When "Less Is More": How Statistical Discrimination Can Decrease Predictive Accuracy

Felipe A. Csaszar
Stephen M. Ross School of Business
University of Michigan
fcsaszar@umich.edu

Diana Jue-Rajasingh
Stephen M. Ross School of Business
University of Michigan
djuerjs@umich.edu

Michael Jensen
Stephen M. Ross School of Business
University of Michigan
michjen@umich.edu

August 11, 2022

Forthcoming in *Organization Science*

Acknowledgements: The authors thank Jerker Denrell and three anonymous referees for their insightful and constructive reviews. For valuable suggestions, the authors thank Cha Li, David Ross, and Dashun Wang. The authors also thank the seminar participants at the Academy of Management Meeting, Nanyang Business School, and the University of Michigan. Errors remain the authors' own.

When "Less Is More": How Statistical Discrimination Can Decrease Predictive Accuracy

Abstract

Discrimination is a pervasive aspect of modern society and human relations. Statistical discrimination theory suggests that profit-maximizing employers should use all the information about job candidates, including information about group membership (e.g., race or gender), to make accurate predictions. In contrast, research on heuristics in psychology suggests that using less information can be better. Drawing on research on heuristics, we show that even small amounts of inconsistency can make predictions using group membership less accurate than predictions that do not use this information. That is, whereas statistical discrimination theory implies that better predictions can be achieved by using all available information about an individual (including group characteristics that may be correlated with but do not cause performance), our model shows that using all available information only improves predictive accuracy under a very specific set of conditions, thus suggesting that statistical discrimination often results in worse predictions. By understanding when statistical discrimination improves or worsens predictions, our work cautions decision makers and uncovers paths toward reducing the occurrence of situations in which statistical discrimination benefits predictive accuracy, thus reducing its pervasiveness in society.

Keywords: Discrimination, labor market discrimination, statistical discrimination, heuristics

1 Introduction

When is more information less helpful in hiring decisions? To predict who will perform well in a particular job, is it always better for an employer to use as much information as possible about job candidates? Statistical discrimination theory suggests that profit-maximizing employers should use all the information about job candidates that correlates with job productivity to determine whom to hire (Arrow 1973, Bertrand and Duflo 2017, Phelps 1972). This is especially true when candidate-level data about productivity is hard to observe, since observable group characteristics such as race and gender that correlate with (but do not cause) productivity become valuable predictors of individuals' productivity (Aigner and Cain 1977, Correll and Benard 2006). However, statistical discrimination's use of non-causal information to make predictions of individual performance can lead to the unequal treatment of equally productive individuals and create great concerns about fairness (Roemer 1998). This type of discrimination is particularly insidious, as statistical discrimination has been argued to improve predictive accuracy and, hence, is likely to be rewarded by the market.[1] Indeed, statistical discrimination theory rationalizes not only the use of group-specific information like race and gender to make individual hiring decisions but also the persistence of group disparities in labor market outcomes (Tilcsik 2021).

While statistical discrimination theory implies that more information is better to make accurate predictions, research on heuristics suggests that less information can be better.[2] Specifically, research on heuristics—simple decision-making rules that ignore information—has shown that they can make more accurate predictions than rules that use more information (Gigerenzer and Brighton 2009, Gigerenzer and Gaissmaier 2011). For example, heuristics can increase predictive accuracy because of decision makers'

---

[1] Statistical discrimination is a dominant perspective in economics. Other perspectives on discrimination exist such as taste-based discrimination in economics (Becker 1957), status-based and institutional discrimination in sociology (see Correll and Benard 2006, Pager and Shepherd 2008) as well as stereotype and prejudice in psychology (Allport 1954, Park and Judd 2005). Given our focus on integrating statistical discrimination and heuristics research, it is beyond the scope of our study to provide comprehensive reviews of these alternative perspectives.

[2] The less-is-more effect in hiring decisions is not limited to making accurate predictions of future productivity, the focus of our study. Recent research suggests that using random selection in the hiring process might also increase the chances of achieving a diversity performance bonus by increasing the likelihood that employees have non-overlapping cognitive repertoires (Liu 2021).

limited cognitive ability to distinguish useful information from environmental noise or optimally incorporate useful information (Csaszar and Ostler 2020, Gigerenzer and Brighton 2009).

Drawing on research on heuristics, we develop a model of discrimination that shows that using less information can improve predictive accuracy.[3] While statistical discrimination theory only focuses on two aspects of the decision problem (the *unobservability* of true individual productivity and the *correlation* between an observable group characteristic and productivity), our model adds realism by incorporating three additional parameters: *uncertainty* in the environment, decision maker *inconsistency*, and *diagnostic bias* of the productivity signal. Among other results, we show that even slight deviations from the optimal use of information implied by statistical discrimination can make the predictions of a statistical discriminator less accurate than the predictions of someone who does not take into account group characteristics to predict a candidate's future performance. This is a finding that runs counter to statistical discrimination theory. Thus, we show how the idea that "less is more" can be extended to further understand and challenge one of the leading conceptualizations of discrimination.

Several contributions emerge from our behavioral model of discrimination. First, we demonstrate that the less-is-more effect, a well-known insight from the heuristics literature (Gigerenzer and Brighton 2009, Gigerenzer and Gaissmaier 2011), has important implications for statistical discrimination. We do so by adding realism to statistical discrimination theory through the consideration of uncertainty, inconsistency, and diagnostic bias in the decision-making process. This highlights important boundary conditions of statistical discrimination, thus qualifying previous work. In particular, we show that a decision maker interested in maximizing predictive accuracy seldom benefits from using discriminatory information even if this is correlated with individual productivity. Second, we extend upon statistical

[3] We define the costs and benefits of not discriminating from the perspective of the individual decision maker in terms of his or her predictive accuracy. Non-discrimination is costly if it decreases predictive accuracy and beneficial if it increases predictive accuracy. We do not include other important economic, social, and legal costs of discrimination, such as consumer backlash, inability to recruit and retain employees, and litigation. We focus on predictive accuracy because predictive accuracy is relevant for all decision making regardless of the other costs and benefits of non-discrimination whose importance vary more from context to context. Our emphasis on predictive accuracy makes our model comparable to those of statistical discrimination, which focus on the same outcome.

discrimination theory by incorporating mechanisms that are prevalent in other disciplines outside of economics. Third, we provide an efficiency argument for abstaining from statistical discrimination, as we show how statistical discrimination can lead to worsened, not improved, predictive accuracy. More generally, we contribute to a growing understanding of how human behavior can lead to unequal outcomes for equally productive individuals by providing a stylized yet realistic way of modeling it. In practical terms, our model helps characterize the situations under which discrimination increases predictive accuracy the most, which suggests interventions that can reduce the likelihood of these situations occurring.[4]

To show that using more information in hiring decisions can be detrimental in hiring decisions, we start by comparing the use of information in statistical discrimination versus heuristics research.

2 Theoretical Background and Motivation

In this section, we first review research on statistical discrimination, which emphasizes that more information allows for more accurate predictions of individual productivity, and research on heuristics, which points in the opposite direction—that less information can lead to more accurate predictions. We then combine the contingencies present in statistical discrimination with those in heuristics to distill the five parameters in our model of discrimination.

2.1 Statistical Discrimination: More Information is Better

Statistical discrimination theory stresses the benefits of using more information to identify high-productivity individuals. Bertrand and Duflo (2017, p. 311) note: "As a profit-maximizing prospective employer … tries to infer the characteristics of a person … they use all the information available to them. When the person-specific information is limited, group-specific [characteristics] may provide additional

[4] Like other models in management and strategy research (e.g., Levinthal 1997), our model is both descriptive and normative. It is descriptive in the sense that it describes how decision makers discriminate given well-accepted behavioral assumptions. It is normative in the sense that it can be used to derive optimal behavior and make recommendations on how decision makers should behave in a given situation.

valuable information about expected productivity." Regardless of the actual specification of the statistical discrimination model, statistical discrimination theory suggests that employers use observable group characteristics including race or gender to proxy or fill in for unobservable individual productivity (Arrow 1973, Bielby and Baron 1986, Phelps 1972). In contrast, non-discriminators do not use any information about group characteristics.

Prior research documents the presence of statistical discrimination by showing that decision makers rely on group characteristics to make decisions when more individual information is not available, thus resulting in unequal treatment of individuals from discriminated and non-discriminated groups that are otherwise similar. The presence of individual-level information has the effect of equalizing treatment. For example, Rissing and Castilla (2014) show that when detailed employment information is not available for immigrants applying for employment-based green cards, labor certification approvals for immigrants differ significantly depending on foreign citizenship. In contrast, when detailed employment information is provided, labor certification approvals are equally likely across foreign citizenship. Similarly, Fernandez and Greenberg (2013) show that when employers make use of referrals, which can inexpensively signal the high quality of job applicants, referred minority applicants are more likely to receive a job offer than non-referred minority applicants. Moreover, referred minority applicants are treated similarly to referred majority (white) applicants. These studies show that individual-level information can reduce unequal treatment between individuals from discriminated and non-discriminated groups, implying that decision makers are statistically discriminating when individual-level information is unavailable or unobservable.

Most statistical discrimination theory relies on two assumptions: high unobservability of the true productivity of a candidate before hiring and high correlation between group characteristics and productivity. Because of these two assumptions, statistical discrimination theory implies that it is rational for employers to use more information—observable group characteristics—as a proxy for individual productivity. However, unobservability and correlation in statistical discrimination theory are limited in their ability to describe more realistic decision-making situations. As we demonstrate, when more realism is added to models of statistical discrimination, it may not be better to use more information (i.e., the group characteristics of a particular candidate) to improve predictions about productivity.

2.2 Heuristics: Less Information is More

A heuristic is a "strategy that ignores part of the information" in order to make decisions "more quickly, frugally, and/or accurately than more complex methods" (Gigerenzer and Gaissmaier 2011, p. 454). The use of heuristics to make decisions challenges the prescription to use all the available information in statistical discrimination theory. In contrast with the view that predictive accuracy increases as more information is considered in more complex ways (Payne et al. 1993), heuristics oftentimes achieve higher predictive accuracy than more complex models such as linear regressions and decision trees. The increase in accuracy while using less information, computation, and time is known as the "less-is-more" effect (Gigerenzer and Brighton 2009). For example, the take-the-best heuristic uses only the most valid cue (i.e., piece of information) to discriminate between two alternatives and ignores all the other cues to make decisions. Another example is the minimalist heuristic, which uses a randomly selected cue to assess two alternatives. For some tasks (e.g., predicting which of two German cities has a larger population), both the take-the-best heuristic and the minimalist heuristic can match or outperform the accuracy and speed of models that use more information (Gigerenzer and Goldstein 1996, 1999).

