with Denise, it would also on average decrease case clearance, civilian satisfaction, and other core police functions. In our view, Title VII did not require such a tradeoff.

Today, we extend that holding to the use of BIs in decisions pertaining to hiring as well as to promotion. The strong consensus that the Bellwether Index is an appropriate measure of job performance in one context makes it an appropriate target variable in the other. We recognize that there is a difference between hiring and promotion decisions: applicants are hypothetical apples and officers are actual oranges. Applicants' BIs must be predicted, while officers' BIs can be measured. But any rational hiring process must rest upon some kind of predictions about future job performance, so prediction is inherent in the nature of the task. The League has much to say about the difficulties of prediction, and so will we. We do not believe these arguments go to the target variable itself. The League accepts that the ZPD may hire new officers with a view toward their future job performance, and the Bellwether Index remains the best legally cognizable standard anyone has come up with to define job performance for an officer of the ZPD.

Other target variables may be problematic to the point of being illegal; this one is not.

B

But as we said, this is not a case about a target variable. This is a case about the use of a model to *predict* a target variable. The considerations applicable are different in a subtle but significant way.

As a starting point, suppose by way of example that the ZPD had observed the correlation between species and BI and instituted a blanket policy of hiring only carnivores. Such a policy would be a *per se* violation of Title VII. There is a broad social consensus, reflected in Title VII's disparate-treatment theory, that where a prohibited characteristic does no more than point to pre-existing patterns its use in employment decisions is illegitimate. Even though carnivores are statistically more successful as ZPD officers as measured by the Bellwether Index, Title VII forbids projecting that statistical correlation onto individual applicants on the basis of their species. Given the ugly history of discrimination

and oppression that has often accompanied them through the centuries, such characteristics may not be used as proxies for employment criteria, however strong the correlation. Their use is intrinsically harmful, even if unintentional.

The ZPD asserts that it has done something different: used data mining algorithms to create a model that enables it to hire on the basis of likely BI, rather than on the basis of species. The model is indeed predictive of BIs. A 1-point increase in the model score predicts a 1-point increase in BI Here, the two officers with model scores of 100 had an average BI of 100; the two officers with model scores of 150 had an average BI of 150.

The problem is that maybe the ZPD's model score is successfully identifying species *and only species*. The model score is *moderately* correlated with BI, to be sure, but it is *perfectly* correlated with species. The two carnivores (Bert and Denise) received model scores of 150; the two herbivores (Amy and Charles) received model scores of 100. Knowing an officer's model score does not provide any more information about his or her BI than just knowing his or her species would. It might be the case that the ZPD's model is effectively using all of the data-mined factors to fill in the missing "species" column that Officer Clawhauser deleted from the personnel files—and then inferring a predicted BI from an applicant's species.

Consider another example. Suppose that the ZPD's model had been based on a single factor: home address. This would not strictly speaking involve considering an applicant's species. Still, home address is highly predictive of species; tigers and lions do not live in Little Rodentia. A model that gave a score of 150 to applicants from Tundratown (where Bert the polar bear lives) and from Outback Island (where Denise the dingo lives) and 100 to all others would perfectly replicate the scores of ZPD's actual model on the subset of training data listed above. It would thus be just as correlated with BI as the actual model is. To the extent it was predictive of job performance, it would be right for the wrong reason. Such a model is not actually measuring anything relevant to job performance; instead, it measures a different but prohibited factor that is itself correlated with job performance.

The ZPD argues at great length that this is not what it is doing, that its model really is something other than a proxy for species. We are unconvinced. We could be convinced. But it will take something more than the evidence presently in the record. While the ZPD may be *trying* to predict BIs rather than species, it has not (yet) shown that it has succeeded in doing so. Unless and until it does, we believe it has failed to carry its burden of showing that the model it uses is a genuine business necessity. We will reverse and remand to give the ZPD an opportunity to make such a showing, on a proper evidentiary record and under a proper legal standard.

C

We add a few remarks about arguments the District Court may consider on remand. The ZPD could rehabilitate its model by showing that the connection between the target variable and the factors the model relies on is more than just coincidental or correlational. It could, for example, identify factors with particularly high weights in the model and explain why those factors meaningfully relate to an animal's ability to ably discharge his or her duties as a sworn officer of the ZPD. Notably, such a showing would go beyond purely statistical proof; it would require an explanation in terms of the chains of causation by which one state of affairs in the world leads to another. Sometimes such explanations come readily to hand: animals who can run faster and for longer will have an easier time apprehending criminals in hot pursuit. In other cases, the explanations may be more obscure. We do not currently see how an officer's preference among Gazelle songs could causally relate to solving crimes—but perhaps there is such a connection, and one that the ZPD can articulate to the District Court on remand. For example, empirically validated evidence that listening to up-tempo songs causes officers to be more motivated to exercise, thereby significantly increasing their ability to exert themselves when needed on the job, would show that musical preferences are not just predictive but predictive for the right kind of reason. A good explanation of this sort is one that identifies the hidden and non-discriminatory variables connecting the observed factors with the predicted target variable. In the previous example, an officer's motivation to exercise is the relevant hidden variable. Up-tempo musical tastes are causally connected to motivation to exercise, which is causally connected to success in apprehending criminals.