The less-is-more effect indicates that "ignoring large amounts of the available information can pay off" in many situations (Brighton and Gigerenzer 2012, p. 35). For instance, when the environment is characterized by high uncertainty, simple heuristics tend to make better predictions than complex models using many parameters. To understand why, imagine that two models are learning a true, underlying function from noisy data generated by the function. The complex model is allowed to closely fit the noisy data, and the simple model (i.e., the heuristic) is restricted from doing so. When the models are used to make predictions, the complex model incurs low bias and high variance; that is, predictions are close to the true values on average, but individual predictions may be highly inaccurate because the model fits the noise. The simple model, in contrast, incurs high bias and low variance—predictions are far from the true values on average, but individual predictions are less erratic and exhibit relatively stable error. As a consequence, heuristics perform better than complex models when there is “greater uncertainty” and a “less stable environment” (Artinger et al. 2015, p. S38) because “uncertainty is often best dealt with by ignoring information, and being biased” (Brighton and Gigerenzer 2012, p. 60). Because of the high variance in uncertain environments, complex models with many adjustable parameters are likely to overfit the data and result in greater prediction error (for further explanation of how underspecified models can outperform truer models, see Csaszar and Ostler 2020, Shmueli 2010, and Wu et al. 2007).

Heuristics can also outperform complex models if the decision maker is inconsistent. As Kahneman et al. (2016, p. 40) explain: “Organizations expect consistency from [decision makers]: Identical cases should be treated similarly, if not identically.” Inconsistency manifests as disagreements within a single evaluator’s assessments (Brehmer 1981, 1994). Kahneman et al. (2021, pp. 200–209) refer to inconsistency as “pattern noise” and explain that inconsistency has many sources, including personality and values affecting reactions to a particular case, incompetence and inexperience for a certain type of decision, and extraneous and irrelevant factors like the weather, time of day, distractions, hunger, or moodiness at the moment of evaluation. This noise can lead to paying attention to the wrong information (Gigerenzer and Goldstein 1996, Goldstein and Gigerenzer 2002) or lead to a “trembling hand” when applying weights. Several empirical studies have documented the presence of inconsistency (Brehmer et al. 1980; see also Hammond and Summers 1972 and Mitchell et al. 2011).

Compared to heuristics, complex models offer more opportunities for decision makers to make mistakes, require more skill to use, and are more difficult to properly execute. For example, Hogarth and Karelaia (2007) compare the accuracy of predictions made by an inconsistent decision maker using a linear model versus different heuristics. They find that the ability of the decision maker to use information consistently must be “quite high” before the more complex linear model can compete with simpler heuristics: “unless [decision maker consistency] is high, decision makers are better off using simple heuristics, provided that they are able to implement these correctly” (Hogarth and Karelaia 2007, p. 741).

Hence, research on heuristics suggests that statistical discrimination's logic of always using more information to make predictions is not always optimal. In conditions of environmental uncertainty or decision maker inconsistency, complex models can underperform heuristics. Thus, incorporating these variables into statistical discrimination may show that statistical discrimination's logic—using all information about a candidate, including group characteristics that are correlated with but do not cause productivity—may be misleading. The goal of our work is to point out statistical discrimination's boundary conditions: the specific situations under which using less information can lead to better predictions about individual productivity.

2.3 Contextualizing Discrimination

To contextualize and parameterize discrimination, we combine the key insights from statistical discrimination and heuristics research discussed above. Specifically, statistical discrimination theory emphasizes two key contingencies: (i) unobservability of a candidate's true productivity and (ii) correlation between observable group characteristics and productivity in the population. *Unobservability* describes the degree to which the decision maker can see the candidate's true productivity or ability, which can be obscured by noise in the productivity signal. *Correlation* describes the degree to which a group characteristic (e.g., skin tone or gender) is correlated with ability in a population. When unobservability and correlation are high (as assumed in statistical discrimination theory), then it is rational for employers to base their predictions on all available information about the candidates, including group characteristics.

Research on heuristics suggests two contingencies that add realism to a model of statistical discrimination: (iii) environmental uncertainty and (iv) decision maker inconsistency. *Uncertainty* describes the unpredictability of the circumstances in which predicted outcomes occur. *Inconsistency* describes the non-optimal use of information. That is, the decision maker knows how to use the information optimally, but the final prediction does not purely reflect this knowledge. When uncertainty and inconsistency are high, heuristics' less-is-more effect proposes that using less information (i.e., ignoring information about a candidate's group characteristics) can produce better predictions than using more information.

In addition to the four contingencies mentioned so far, which stem from classic research on statistical discrimination and heuristics, we also incorporate *diagnostic bias*, a contingency that has been highlighted by recent empirical research on discrimination (Bordalo et al. 2019, Correll and Benard 2006).[5] Diagnostic bias describes the degree to which the true ability of individuals in one group is systematically overestimated and the true ability of individuals in another group is systematically underestimated. Evaluators might introduce diagnostic bias in the evaluation process because they rely on stereotypes that affect how well they can "see" an individual's true productivity. For example, even when women perform as well as men on mathematics tasks, their ability is consistently underestimated (Reuben et al. 2014). Correll and Benard (2006, p. 92) explain that because good performances are inconsistent with expectations about the discriminated actors, their performances face increased scrutiny and stricter assessment, and they are "less likely to be judged as demonstrating task ability or competence." Diagnostic bias may also emerge because evaluation systems might themselves be biased (Borjas and Goldberg 1978). For example, questions in standardized tests may be phrased in terms that are hard to understand by a disadvantaged group simply due to lack of familiarity with a cultural or linguistic norm (Freedle 2003, Santelices and Wilson 2010).

We next describe how the contingencies discussed so far—unobservability, correlation, uncertainty, inconsistency, and diagnostic bias—are captured by our behavioral model of how discrimination affects predictive accuracy.

3 Model

Our model of discrimination focuses on the task of predicting the future performance of a candidate based on two cues or characteristics of the candidate: a discriminatory cue and a causal cue. These cues are

[5] We would like to thank the editor and reviewers for their suggesting the addition of this contingency.

correlated with each other, but they differ because the discriminatory cue is perfectly observable yet has no causal effect on performance, while the causal cue cannot be perfectly observed but causes performance.[6] We study the four decision rules corresponding to every possible subset of the two cues; that is, we study rules that use: (i) both cues, (ii) just the discriminatory cue, (iii) just the non-discriminatory cue, and (iv) neither of the cues. We respectively call these rules Discrimination (D), Extreme Discrimination (ED), Non-Discrimination (ND), and Zero Cues (ZC).

Each of these rules makes optimal use of the information that is available to it, meaning that they use the available cue(s) in a way that minimizes the expected discrepancy between the predicted and the real performance of the candidates under assessment.

To understand the basic elements of our model, consider the example of a hiring manager who is assessing a job candidate for a computer programming position. She has two pieces of information about the job candidate from which she must infer his future performance: skin tone (the discriminatory cue) and programming ability (the causal cue). Skin tone is perfectly visible but *does not* affect performance; programming ability is to some degree unobservable (because it is measured through an imperfect instrument, such as a portfolio of past projects) but *does* affect performance. Skin tone and programming ability may be correlated, however. For example, Monk (2014) discusses how skin tone affects the life chances of black Americans, indicating that individuals with darker skin have fewer opportunities to develop skills (such as programming) in school, informally, or in a job. Since programming ability is partly unobservable, the manager can more accurately predict performance by taking skin tone into account—even though skin tone itself has no causal effect on performance. However, taking skin tone

[6] In our model, what makes a cue "discriminatory" is its use by employers to discriminate. Note that discrimination in labor markets is conventionally defined as "the presence of different [treatment] for workers of the same ability" (Aigner and Cain 1977, p. 175) and statistical discrimination is defined in terms of employers evaluating workers "on the basis of easily observable variables that are correlated with productivity" (Altonji and Pierret 2001, p. 314). We call this correlated but non-causal variable the "discriminatory cue." Thus, our use of the term is consistent with how discrimination has been conceptualized by previous literatures. However, note that there could be cases where discrimination is free from negative moral and socio-political connotations, such as when medical recommendations are made based on prevalent conditions affecting a patient's race.

into account could result in discrimination (i.e., the unequal treatment of equally productive individuals), by lowering the likelihood of hiring equally productive individuals with darker skin tone compared to those with lighter skin tones. The four decision rules we model correspond to deciding who to hire based on: (i) skin tone and programming ability (i.e., Discrimination rule), (ii) solely on skin tone (i.e., Extreme Discrimination rule), (iii) solely on programming ability (i.e., Non-Discrimination rule), and (iv) not paying attention to any cues, just the population average (i.e., Zero Cues rule).

As mentioned before, contextualizing discrimination calls for accounting not just for the role of unobservability and correlation but also for the role of uncertainty and inconsistency (and diagnostic bias, which we incorporate later for the sake of clarity). Continuing with the example, the predictive accuracy of the hiring manager is influenced by uncertainty in the environment, which may randomly affect the future performance of the hired candidate. For example, new projects may require skills that were not known or evaluated at the time of hiring. The predictive accuracy of the hiring manager is also affected by her inconsistency; for example, she may place too much value on the unobservable cue of programming ability because one of the job candidate's past projects resembles a project she had worked on. Or the hiring manager may be distracted by interruptions or fatigue during some interviews. These four aspects of the decision-making process—unobservability, correlation, uncertainty, inconsistency—affect the predictive accuracy of the hiring manager's final assessment.

The workings of our model can be efficiently described using Brunswik's (1952) lens model framework (for a survey of studies using this model, see Karelaia and Hogarth 2008), which describes prediction tasks by modeling both the environment and the decision maker (see Figure 1). The left-hand side of Figure 1 describes how the actual performance of the individual (y) depends on two cues (x_D as the discriminatory cue and x_C as the causal cue) and a random factor ($\tilde{\mu}$). The dotted line connecting x_D and x_C depicts the *correlation* between these two cues (denoted ρ). The variability of the random factor (denoted σ_μ) determines the *uncertainty* of the environment. This means that as σ_μ increases, the relationship between the cues and performance becomes less deterministic. The β's on the left-hand arrows express the weight each factor has on performance (i.e., $y = 0x_D + 1x_C + 1\tilde{\mu} = x_C + \tilde{\mu}$).[7]

Insert Figure 1 about here

Note that the fact that β_D is set to zero implies that x_D, the discriminatory cue, does not have a causal effect on y. This encapsulates the idea of discrimination: that equally productive individuals are treated differently based on characteristics that do not affect (i.e., cause) their job performance. More precisely, $\beta_D = 0$ implies that if one could manipulate x_D, y would not change—which is precisely how lack of causation is formally defined (Pearl 2009, p. 417). The model is also consistent with the statistical hallmark of lack of causation: that, given x_C, y and x_D are conditionally independent (known as the "screening off" criterion; Reichenbach 1956, p. 189).

The right-hand side of Figure 1 describes how the decision maker predicts performance; that is, how predicted performance (denoted $\hat{y}$) depends on the discriminatory cue (x_D) and on a noisy measurement of the causal cue (denoted x_C' and defined as $x_C + \tilde{\varepsilon}$). The variability of the measurement error (denoted σ_ε) determines the unobservability of the causal cue, which is the degree to which the causal cue is unobservable. The $\hat{\beta}$'s on the right-hand arrows express how the decision maker weighs the observed cues. Under the Discrimination rule, the two cues are weighted optimally: $\hat{\beta}_D$ and $\hat{\beta}_C$ are chosen so as to minimize the expected prediction error, which measures the discrepancy between the predicted and actual performance.[8] Under the Extreme Discrimination rule, only the first cue is weighted optimally, and the second cue cannot be used. Under the Non-Discrimination rule, the first cue cannot be used, so only the second cue is weighted optimally. Lastly, under the Zero Cues rule, neither the first nor the

[7] It is important to note that the fact that $\beta_D = 0$ does not imply that corr(x_D, y) = 0. In fact, if $\rho > 0$, x_D and y will also be positively correlated. For instance, a high x_D is likely to be accompanied by a high x_C and, hence, by a high y (as y is a function of x_C). For this reason, our model is fundamentally different from the model used by Conger and Jackson (1972) to describe "suppressor effects."

[8] To compute the optimal weights, the model assumes that the decision maker knows the probability distributions of the cues (x_D and x_C), the uncertain factor ($\tilde{\mu}$), and the measurement error ($\tilde{\varepsilon}$). The knowledge of these distributions could stem from learning (experientially or vicariously) from a large number of previous cases.

second cue are used.

The four rules so far described make optimal use of the information they observe and, thus, assume perfect rationality. Although perfect rationality is useful as a benchmark, it may not be behaviorally plausible (March and Simon 1958) because "humans are unreliable decision makers" (Kahneman et al. 2016, p. 40). As mentioned previously in our contextualization of discrimination, to account for a more behaviorally plausible use of the four rules, our model includes inconsistency, which captures different degrees of deviation from optimal behavior. We model inconsistency as random noise that is added to the optimal coefficients for the cues used by different decision rules.[9]

In sum, our model allows us to explore the behavior of Discrimination, Extreme Discrimination, Non-Discrimination, and Zero Cues as a function of four main contingencies: (i) the degree to which the causal cue is unobservable (σ_ε); (ii) the correlation between the discriminatory and causal cues (ρ); (iii) the uncertainty of the environment (σ_μ); and (iv) the degree of inconsistency with which the optimal weights are used (denoted σ_ι). We later add diagnostic bias as a fifth contingency. The rest of this section describes the elements of the model in more detail.

3.1 The Environment

Consistent with the Brunswikian tradition, we model the environment (i.e., the left-hand side of Figure 1) as a data generating process. The environment consists of prospects (i.e., candidates being judged), each characterized by a measure of performance (y) and two cues (x_D, which is the discriminatory cue, and x_C, which is the causal cue). Prospects are drawn from a population whose cues are normally distributed (with mean 0 and standard deviation 1) with cross-cue correlation ρ. That is,

[9] This modeling of inconsistency is consistent with the Brunswikian approach and draws on prior research on decision makers' use of non-optimal weights. For example, Dawes (1979) describes many situations in which decision makers do not use optimal weights. Hogarth and Kareleia (2007) model "linear cognitive ability" (i.e., a decision makers' ability to utilize a linear judgment model) which partly consists of the degree to which the decision maker's weights are (non-)optimal (Hogarth and Karelaia 2007, pp. 737–738).

$$\begin{pmatrix} x_D \\ x_C \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right) \tag{1}$$

As previously mentioned, the prospect's performance (y) only depends on its causal cue (x_C) and on some random environmental factor ($\tilde{\mu}$). We allow for a tunable degree of uncertainty by drawing $\tilde{\mu}$ from a Normal distribution with mean 0 and standard deviation σ_μ. More formally,

$$y = x_C + \tilde{\mu}, \text{where } \tilde{\mu} \sim N(0, \sigma_\mu^2). \tag{2}$$

We model unobservability as how inaccurate the measurement of the causal cue is. The decision maker does not observe x_C but a noisy version of it, which we call x'_C. Mathematically,

$$x'_C = x_C + \tilde{\varepsilon}, \text{where } \tilde{\varepsilon} \sim N(0, \sigma_\varepsilon^2). \tag{3}$$

As σ_ε increases, the decision maker's perception of the causal cue (x'_C) becomes a less accurate reflection of its true value (x_C). Unobservability can be interpreted as the decision maker's inability to properly "read" the causal cue.

We now introduce the contingency of diagnostic bias. Note that how unobservable the causal cue is may depend on the value of the discriminatory cue. For example, due to the cultural bias embedded in exam questions, standardized test scores may be less reflective of ability for members of a discriminated group (Freedle 2003). Our model allows for this diagnostic bias by introducing a correlation between $\tilde{\varepsilon}$ and x_D, which we call ρ_2. As this correlation increases, those with high values in x_D become also likely to have a high value in $\tilde{\varepsilon}$ and, hence, be perceived as having a higher ability than they really have (in Freedle's (2003) example, high x_D would correspond to white candidates).

3.2 Discrimination Rule

Under the Discrimination rule, the decision maker uses both the discriminatory cue (x_D) and the noisy observation of the causal cue (x'_C). To determine the optimal way of combining these two cues in a given environment (which is characterized by ρ, ρ_2, σ_μ, and σ_ε), we find the function that minimizes the expected discrepancy between y and $\hat{y}$. For simplicity and analytical tractability, we measure discrepancy as square loss ($(y-\hat{y})^2$). Minimizing this expected loss over the distribution of prospects given in Equation (1) leads to the optimal discrimination rule (see the Appendix for the derivation). For clarity of the exposition, we first present this optimal rule assuming that $\rho_2 = 0$ (i.e., no diagnostic bias). In such a case, the optimal discrimination rule is

$$\hat{y}_{\mathrm{D}} = \frac{\rho\sigma_\varepsilon^2}{1-\rho^2+\sigma_\varepsilon^2} x_D + \frac{1-\rho^2}{1-\rho^2+\sigma_\varepsilon^2} x'_C. \tag{4}$$

We analyze this rule in detail in the Results section, but to start developing an intuition for it, note that when the correlation between the causal and discriminatory cues is perfect ($\rho = 1$), the second coefficient becomes zero. Hence, the rule relies solely on the discriminatory cue (x_D). This makes sense, as perfect correlation implies that x_D (the discriminatory cue) is a noiseless signal of x_C (the causal cue) while x'_C only imperfectly reflects x_C. Another observation from this rule is that the optimal weights are not a function of uncertainty (σ_μ). This is so because the effect of uncertainty on y is completely random (and hence cannot be better predicted by making any changes to the decision rule).

When diagnostic bias is different from zero, the optimal discrimination rule becomes:

$$\hat{y}_{\mathrm{D}} = \frac{\rho\sigma_\varepsilon^2 - \rho_2\sigma_\varepsilon}{1-\rho^2+\sigma_\varepsilon^2-\rho_2\sigma_\varepsilon(2\rho+\rho_2\sigma_\varepsilon)} x_D + \frac{1-\rho^2-\rho\rho_2\sigma_\varepsilon}{1-\rho^2+\sigma_\varepsilon^2-\rho_2\sigma_\varepsilon(2\rho+\rho_2\sigma_\varepsilon)} x'_C. \tag{5}$$

Note that this rule can be read as a weighted average between x_D and x'_C. That is, $\hat{y}_{\mathrm{D}}$ has the form $\hat{\beta}_D^* x_D + \hat{\beta}_C^* x'_C$ with each coefficient being a non-negative number. This suggests a simple way of characterizing the behavior of the discrimination rule: measuring the fraction of the cue weight that is given to x_D. More formally,

$$\text{Optimal reliance on the discriminatory cue} = \frac{\hat{\beta}_D^*}{\hat{\beta}_D^* + \hat{\beta}_C^*} = \frac{\rho\sigma_\varepsilon^2 - \rho_2\sigma_\varepsilon}{1-\rho^2+\rho\sigma_\varepsilon^2-\rho_2\sigma_\varepsilon(1+\rho)}. \tag{6}$$

Using this formula one can say, for instance, that when $\rho = 0.75$, $\sigma_\varepsilon = 0.5$, and $\rho_2 = 0$, the optimal discrimination rule relies 30 percent ($= \frac{0.75 \times 0.5^2}{1-0.75^2+0.75 \times 0.5^2}$) on the discriminatory cue (i.e., puts 30 percent of the cue weight on x_D and 70 percent on x'_C).

3.3 Extreme Discrimination Rule

Under Extreme Discrimination, the decision maker only uses the non-causal discriminatory cue (x_D) to predict the prospect's performance.[10] To derive this rule, we solve a minimization problem analogous to the one described for the discrimination rule (see Appendix for the derivation), which results in the following rule:

$$\hat{y}_{\text{ED}} = \rho x_D. \tag{7}$$

Hence, under the Extreme Discrimination rule, if correlation is perfect ($\rho = 1$), the best prediction of performance is x_D. This is because the discriminatory cue is perfectly correlated with the causal cue and, therefore, perfectly substitutable for the causal cue (x_C) that this rule cannot use directly. As correlation decreases, it becomes less useful to rely on x_D, and the predicted outcome moves closer to zero. The rationale is that as x_D becomes less informative, it is more accurate to predict that performance will be closer to the population's average (which is 0, as each distribution that feeds into y is zero-centered; see Equations 1 and 2).

3.4 Non-Discrimination Rule

Under Non-Discrimination, the decision maker only uses the noisy observation of the causal cue (x'_C) to predict the prospect's performance. Solving for the optimal non-discrimination rule results in

$$\hat{y}_{\text{ND}} = \frac{1}{1 + \sigma_\varepsilon^2} x'_C. \tag{8}$$

Note that under this rule, if observability is perfect ($\sigma_\varepsilon = 0$), the best prediction of performance is x'_C; from then on, as unobservability increases, the prediction becomes progressively closer to zero, the population's average performance. The logic is akin to the one described for Extreme Discrimination: as the available cue (in this case x'_C) becomes less informative, the rule progressively switches to predicting values that are closer to the population average.

[10] We would like to thank the editor and reviewers for suggesting we examine Extreme Discrimination and Zero Cues.

3.5 Zero Cues Rule

Under this rule, the decision maker does not use any cues to predict the prospect's performance. Only the population mean is used to make a prediction. Given that the population mean for y is 0, this rule simply corresponds to

$$\hat{y}_{\mathrm{ZC}} = 0. \tag{9}$$

3.6 Inconsistent Application of Rules

Recall from our discussion of contextualizing discrimination that the perfect rationality implicit in the rules described so far may not be realistic. A real-world decision maker may have an intuition about what weights to put on the different cues, but he or she will most likely use weights that deviate from the optimal ones. We model this deviation as independent and identically distributed (i.i.d.) random noise that is added to the optimal weights.[11] These more realistic, noisy versions of the rules—which we call Inconsistent Discrimination (ID), Inconsistent Extreme Discrimination (IED), Inconsistent Non-Discrimination (IND), and Inconsistent Zero Cues (IZC)—are defined as follows:

$$\hat{y}_{\mathrm{ID}} = \left(\hat{\beta}_D^* + \tilde{\iota}_D\right)x_D + \left(\hat{\beta}_C^* + \tilde{\iota}_C\right)x_C' \tag{10}$$

$$\hat{y}_{\mathrm{IED}} = \left(\hat{\beta}_D^{**} + \tilde{\iota}_D\right)x_D \tag{11}$$

$$\hat{y}_{\mathrm{IND}} = \left(\hat{\beta}_C^{***} + \tilde{\iota}_C\right)x_C' \tag{12}$$

$$\hat{y}_{\mathrm{IZC}} = 0 \tag{13}$$

The noise terms ($\tilde{\iota}$'s) are independent draws from a $N(0, \sigma_\iota^2)$, and the starred $\hat{\beta}$'s are the optimal weights derived above for each of the rules. The standard deviation of the noise terms (σ_ι) corresponds to the degree of inconsistency with which the decision maker uses the rules. Note that as σ_ι approaches zero, the inconsistent rules converge to the optimal rules. Note also that Zero Cues (Equation 9) and

[11] Because the noise is zero-centered, inconsistency does not introduce bias in the decision process. In our model, bias is an emergent property (e.g., whenever to minimize prediction errors, Discrimination ends up using a nonzero $\hat{\beta}_D$). It is also worthwhile to note that one could model inconsistency as multiplicative noise (rather than additive); doing so produces results that are qualitatively similar to the ones we present.