We do not hold that this second type of showing is absolutely required when a Title VII defendant asserts a business necessity for a data-mined model with a disparate impact. That said, there are good policy reasons to prefer it. One is intelligibility; such explanations make black-box algorithms more comprehensible to those who use them, those who are affected by them, and those who oversee them.[8] Anti-discrimination law is made by animals for the benefit of animals; algorithmic decision-making must ultimately be accountable to animals, just as much as animal decision-making is.

Our revised test for business necessity, we hope, will increase this understanding of how these algorithms work. By shifting the burden of proof to explain why a hiring model works in cases where that algorithm also has troubling side effects, the test encourages the party best positioned to understand the model to understand it.[9] The ZPD has full access to the training data, the

8. *See generally* FRANK PASQUALE, THE BLACK BOX SOCIETY: THE SECRET ALGORITHMS THAT CONTROL MONEY AND INFORMATION (Harvard University Press, 1st ed. 2015).

9. *Cf.* Kim, *supra* note 5 ("Similarly, in disparate impact cases courts tend to defer to employer judgments about what abilities or skills are necessary for a job when evaluating employer justifications

data-mining algorithms, and the model's decisions. It should bear the costs of obscurity—not the applicants whose fates are in the model's hands but who have no ability to observe the model's development or inner workings. And even if the ZPD fails in developing a convincing explanation, effort devoted to understanding its model is likely to bear useful fruit. If a model predicts job performance by way of species because it accurately identifies where animals live, pinpointing that fact might enable an employer to prune factors that are predicting job performance for the wrong reasons. Or consider again the musical tastes example. Perhaps some officers listen to up-tempo music because they are already motivated to run hard when necessary. Then musical tastes do not cause job success, but identifying motivation as the connecting causal hidden variable might allow the ZPD to redesign its model to assess motivation more directly, thereby coming closer to hiring officers for the right reasons.

Our revised test will also, we hope, encourage parties who use algorithms to make sensitive decisions to *disclose* important facts about how those algorithms work. We place the burden on the ZPD to build a public record establishing that its algorithms work as described and for the right reasons. The ZPD introduced a complex data-mined model into its hiring process; if the progress of public understanding lags behind the progress of technology, the ZPD is best positioned to close the gap, either by mining less or by explaining more.

At various stages of this litigation, the ZPD has maintained that it is legally prohibited from saying more about its model and the data-mining process that produced it because they are protected personnel records that could reveal sensitive private information about ZPD officers, and because they are trade secrets of the private-sector information-technology vendors whose services Officer Clawhauser and the ZPD have employed. While we respect the ZPD's operational needs for confidentiality, we do not believe it is appropriate to shield algorithms from accountability on these grounds. We will not compel the ZPD to publicize the inner workings of its hiring algorithms. But unless it is willing to allow some sunlight into the workings of a model with a disparate impact on a protected class of animals, we believe it should be from precluded on relying on the model's predictive accuracy in defending its conduct. Any other rule would allow employers to accept the benefits of algorithmic decision-making while sloughing off the corresponding responsibilities.

D

Our holding today is not without precedent. Other bodies of law require explanations for *why* models behave the way they do when making important decisions. The Fair Credit Reporting Act, for example, requires disclosure of "all

for a practice. However, data mining models often rely on 'discovered' relationships between variables, rather than measuring previously identified job-related skills or attributes.").

relevant elements or reasons adversely affecting the credit score for the particular individual, listed in the order of their importance" when an animal is denied credit based on a credit score.[10] This is a start, but it stops short of requiring that the factors used by the credit-scoring model need to actually be relevant, rational, or intelligible. Disclosure that one's credit score was driven down by the phase of the moon would suffice—even if that arbitrary choice had catastrophic consequences for capybaras.

The Office of the Comptroller of the Currency (OCC) recognized this gap in its prescient 1997 guidance on compliance issues posed by credit scoring models. Those compliance issues include discrimination on the basis of prohibited characteristics in granting credit; the issues are essentially the same as in the employment context. The OCC stated that it would conclude that a variable used in a credit scoring system "is justified by business necessity and does not warrant further scrutiny if the variable is statistically related to loan performance, and *has an understandable relationship* to an individual applicant's creditworthiness." [11] Today, we endorse the OCC's italicized phrase; we will not accept an incomprehensible model as a business necessity.