Inconsistent Zero Cues (Equation 13) are identical, as these rules do not depend on coefficients that can be used inconsistently.

How we model inconsistency—as putting too much or too little weight on the different characteristics of a problem—is aligned with the understanding of inconsistency as stemming from imperfectly applying a decision rule (Brehmer 1994, pp. 143–144) and with empirical evidence showing that inconsistency increases with the number of cues in the decision rule (Brehmer et al. 1980, Einhorn 1971). However, because not much research has empirically examined inconsistency (Kahneman et al. 2021, pp. 200–209), future empirical research could refine our distributional assumptions.

3.7 Measuring the Expected Errors of the Rules

To compare how well the different decision rules (equations 10–13) perform at making predictions under different situations, we compute the mean square error (MSE) of each rule. The appendix describes how to derive these errors in closed-form. Table 1 summarizes the prediction formula ($\hat{y}$) and error formula (MSE) for each rule.

Insert Table 1 about here

Note that discrimination is the only rule whose formulas depend on diagnostic bias (ρ_2; the correlation between $\tilde{\varepsilon}$ and x_D). Because of this and because the formulas for this rule are somewhat unwieldy, Table 1 presents the formulas for the Discrimination rule with and without $\rho_2 = 0$.

It is worth noting that all the error formulas in Table 1 are equally affected by environmental uncertainty: all the MSEs only include σ_μ^2 as an additive term. Because environmental uncertainty affects all rules equally, we will ignore this parameter when comparing the relative performance of the rules.

To summarize, our model has defined optimal and inconsistent versions of the Discrimination, Extreme Discrimination, Non-Discrimination, and Zero Cues rules, as well as measures to compare their predictive errors (MSE). In the next section we will study how the rule's predictive errors depend on the model's contingencies.

4 Results

In this section we analyze the model using graphical plots that intuitively and precisely convey the main results. Each of these plots shows how a dependent variable (plotted as a color) depends on correlation ρ (on the x-axis) and unobservability σ_ε (on the y-axis) for fixed values of inconsistency σ_ι and diagnostic bias ρ_2. We chose the parameter values used to generate these plots after exploring the model exhaustively and verifying that these parameter values qualitatively illustrate the behavior of the model under any plausible set of parameter values. The parameter values used correspond to varying correlation (ρ) and unobservability (σ_ε) continuously from 0 to 1 and having inconsistency (σ_ι) and diagnostic bias (ρ_2) and take specific values between 0 and 1 ($\sigma_\iota =$ 0, 0.1, 0.3, 0.7 and $\rho_2 =$ 0, 0.25, 0.5). Understanding these plots is, in other words, sufficient to understanding the whole behavior of the model. We first examine a baseline case in which there is no inconsistency and no diagnostic bias to understand the basic set of mechanisms that are driving our model of discrimination before exploring how the rest of the model parameters affect predictive accuracy.

4.1 Baseline Case

To start developing intuitions about the model, we begin by studying the Discrimination rule under the baseline case of no inconsistency ($\sigma_\iota = 0$) and no diagnostic bias ($\rho_2 = 0$), which matches the assumptions of the statistical discrimination literature. In this case, the optimal reliance on the discriminatory cue increases with correlation and unobservability (see Figure 2, where the optimal reliance on the discriminatory cue increases as one moves from the bottom-left to the top-right corner). To understand this behavior, we compare the extreme cases. In the bottom-left corner of Figure 2, there is no correlation and the causal cue is completely observable. In this situation, the decision maker can perfectly predict performance by only using the causal cue; hence, the optimal weight to put on the discriminatory cue is zero. In the top-right corner of Figure 2, however, the causal cue is completely unobservable, and it is highly correlated with the discriminatory cue. In this case, the rational decision maker should rely 100 percent on the discriminatory cue. The discriminatory cue would essentially substitute for the highly correlated, unobservable causal cue to achieve an optimal prediction. As one moves from the bottom-left corner to the top-right corner of Figure 2, the optimal reliance on the discriminatory cue increases.

Insert Figure 2 about here

The previous observations are consistent with Phelps' (1972) model of statistical discrimination. Employers in Phelps' model rely more on the average productivity of a worker's social group (i.e., race or gender) when determining the wage for a worker whose signal of productivity is noisy. Similarly, decision makers in our model rely more on the discriminatory cue when the causal cue is unobservable.

Interestingly, for most situations represented by Figure 2, we can see that the optimal reliance on the discriminatory cue is less than 20 percent (the lower-left area of Figure 2 falls below the 20 percent contour line). Cases in which the optimal reliance is higher than 20 percent are the minority. In other words, putting fairness considerations aside, most circumstances depicted in the baseline case only call for a relatively small reliance on the discriminatory cue.

Now that we understand how the optimal reliance on the discriminatory cue changes with correlation and unobservability, we move on to compare the performance of Discrimination versus Non-Discrimination. We compare these rules by looking at the relative difference between their MSEs. Using the MSE of Discrimination as the reference point, this relative difference can be interpreted as the cost of using the Non-Discrimination rule rather than the Discrimination rule; namely,

$$\text{Cost of not discriminating } = \frac{\mathrm{MSE_{ND}} - \mathrm{MSE_D}}{\mathrm{MSE_D}} = \frac{\mathrm{MSE_{ND}}}{\mathrm{MSE_D}} - 1. \quad (14)$$

A positive cost means that in a given situation, Non-Discrimination yields a higher error than Discrimination. Conversely, a negative cost means that Non-Discrimination produces smaller errors than Discrimination and, hence, there is a benefit to not discriminating.

Figure 3 shows that the cost of not discriminating as a function of correlation and unobservability increases under the same situation described in Figure 2 ($\sigma_\iota = 0$ and $\rho_2 = 0$). The pattern is similar to the one observed in Figure 2. As one moves right along the x-axis (correlation ρ) and up along the y-axis (unobservability σ_ε) in Figure 3, one reaches the maximum cost of not discriminating. The mechanism driving the cost of not discriminating is the following: as unobservability of the causal cue and correlation between the causal and discriminatory cue increase, it becomes more beneficial to rely on the discriminatory cue (and, therefore, more costly to not use it) when predicting performance.

Insert Figure 3 about here

In the vast majority of the cases depicted in Figure 3, the cost of not discriminating leads to an increase in expected error that is below 25 percent.[12] Moreover, in roughly half the cases in this figure, the cost of not discriminating is less than three percent. While models of statistical discrimination assume that it is costly for employers not to discriminate, our model shows that the cost is often minimal (it is important to note, however, that this is a relative cost and that the material significance of the cost's magnitude depends on real-world context). In about half of the cases depicted in Figure 3, all that an employer loses from not discriminating is five percent in error. There are, however, situations in which it is relatively very costly for decision makers to not discriminate. These are the situations that are near the top-right corner of Figure 3, where both unobservability and correlation are high.

4.2 Expanding the Analysis to Incorporate More Rules and Contingencies

So far, our examination has been limited to comparing two rules (Discrimination vs. Non-Discrimination) in the specific case of no inconsistency ($\sigma_\iota = 0$) and no diagnostic bias ($\rho_2 = 0$). We now expand our

[12] Recall that the cost of not discriminating (Equation 14) depends on the ratio of the *MSE of Non-Discrimination* over the *MSE of Discrimination*. Whether this is a large or a small error will depend on specifics of the situation (e.g., the competition). In some situations, a 25% difference in predictive accuracy is critical, but in others, it may not be. Though outside of the scope of the present research, it would be valuable to empirically explore the real-world significance of these costs in addition to the relative cost.

analysis to include the two additional rules (Extreme Discrimination and Zero Cues) as well as different levels of inconsistency and diagnostic bias.

Figure 4 will guide our analysis of this larger space of rules and parameters. This figure shows the rule that achieves the lowest MSE in each region of the parameter space. As before, the *x*- and *y*-axes in each panel correspond, respectively, to correlation (ρ) and unobservability (σ_ε). The columns of panels represent increasing levels of inconsistency (σ_ι) and the rows of panels represent increasing levels of diagnostic bias (ρ_2).[13]

Insert Figure 4 about here

Panel (a) in Figure 4 describes the same situation as Figures 2 and 3 (i.e., $\sigma_\iota = 0$ and $\rho_2 = 0$). In this panel, Discrimination is always preferable over the other rules. This is consistent with Figure 3 (which showed that not discriminating always had a cost, albeit oftentimes small). It is also reasonable that the other rules (Extreme Discrimination and Zero Cues) trail Discrimination, as the absence of inconsistency in this panel allows Discrimination to benefit from using more information. Overall, panel (a) in Figure 4 is consistent with the literature on statistical discrimination, as this panel shows that using the discriminatory cue in addition to the causal cue increases predictive accuracy.

Strikingly, the fundamental result of the statistical discrimination literature is not generally true as one moves away from panel (a): in all other panels in Figure 4, there are plenty of regions where Discrimination is not the preferable rule. In fact, the majority of the area in the other panels of Figure 4 is dominated by Non-Discrimination (shaded in green). The contingency that explains most of the variation across the panels in Figure 4 is inconsistency (σ_ι), thus, we now concentrate on understanding what

[13] A careful observer may note that the plots in the second and third rows of panels end before reaching $\rho = 1$. This is not a plotting flaw but a consequence of the distributional assumptions of the model. For instance, if the correlation between *a* and *b* is very high and the correlation between *b* and *c* is also very high, the correlation between *a* and *c* cannot be any arbitrary number, it must be high too. When using a multivariate normal distribution like we do, all such consistency relationships are captured by the property that the covariance matrix must be positive definite. For our model, this property implies that $\rho < \sqrt{1 - \rho_2^2}$. Hence, that is the maximal value plotted in the panels' *x*-axes.

happens as one increases inconsistency starting from panel (a) and moves to the right into panels (b)–(d). When considering the effects of inconsistency, recall that the inconsistent decision maker introduces error with each cue that is utilized, as random noise is added to the optimal weight of each cue (Equations 10–13).

A first observation from comparing panel (a) to panels (b)–(d) is that Extreme Discrimination becomes the preferable rule when correlation and unobservability are high (i.e., in the top-right corner of panels (b)–(d)). This is because in this region, the optimal reliance on the discriminatory cue is very high (see Figure 2), which is the same as saying that in this region the optimal weight on the *causal* cue is very small. Thus, if the decision maker is inconsistent, then using the causal cue comes at the expense of introducing error through the random noise that is added to the optimal weight. Since the optimal reliance on the causal cue is small in this region, it is better for the decision maker to not use the causal cue at all to avoid bringing in the random noise, which is what the extreme discriminator does. This logic also explains why the area covered by Extreme Discrimination increases with inconsistency (σ_ι); that is, as the random noise increases, any benefit from using the causal cue becomes progressively dwarfed by the random noise it introduces and, hence, the region dominated by Extreme Discrimination grows.

A second observation is that as inconsistency increases, Non-Discrimination becomes the preferable rule when correlation and unobservability are low or medium (i.e., in the area that is closer to the bottom-left corner of panels (b)–(d)). The explanation is analogous to the one for Extreme Discrimination: in this region, the optimal reliance on the discriminatory cue is relatively small (see the bottom-left corner of Figure 2) and, hence, as inconsistency increases, it becomes better for the decision maker to ignore the discriminatory cue to avoid bringing in random noise—that is, Non-Discrimination becomes preferable.

A third observation is that when inconsistency is very high (in panel (d)), there is a region where Zero Cues dominates (i.e., near the top-left corner of panel (d)). This is because a sufficiently large level of inconsistency can make all the rules that are sensitive to it less predictive than simply predicting the population average. Zero Cues shows up on the top-left corner of panel (d), as that is where predictions are the noisiest (as there the causal cue is hard to observe and the discriminatory cue adds little predictive power). If one were to increase inconsistency further, Zero Cues would eventually dominate all other rules regardless of the value of the other contingencies.

An overall observation from panels (a)–(d) is that the area where Discrimination is preferable becomes smaller and smaller as inconsistency increases. Discrimination dominates all of panel (a); yet, by panel (d), it has completely vanished. This overall observation is significant, as it qualifies the logic of statistical discrimination—that as long as correlation and unobservability are greater than zero, using the discriminatory cue can only improve predictive accuracy. In contrast, the previous analyses show that the logic of statistical discrimination only applies in a small region of the parameter space (panel (a)). What happens is that, although Discrimination uses more information, this rule is more sensitive to inconsistency than the other rules (as the discriminator has two ways of being inconsistent while the other rules have either one or zero sources of inconsistency). Thus, which rule is better at making predictions depends on the trade-off between the costs and benefits—the noise and the signal—introduced by each cue. And those depend in nuanced ways on the contingencies we study.

The remaining rows of Figure 4 follow a similar pattern as panels (a)–(d). This means that introducing diagnostic bias (i.e., using $\rho_2 > 0$) does not generally change which rule is preferable. The main difference between the three rows of panels is in the second column: as one moves from panels (b) to (f) to (j), Discrimination begins to dominate part of the middle ground of panels (f) and (j). What happens is that when the diagnostic bias increases (e.g., one moves from panel (b) to (f)), Discrimination is able to use the discriminatory cue to counteract the effect of the biased causal cue it perceives (x_C'). This counteracting effect ceases to be significant when inconsistency is larger (i.e., in the panels to the right of (f) and (j)), as Discrimination, which depends on two cues, becomes more inaccurate than the other simpler rules.

4.3 How Costly Is It to Use Non-Discrimination?

In addition to identifying the optimal rule, we can also analyze the cost of not discriminating vis-à-vis the other rules. Figure 5 follows the same layout as Figure 4 but instead of plotting the preferred rule, it plots the cost of not discriminating.

Insert Figure 5 about here

Because unlike Figure 3 which only compared two rules, we now compare four rules, instead of comparing Non-Discrimination vis-à-vis Discrimination we now compare Non-Discrimination vis-à-vis the best of the other rules. That is,

$$\text{Cost of not discriminating}' = \frac{\mathrm{MSE}_{\mathrm{ND}}}{\min\left(\mathrm{MSE}_{\mathrm{D}}, \mathrm{MSE}_{\mathrm{ED}}, \mathrm{MSE}_{\mathrm{ZC}}\right)} - 1. \quad (15)$$

Panel (a) in Figure 5 is identical to Figure 3. One can see that as inconsistency increases (i.e., as one moves to the right from panel (a)), the cost of not discriminating becomes mostly negative. In these situations, Non-Discrimination is the preferred rule (see Figure 4). Moreover, even when the cost of not discriminating is positive, it is not very large in most cases described by Figure 5. For example, consider Figure 4(j). About half of the plot shows Discrimination as the preferred rule. However, Figure 5(j) reveals that the costs of not discriminating are relatively low; the increase in error is less than 13 percent in the vast majority of situations. Hence, even when Non-Discrimination is *not* the preferred rule, a decision maker may still consider implementing Non-Discrimination without incurring too great of a cost. In other words, in the great majority of cases depicted in Figure 5, using meritocracy (i.e., just considering the causal cue) is either the best decision or does not carry a very large cost.

5 Discussion

Statistical discrimination, in which a job candidate's non-performance-causing group characteristic is used to proxy for an unobservable causal cue, is often viewed as a "rational" baseline in research on diversity and equity in labor markets (Altonji and Pierret 2001, p. 316, Pager and Karafin 2009, pp. 73–

75, Tilcsik 2021). Statistical discrimination's main message is that decision makers can better predict the performance of a candidate when using more information rather than less information about the candidate. In contrast, research on heuristics argues that "less-is-more": that due to cognitive and environmental constraints, more accurate predictions can be achieved by ignoring some information. We show that this is indeed the case in the context of discrimination, where decision makers are oftentimes better off by ignoring group characteristics when evaluating candidates.

More specifically, our model shows when and why the less-is-more effect holds in the context of statistical discrimination under conditions of increased realism. We compare how using or not using two different pieces of information about a candidate (i.e., a non-causal discriminatory cue like skin tone and a causal cue like programming ability) affects the accuracy of predictions made under five contingencies, the latter three of which extend upon models of statistical discrimination: (i) unobservability of cues, (ii) correlation between cues, (iii) uncertainty in the environment, (iv) inconsistency of the decision maker, and (v) diagnostic bias of the causal cue. Overall, the results show that not discriminating—that is, not utilizing the discriminatory cue in making predictions—achieves the most accurate predictions in many situations. Moreover, even under the very specific conditions in which discrimination provides the better prediction, the gains from discriminating in terms of predictive accuracy are usually not very large, though their material significance depends on real-world context. That is, our work overturns the main result of statistical discrimination by pointing out that when behavioral realism is added to statistical discrimination, there are many situations in which the main prediction of the classic literature on statistical discrimination is overturned.

5.1 Theoretical Contributions

Our model makes three theoretical contributions: demonstrating critical implications of heuristics' less-is-more effect for statistical discrimination, extending statistical discrimination theory, and providing an efficiency argument for abstaining from statistical discrimination.

First, our model demonstrates that the less-is-more effect, a well-known insight from the heuristics literature (Gigerenzer and Brighton 2009), has important implications for statistical discrimination theory. While statistical discrimination advocates for decision makers to “use all the information available to them” (Bertrand and Duflo 2017, p. 311) to make improved predictions about candidates, the heuristics literature proposes that ignoring information can lead to more accurate predictions. Indeed, our model demonstrates that in most situations described by the parameters in our model, *not using* information about a candidate’s group characteristics can lead to better predictions than using such information. Our model also delineates the situations where it is preferable to use *all* available information, *only* information about a candidate’s group characteristic, *only* imperfect information about a candidate’s ability, and *zero* information about the individual. Therefore, our work qualifies previous work on statistical discrimination. While statistical discrimination theory justifies employers using all available information, our model shows that doing so can be more harmful than helpful to predictive accuracy when behavioral realism is taken into account. The inclusion of more behaviorally realistic contingencies can even overturn core predictions made by statistical discrimination.

Second, we extend upon statistical discrimination theory by modeling mechanisms that are more prevalent in disciplines outside of economics. To bridge our account of statistical discrimination with social psychology’s work on stereotypes (Fiske 1998, Hilton and von Hippel 1996, Judd and Park 1993), our model of statistical discrimination includes the contingency of diagnostic bias, which can reflect decision makers’ reliance on stereotypes that would generate an “overreaction to diagnostic information” (Bordalo et al. 2016, p. 1759). To take into account the constraint on decision maker’s rationality, the limited “computational might” described by the heuristics literature (Gigerenzer and Goldstein 1996, p. 651), our model of statistical discrimination incorporates the contingency of inconsistency, which captures the decision maker’s inability to set optimal beliefs. These departures from statistical discrimination theory contribute to a broader conversation about the fallibility of statistical discriminators within economics (e.g., Bohren et al. 2019, Bordalo et al. 2016, 2019, Coffman et al. 2021).

Third, we provide an efficiency argument for abstaining from statistical discrimination. In his essay about how the idea of statistical discrimination has rationalized the use of stereotypes, Tilcsik (2021, p.

100) writes that advocates of statistical discrimination “characterize the practice as fair and morally defensible” because it is profit-maximizing, does not stem from prejudice, and treats people with the same expected productivity equally. Similarly, Bertrand and Duflo (2017, p. 312) write: “While taste-based discrimination is clearly inefficient…, statistical discrimination is theoretically efficient and, hence, more easily defendable in ethical terms under the utilitarian argument.” We demonstrate, however, that these arguments for the practice of statistical discrimination are weakened, if not disproven, when the limitations of human decision makers are taken into consideration.

5.2 Practical Implications

A practical implication of our work is that decision makers should acknowledge that discrimination—apart from being morally abject—only improves predictive accuracy under a very specific set of conditions. Non-discrimination is the optimal rule in most situations defined by our model. Hence, if the exact values of the contingencies faced by an employer are unknown, it may be rational for a decision maker to abstain from discriminating. Moreover, even if the decision maker knows the exact scenario he or she is facing but is inconsistent in applying optimal weights, it may still be rational to abstain from discriminating, since in many situations even the slightest non-optimal use of information can result in discrimination decreasing predictive accuracy. We conjecture that the pervasiveness of discrimination may be partly explained by decision makers being ignorant of their inconsistency and erroneously thinking that they benefit from using information contained in discriminatory cues like a candidate’s race or gender. Such erroneous beliefs may have been reassured by statistical discrimination.

Another practical implication is to suggest ways of reducing discrimination in society. Here, we assume that decision makers tend to use the decision rule that maximizes their predictive accuracy (akin to the profit maximization assumption in economics) and, hence, our suggestions revolve around moving the decision makers toward situations where discrimination is not preferable. One way of doing so is to prevent situations where using the discriminatory cue entails the largest gains in predictive accuracy (i.e., when correlation and unobservability are both high). Governments can make such situations less prevalent by decreasing correlation (e.g., by training the disadvantaged group; Heckman 2006) and decreasing unobservability (e.g., by creating certification programs that are equally accessible by all groups; Harris 2001). Another way of reducing the benefits of discrimination applies to situations where there is no inconsistency and, thus, discrimination maximizes predictive accuracy (see the first column of panels in Figure 4). These situations may become more prevalent because algorithms, which unlike humans do not exhibit inconsistency, are increasingly used to make decisions. Hence, if left unregulated, artificial intelligence could usher in a period of increased discrimination. One way of minimizing the chances of this scenario occurring is by requiring that algorithms be open to inspection. Doing so would increase scrutiny over these algorithms, making discriminatory behavior more easily detectable.