We therefore also reject the suggestion of the Uniform Guidelines on Employment Selection Procedures that purely statistical analysis can suffice to establish a model's validity.[12] Two of the three ways that the Guidelines contemplate validating an employment selection procedure include some kind of causal explanation. Content validity, in which an applicant is assessed for "a representative sample of the content of the job," directly measures the activities and behaviors that they will perform on the job.[13] Here, there is less danger that the model is serving as a proxy for a prohibited characteristic because the inferential chain (from doing something well while being assessed to doing the same thing well while on the job) is too short to include the protected characteristic. On the other hand, construct validity involves "identifiable characteristics which have been determined to be important in successful performance in the job."[14] Here, there is less danger that the model is serving as

10. 15 U.S.C. § 1681g(f)(2)(B).

11. Office of the Comptroller of the Currency, Credit Scoring Models, OCC Bull. No. 97-24, app. 11 (May 20, 1997), http://www.occ.gov/news-issuances/bulletins/1997/bulletin-1997-24.html [https://perma.cc/R6GK-XQNK] (emphasis added). *See also* National Consumer Law Center, Credit Discrimination 137 n.116 (6th ed. 2013) (stating that the OCC's guidance "may be indicative of how other federal regulators will view this issue.") We are grateful to *amici curiae* David Robinson and Aaron Rieke for drawing our attention to these authorities.

12. *See* 29 CFR §§ 1607.5(B) ("[A] criterion-related validity study should consist of empirical data demonstrating that the selection procedure is predictive of or significantly correlated with important elements of job performance."), 14(B)(5) ("Generally, a selection procedure is considered related to the criterion, for the purposes of these guidelines, when the relationship between performance on the procedure and performance on the criterion measure is statistically significant at the 0.05 level of significance. . . .").

13. *Id.* § 1607.14(C)(1).

14. *Id.* § 1607.5(B).

a proxy because those "identifiable characteristics" serve as the missing link in the inferential chain (from observations about the applicant to the identifiable characteristics to job performance). In each case, there is a plausible causal story with a reasonable evidentiary foundation that does not involve a protected characteristic. With content validity, the causal chain is too short for such characteristics to sneak in; with construct validity, the chain is longer but properly connected. Both strike us as sufficient.

E

We close with an observation about the structure of Title VII. Our modification to the burden-shifting framework of disparate impact borrows from disparate treatment doctrine. In our view, Title VII does not permit an employer to do indirectly what it could not do directly. An employer that explicitly selects applicants on the basis of species violates Title VII under a disparate treatment theory, regardless of whether species is correlated with job performance, and regardless of whether it bears animus against particular species. It is the selection "on the basis of" species that is the problem. An employer that uses home address to infer applicants' species and then selects applicants from particular species does exactly the same, only in two steps rather than one. This too is a form of disparate treatment. The ZPD's model is more complex, but if that model selects applicants based purely on their species, the ZPD is still effectively engaged in disparate treatment, regardless of whether it realizes that that is how its model works.[15]

We regard this indifference to the employer's knowledge about the inner workings of a discriminatory model as a virtue of the test we announce today. We do not want to deter employers from examining closely the models and algorithms they use. A test that turned only on the employer's knowledge of how its model functions would discourage employers from looking too closely at models that superficially seemed to work. Where a model has a disparate impact, our test in effect requires an employer to explain why its model is not just a mathematically sophisticated proxy for a protected characteristic.

We do not regard this as an unreasonable burden on employers. They are not required to produce such explanations in all cases, but only when (a) they have delegated their selection procedures to an algorithmically derived model and (b) that model yields *prima facie* discriminatory results. An employer that develops a selection procedure the old-fashioned way, with animalian design and supervision, can ask the procedure's designers to explain it. And an employer that develops a model algorithmically is in the clear if that model does not have a disparate impact.

---

15. On the (difficult) relationship between disparate impact and disparate treatment doctrine in this context, *see generally* Barocas and Selbst, *supra* note 1; Kim, *supra* note 5; George Rutherglen, *Disparate Impact, Discrimination, and the Essentially Contested Concept of Equality*, 74 FORDHAM L. REV. 2313 (2006).

III

Our holding today is simple. *Incomprehensible discrimination will not stand*. Applicants who are judged and found wanting deserve a better explanation than, "The computer said so." Sometimes computers say so for the wrong reasons—and it is employers' duty to ensure that they do not.

The judgment of the Court of Appeals is reversed, and the case is remanded for further proceedings consistent with this opinion.

It is so ordered.