5.3 Limitations and Further Research

Like all models, our work has limitations that could be addressed by future work. First, we used predictive accuracy as our model's dependent variable given that in many situations this is an important determinant of profits. However, in some situations, decision makers may be willing to accept lower predictive accuracy in exchange of other objectives (e.g., equal representation). Thus, examining other dependent variables would make sense. One example is examining candidates' rankings (see, e.g., Denrell and Liu 2012), as predictive accuracy may not change who is shortlisted or hired for a job. One could also examine omission and commission errors, as in some settings, rejecting a good candidate may have a much smaller cost than accepting a bad one. Our model implicitly assigns equal importance to the two types of errors, even though errors of omission and commission may have different costs depending on the situation (Csaszar 2013).

Second, further research could extend our model to study additional mechanisms. For instance, one could model the learning dynamics by which decision makers learn about the environment and determine the decision weights they use (this could be modeled by having decision makers who run regressions on previous observations to determine what weights to use; see Csaszar and Ostler 2020 for a model along these lines). Examining the learning dynamics of discrimination could shed light on the conditions under which policies like affirmative action are more likely to lead to less discrimination. Future research could also examine the effect of using other distributional assumptions (we assumed all distributions to be Normal, as this is a standard assumption and allowed us to derive all the decision rules in closed-form). One could ask, for example, how robust are the model predictions if the cues or inconsistency followed different distributions.

Third, empirical research could test the predictions of our model and establish which regions of the parameter space are more common in practice. Finally, a more general way in which future research could build on our work is by understanding other organizational phenomena in light of the key mechanism underlying our model: that sometimes it is worthwhile to use some cues to proxy for the effect of others (Brunswik 1952, p. 18 called this mechanism "vicarious functioning"). This mechanism not only underlies discrimination but many organizational phenomena. For example, status and reputation are commonly used to proxy for the quality of a product or service (Jensen et al. 2011, Podolny 1993); similarity and homophily, to proxy the strength of a relationship (Granovetter 1973, Rogan and Sorenson 2014); and category membership, to proxy for the legitimacy of an organization (Lamont and Molnár 2002). Looking at such phenomena from the standpoint of our model moves the discussion up the ladder of abstraction, which may allow us to see connections across phenomena and unify seemingly unrelated parts of organization theory. It may also provide fruitful ways to model these phenomena, similar to how it helped us to formally understand discrimination.

5.4 Conclusion

Discrimination seems to be a permanent element characterizing modern society and human relations, not just the labor market. Unlike previous models of statistical discrimination that encourage the use of all information—including group characteristics—to enable more accurate predictions, our model demonstrates that this is only true under a very specific set of conditions. In all other situations, "less is more," such that *not* using characteristics like race or gender results in better predictions. This is particularly true in the presence of decision-maker inconsistency. An overarching implication of our work is that even actors whose sole purpose is to make better predictions can be better off by not discriminating. By providing a theoretical apparatus to more formally reason about discrimination in stylized yet realistic ways, our work shines light on the mechanisms that drive the sources of unequal treatment of equally productive individuals, the situations in which discrimination increases and decreases predictive accuracy, and how discrimination can be reduced—all of which are important steps toward more just organizations and societies.

Appendix: Derivation of the Optimal Rules

A direct way of deriving the optimal rules is to use the projection property of conditional expectation: that $\mathbb{E}[Y \mid X]$ is precisely the function f that minimizes the square loss between Y and $f(X)$ (see, e.g., Shao 2003, p. 40). To use this property, we first write down the relationship between y, x_D, and x_C' as a joint probability distribution. That is,

$$\begin{pmatrix} y \\ x_D \\ x_C' \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1+\sigma_\mu^2 & \rho & 1 \\ \rho & 1 & \rho+\rho_2\sigma_\varepsilon \\ 1 & \rho+\rho_2\sigma_\varepsilon & 1+\sigma_\varepsilon^2 \end{pmatrix} \right).$$

We then derive the rules by computing the corresponding conditional expectations (using the formulas for the multivariate normal; see, e.g., Schott 2017, p. 292) and get:

$$\begin{aligned} \hat{y}_{\mathrm{D}} = \quad \mathbb{E}[y \mid x_D, x_C'] \quad &= (\rho \quad 1) \begin{pmatrix} 1 & \rho+\rho_2\sigma_\varepsilon \\ \rho+\rho_2\sigma_\varepsilon & 1+\sigma_\varepsilon^2 \end{pmatrix}^{-1} \begin{pmatrix} x_D \\ x_C' \end{pmatrix} \\ &= \frac{\rho\sigma_\varepsilon^2 - \rho_2\sigma_\varepsilon}{1-\rho^2+\sigma_\varepsilon^2-\rho_2\sigma_\varepsilon(2\rho+\rho_2\sigma_\varepsilon)} x_D \\ &\qquad + \frac{1-\rho^2-\rho\rho_2\sigma_\varepsilon}{1-\rho^2+\sigma_\varepsilon^2-\rho_2\sigma_\varepsilon(2\rho+\rho_2\sigma_\varepsilon)} x_C'. \end{aligned} \tag{16}$$

$$\hat{y}_{\mathrm{ED}} = \quad \mathbb{E}[y \mid x_D] \quad = \rho x_D \tag{17}$$

$$\hat{y}_{\mathrm{ND}} = \quad \mathbb{E}[y \mid x_C'] \quad = \frac{1}{1+\sigma_\varepsilon^2} x_C'. \tag{18}$$

$$\hat{y}_{\mathrm{ZC}} = \quad \mathbb{E}[y] \quad = 0. \tag{19}$$

To compute the rules' mean square errors (MSEs), we derive an expression for $\mathbb{E}[(y-\hat{y})^2]$ for each of the rules. For instance, the MSE of Extreme Discrimination stems from computing:

$$\begin{aligned} \mathbb{E}[(y-\hat{y}_{ED})^2] \quad &= \mathbb{E}[((x_C+\mu) - ((\rho+\iota_D)x_D))^2] \\ &= \int\int\int\int f(x_C, x_D, \mu, \iota_D)(x_C+\mu-\rho x_D - \iota_1 x_D)^2 \, \mathrm{d}x_C \, \mathrm{d}x_D \, \mathrm{d}\mu \, \mathrm{d}\iota_D \\ &= 1-\rho^2+\sigma_\iota^2+\sigma_\mu^2 \end{aligned}$$

Table 1 contains the resulting formulas. Because performing all the derivations by hand is cumbersome and error-prone, we provide as an online appendix a *Mathematica* notebook that derives from scratch all the formulas used in the paper.

References

Aigner DJ, Cain GG (1977) Statistical theories of discrimination in labor markets. *Industrial and Labor Relations Review* 30(2):175–187.

Allport GW (1954) *The Nature of Prejudice* (Addison-Wesley, Oxford).

Altonji JG, Pierret CR (2001) Employer learning and statistical discrimination. *The Quarterly Journal of Economics* 116(1):313–350.

Arrow KJ (1973) The theory of discrimination. Ashenfelter O, Rees A, eds. *Discrimination in Labor Markets*. (Princeton University Press, Princeton, NJ), 3–33.

Artinger F, Petersen M, Gigerenzer G, Weibler J (2015) Heuristics as adaptive decision strategies in management. *Journal of Organizational Behavior* 36(S1):S33–S52.

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

Bertrand M, Duflo E (2017) Field experiments on discrimination. Banerjee AV, Duflo E, eds. *Handbook of Economic Field Experiments*. Handbook of Field Experiments. (North-Holland, Amsterdam), 309–393.

Bielby WT, Baron JN (1986) Men and women at work: Sex segregation and statistical discrimination. *American Journal of Sociology* 91(4):759–799.

Bohren JA, Haggag K, Imas A, Pope DG (2019) *Inaccurate statistical discrimination: An identification problem* (National Bureau of Economic Research).

Bordalo P, Coffman K, Gennaioli N, Shleifer A (2016) Stereotypes. *The Quarterly Journal of Economics* 131(4):1753–1794.

Bordalo P, Coffman K, Gennaioli N, Shleifer A (2019) Beliefs about gender. *The American Economic Review* 109(3):739–773.

Borjas GJ, Goldberg MS (1978) Biased screening and discrimination in the labor market. *The American Economic Review* 68(5):918–922.

Brehmer B (1981) Models of diagnostic judgments. Rasmussen J, Rouse WB, eds. *Human Detection and Diagnosis of System Failures*. NATO Conference Series. (Plenum Press, New York), 231–239.

Brehmer B (1994) The psychology of linear judgement models. *Acta Psychologica* 87(2–3):137–154.

Brehmer B, Hagafors R, Johansson R (1980) Cognitive skills in judgment: Subjects' ability to use information about weights, function forms, and organizing principles. *Organizational Behavior and Human Performance* 26(3):373–385.

Brighton H, Gigerenzer G (2012) How heuristics handle uncertainty. Todd PM, Gigerenzer G, ABC Research Group, eds. *Ecological Rationality: Intelligence in the World*. (Oxford University Press, Oxford), 33–60.

Brunswik E (1952) *The Conceptual Framework of Psychology* (University of Chicago Press, Chicago).

Coffman KB, Exley CL, Niederle M (2021) The Role of Beliefs in Driving Gender Discrimination. *Management Science*.

Conger AJ, Jackson DN (1972) Suppressor variables, prediction, and the interpretation of psychological relationships. *Educational and Psychological Measurement* 32:579–599.

Correll SJ, Benard S (2006) Biased estimators? Comparing status and statistical theories of gender discrimination. Thye SR, Lawler EJ, eds. *Social Psychology of the Workplace*. Advances in Group Processes. (JAI Press, Oxford), 89–116.

Csaszar FA (2013) An efficient frontier in organization design: Organizational structure as a determinant of exploration and exploitation. *Organization Science* 24(4):1083–1101.

Csaszar FA, Ostler J (2020) A contingency theory of representational complexity in organizations. *Organization Science* 31(5):1198–1219.

Dawes RM (1979) The robust beauty of improper linear models in decision making. *American Psychologist* 34(7):571–582.

Denrell J, Liu C (2012) Top performers are not the most impressive when extreme performance indicates unreliability. *Proceedings of the National Academy of Sciences* 109(24):9331–9336.

Einhorn HJ (1971) Use of nonlinear, noncompensatory models as a function of task and amount of information. *Organizational Behavior and Human Performance* 6(1):1–27.
Fernandez RM, Greenberg J (2013) Race, network hiring, and statistical discrimination. McDonald S, ed. *Networks, Work, and Inequality*. Research in the Sociology of Work. (Emerald Group Publishing Limited, Bingley), 81–102.
Fiske ST (1998) Stereotyping, prejudice, and discrimination. Gilbert DT, Fiske ST, Lindzey G, eds. *The Handbook of Social Psychology*. (McGraw-Hill, New York), 357–411.
Freedle R (2003) Correcting the SAT's ethnic and social-class bias: A method for reestimating SAT scores. *Harvard Educational Review* 73(1):1–43.
Gigerenzer G, Brighton H (2009) Homo heuristicus: Why biased minds make better inferences. *Topics in Cognitive Science* 1(1):107–143.
Gigerenzer G, Gaissmaier W (2011) Heuristic decision making. *Annual Review of Psychology* 62:451–482.
Gigerenzer G, Goldstein DG (1996) Reasoning the fast and frugal way: Models of bounded rationality. *Psychological Review* 103(4):650–669.
Gigerenzer G, Goldstein DG (1999) Betting on one good reason: The take the best heuristic. Gigerenzer G, Todd PM, ABC Research Group, eds. *Simple Heuristics That Make Us Smart*. (Oxford University Press, Oxford), 75–95.
Goldstein DG, Gigerenzer G (2002) Models of ecological rationality: The recognition hypothesis. *Psychological Review* 109(1):75–90.
Granovetter MS (1973) The strength of weak ties. *American Journal of Sociology* 78(6):1360–1380.
Hammond KR, Summers DA (1972) Cognitive control. *Psychological Review* 79(1):58–67.
Harris PM (2001) *The Guide to National Professional Certification Programs* (Human Resource Development Press, Inc., Amherst, MA).
Heckman JJ (2006) Skill formation and the economics of investing in disadvantaged children. *Science* 312(5782):1900–1902.
Hilton JL, von Hippel W (1996) Stereotypes. *Annual Review of Psychology* 47:237–271.
Hogarth RM, Karelaia N (2007) Heuristic and linear models of judgment: Matching rules and environments. *Psychological Review* 114(3):733–758.
Jensen M, Kim BK, Kim H (2011) The importance of status in markets: A market identity perspective. Pearce JL, ed. *Status in Management and Organizations*. (Cambridge University Press, Cambridge, UK), 87–117.
Judd C, Park B (1993) Definition and assessment of accuracy in social stereotypes. *Psychological Review* 100(1):109–128.
Kahneman D, Rosenfield AM, Gandhi L, Blaser T (2016) Noise: How to overcome the hidden cost of inconsistent decision making. *Harvard Business Review* 94(10):38–46.
Kahneman D, Sibony O, Sunstein CR (2021) *Noise: A Flaw in Human Judgment* (Little, Brown Spark, New York).
Karelaia N, Hogarth RM (2008) Determinants of linear judgment: A meta-analysis of lens model studies. *Psychological Bulletin* 134(3):404–426.
Lamont M, Molnár V (2002) The study of boundaries in the social sciences. *Annual Review of Sociology* 28:167–195.
Levinthal DA (1997) Adaptation on rugged landscapes. *Management Science* 43(7):934–950.
Liu C (2021) Why do firms fail to engage diversity? A behavioral strategy perspective. *Organization Science* 32(5):1193–1209.
March JG, Simon HA (1958) *Organizations* (John Wiley & Sons, New York).
Mitchell JR, Shepherd DA, Sharfman MP (2011) Erratic strategic decisions: When and why managers are inconsistent in strategic decision making. *Strategic Management Journal* 32(7):683–704.
Monk EP (2014) Skin tone stratification among black Americans, 2001–2003. *Social Forces* 92(4):1313–1337.

Pager D, Karafin D (2009) Bayesian bigot? Statistical discrimination, stereotypes, and employer decision making. *The Annals of the American Academy of Political and Social Science* 621:70–93.
Pager D, Shepherd H (2008) The sociology of discrimination: Racial discrimination in employment, housing, credit, and consumer markets. *Annual Review of Sociology* 34:181–209.
Park B, Judd CM (2005) Rethinking the link between categorization and prejudice within the social cognition perspective. *Personality and Social Psychology Review* 9(2):108–130.
Payne JW, Bettman JR, Johnson EJ (1993) *The Adaptive Decision Maker* (Cambridge University Press, Cambridge, UK).
Pearl J (2009) *Causality: Models, Reasoning, and Inference* 2nd ed. (Cambridge University Press, Cambridge, UK).
Phelps ES (1972) The statistical theory of racism and sexism. *The American Economic Review* 62(4):659–661.
Podolny JM (1993) A status-based model of market competition. *American Journal of Sociology* 98(4):829–872.
Reichenbach H (1956) *The Direction of Time* (University of California Press, Berkeley, CA).
Reuben E, Sapienza P, Zingales L (2014) How stereotypes impair women's careers in science. *Proceedings of the National Academy of Sciences* 111(12):4403–4408.
Rissing BA, Castilla EJ (2014) House of green cards: Statistical or preference-based inequality in the employment of foreign nationals. *American Sociological Review* 79(6):1226–1255.
Roemer JE (1998) *Equality of Opportunity* (Harvard University Press, Cambridge, MA).
Rogan M, Sorenson O (2014) Picking a (poor) partner: A relational perspective on acquisitions. *Administrative Science Quarterly* 59(2):301–329.
Santelices MV, Wilson M (2010) Unfair treatment? The case of Freedle, the SAT, and the standardization approach to differential item functioning. *Harvard Educational Review* 80(1):106–134.
Schott JR (2017) *Matrix Analysis for Statistics* 3rd ed. (John Wiley & Sons, Hoboken, N.J).
Shao J (2003) *Mathematical Statistics* 2nd ed. (Springer, New York).
Shmueli G (2010) To explain or to predict? *Statistical Science* 25(3):289–310.
Tilcsik A (2021) Statistical discrimination and the rationalization of stereotypes. *American Sociological Review* 86(1):93–122.
Wu S, Harris TJ, Mcauley KB (2007) The use of simplified or misspecified models: Linear case. *The Canadian Journal of Chemical Engineering* 85(4):386–398.

Figures and Table

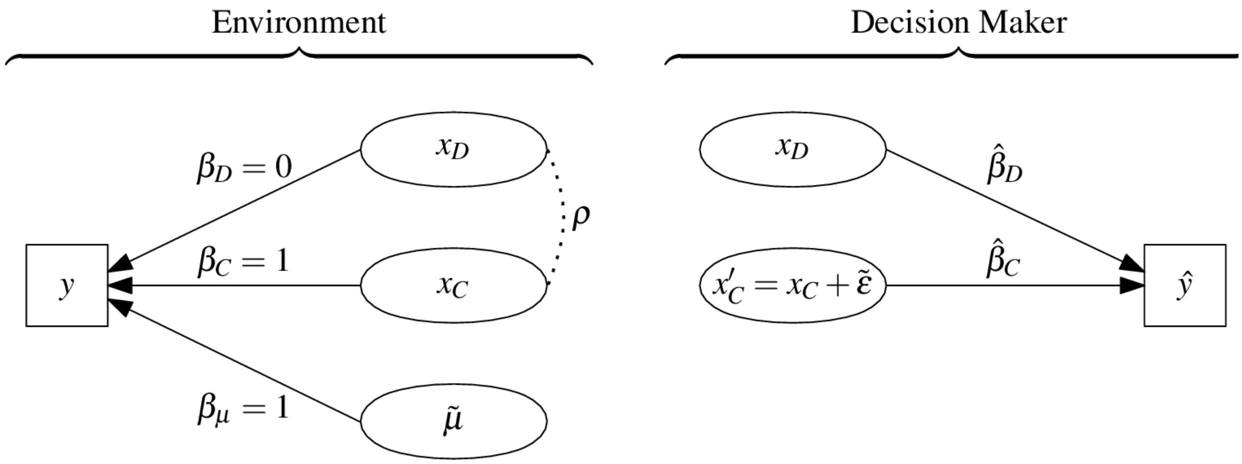

Figure 1: Graphical description of the model (using Brunswik's 1952 lens model framework).

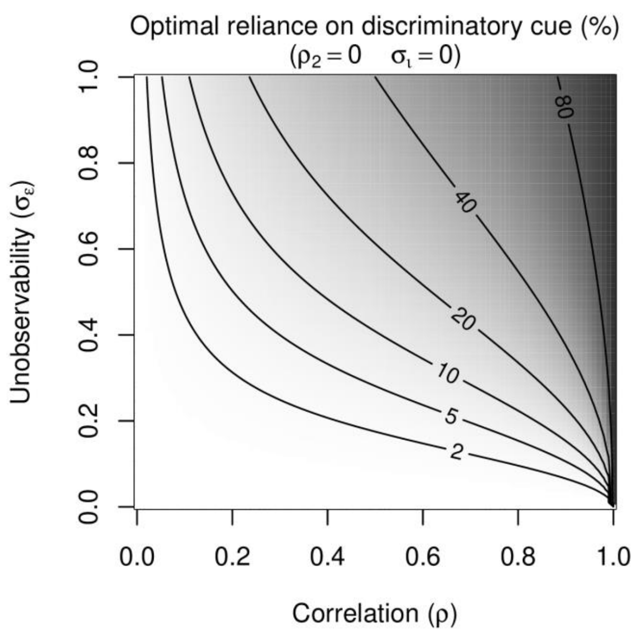

Figure 2: Optimal reliance on discriminatory cue ($\frac{\widehat{\beta}_D^*}{\widehat{\beta}_D^* + \widehat{\beta}_C^*}$) as a function of correlation ρ (in the x-axis) and unobservability σ_ε (in the y-axis).

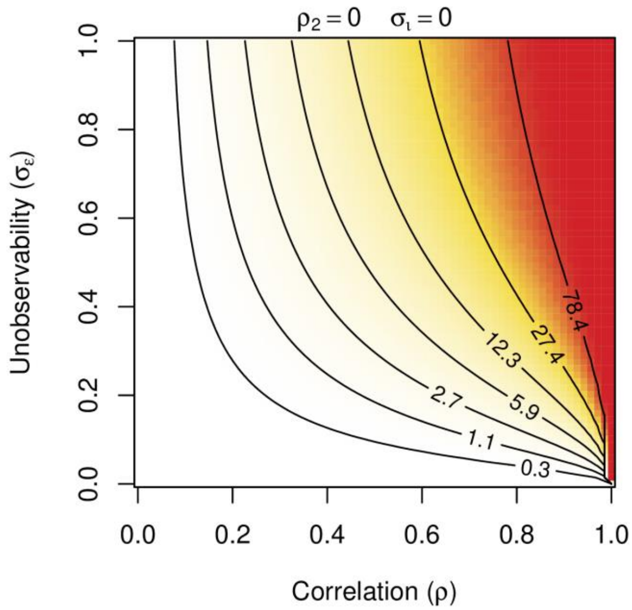

Figure 3: Cost of not discriminating ($\frac{\mathrm{MSE_{ND}}}{\mathrm{MSE_D}} - 1$) as a function of correlation ρ (in the x-axis) and unobservability σ_ε (in the y-axis).

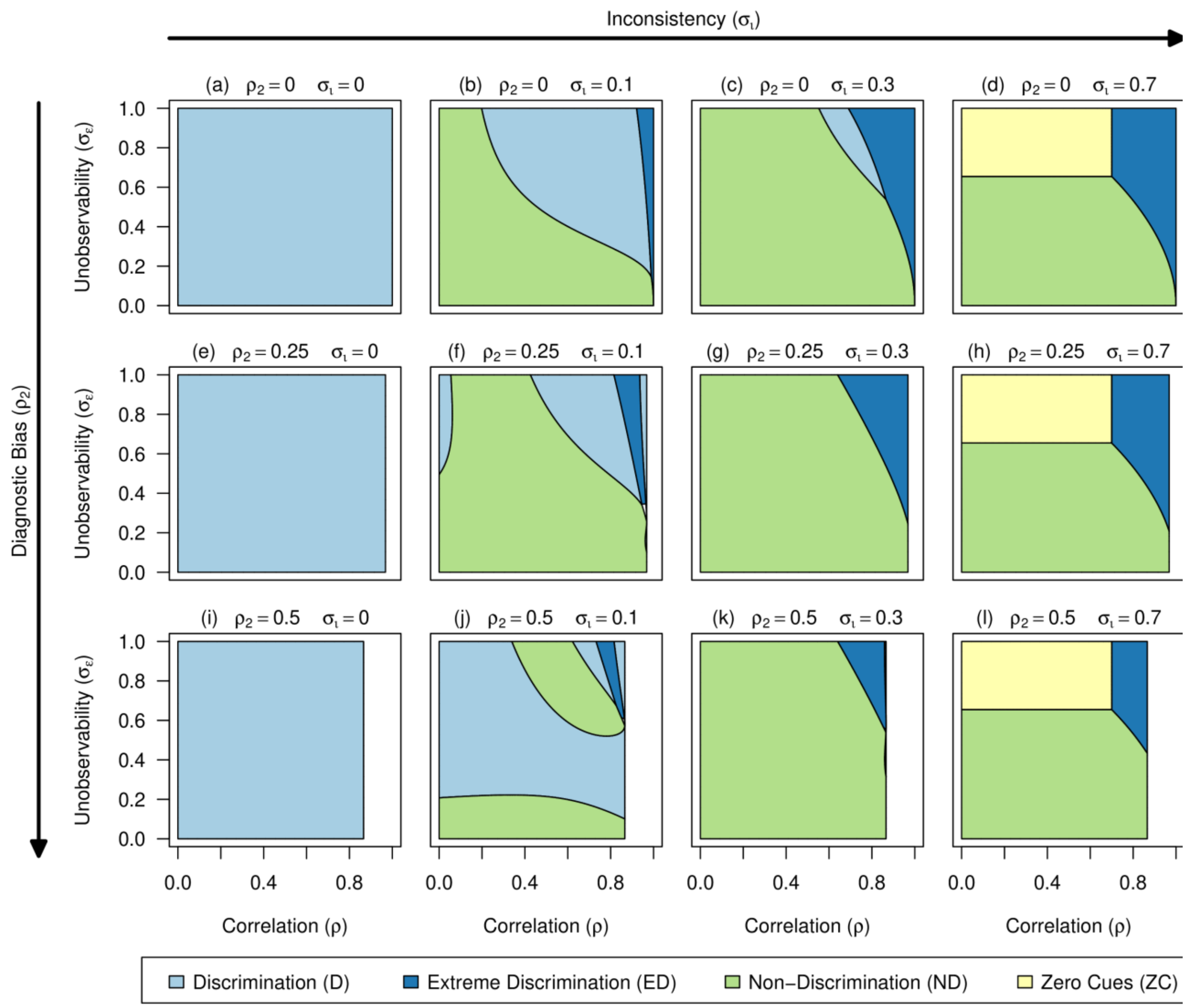

Figure 4: Decision rule that achieves the lowest MSE in each region of the parameter space as a function of correlation ρ (in the x-axis), unobservability σ_ε (in the y-axis), inconsistency σ_ι (across the columns of panels), and diagnostic bias ρ_2 (down the rows of panels).

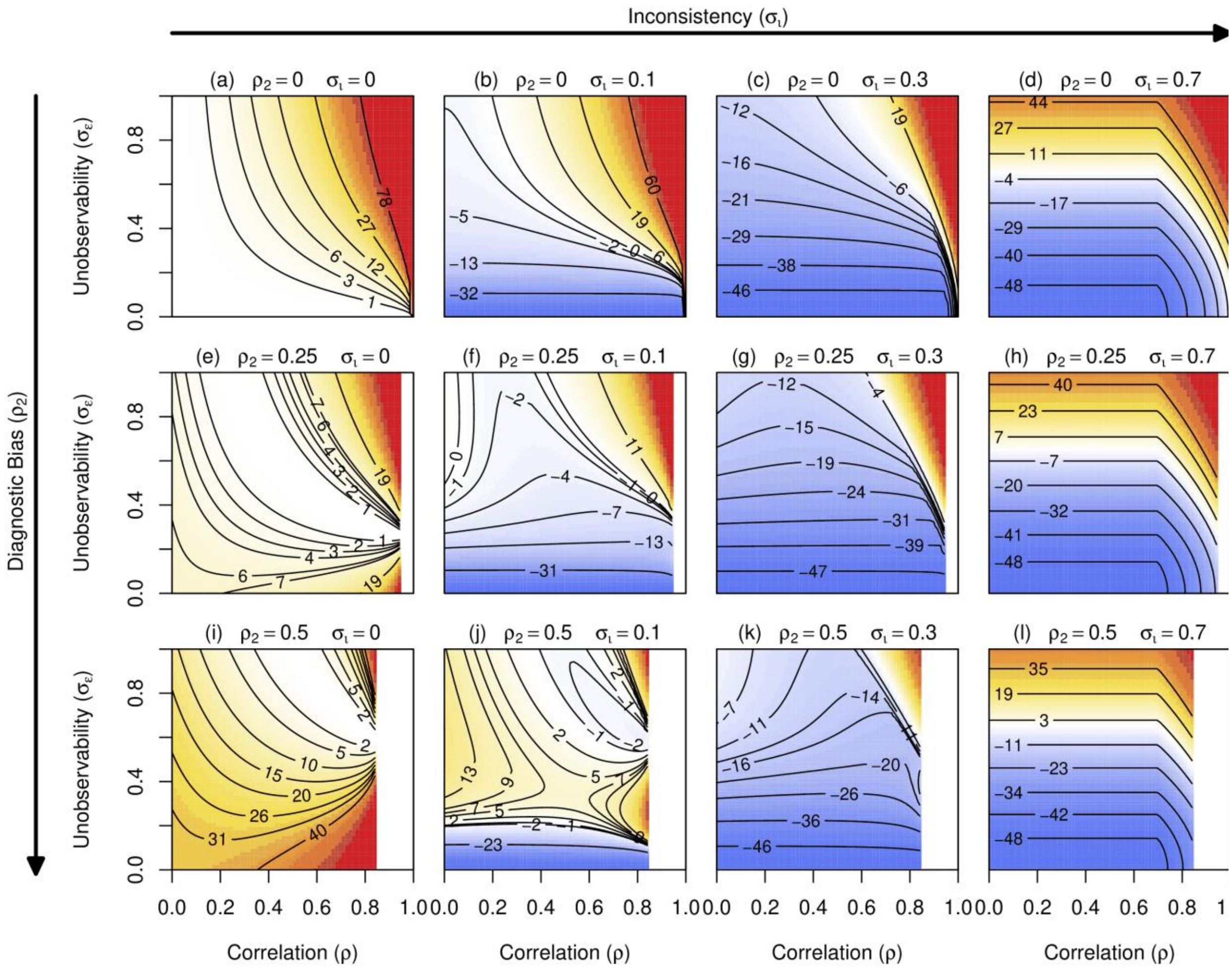

Figure 5: Cost of not discriminating $(\frac{\mathrm{MSE_{ND}}}{\min(\mathrm{MSE_D}, \mathrm{MSE_{ED}}, \mathrm{MSE_{ZC}})} - 1)$ as a function of correlation ρ (in the x-axis), unobservability σ_ε (in the y-axis), inconsistency σ_ι (across the columns of panels), and diagnostic bias ρ_2 (down the rows of panels).

Discrimination (D) under $\rho_2 = 0$. Cues used: $x_\mathrm{D}, x'_\mathrm{C}$	
$\hat{y}_\mathrm{D} =$	$\frac{\rho\sigma_\varepsilon^2}{1-\rho^2+\sigma_\varepsilon^2} x_\mathrm{D} + \frac{1-\rho^2}{1-\rho^2+\sigma_\varepsilon^2} x'_\mathrm{C}$
$\mathrm{MSE}_\mathrm{D} =$	$\frac{\left(\rho^2-1\right)\sigma_\varepsilon^2}{(\rho-\sigma_\varepsilon)(\rho+\sigma_\varepsilon)-1} + \left(2 + \sigma_\varepsilon^2\right)\sigma_\iota^2 + \sigma_\mu^2$
Discrimination (D). Cues used: $x_\mathrm{D}, x'_\mathrm{C}$	
$\hat{y}_\mathrm{D} =$	$\frac{\rho\sigma_\varepsilon^2-\rho_2\sigma_\varepsilon}{1-\rho^2+\sigma_\varepsilon^2-\rho_2\sigma_\varepsilon(2\rho+\rho_2\sigma_\varepsilon)} x_\mathrm{D} + \frac{1-\rho^2-\rho\rho_2\sigma_\varepsilon}{1-\rho^2+\sigma_\varepsilon^2-\rho_2\sigma_\varepsilon(2\rho+\rho_2\sigma_\varepsilon)} x'_\mathrm{C}$
$\mathrm{MSE}_\mathrm{D} =$	$\frac{\sigma_\varepsilon^2\left(\rho_2^2+\rho^2-1\right)}{((\rho_2-1)\sigma_\varepsilon+\rho)(\rho_2\sigma_\varepsilon+\rho+\sigma_\varepsilon)-1} + \left(2 + \sigma_\varepsilon^2\right)\sigma_\iota^2 + \sigma_\mu^2$
Extreme Discrimination (ED). Cues used: x_D	
$\hat{y}_\mathrm{ED} =$	ρx_D
$\mathrm{MSE}_\mathrm{ED} =$	$1 - \rho^2 + \sigma_\iota^2 + \sigma_\mu^2$
Non-Discrimination (ND). Cues used: x'_C	
$\hat{y}_\mathrm{ND} =$	$\frac{1}{1+\sigma_\varepsilon^2} x'_\mathrm{C}$
$\mathrm{MSE}_\mathrm{ND} =$	$\frac{\sigma_\varepsilon^2}{1+\sigma_\varepsilon^2} + (1 + \sigma_\varepsilon^2)\sigma_\iota^2 + \sigma_\mu^2$
Zero Cues (ZC). Cues used: $-$	
$\hat{y}_\mathrm{ZC} =$	0
$\mathrm{MSE}_\mathrm{ZC} =$	$1 + \sigma_\mu^2$

Table 1: Summary of the prediction formula ($\hat{y}$) and error formula (MSE) for the different rules.