Published in The Economic Journal.
Volume 131 Aug 2020, pp. 2018-2032.
https://doi.org/10.1093/ej/ueaa107

A CHARACTERIZATION OF "PHELPSIAN" STATISTICAL DISCRIMINATION

CHRISTOPHER P. CHAMBERS AND FEDERICO ECHENIQUE

ABSTRACT. We establish that a type of statistical discrimination—that based on informativeness of signals about workers' skills and the ability to appropriately match workers to tasks—is possible if and only if it is impossible to uniquely identify the signal structure observed by an employer from a realized empirical distribution of skills. The impossibility of statistical discrimination is shown to be equivalent to the existence of a fair, skill-dependent, remuneration for workers. Finally, we connect the statistical discrimination literature to Bayesian persuasion, establishing that if discrimination is absent, then the optimal signaling problem results in a linear payoff function (as well as a kind of converse).

(Chambers) DEPARTMENT OF ECONOMICS, GEORGETOWN UNIVERSITY
(Echenique) DIVISION OF THE HUMANITIES AND SOCIAL SCIENCES, CALIFORNIA INSTITUTE OF TECHNOLOGY
Date: August 18, 2020.
Echenique thanks the NSF for support through the grants SES-1558757 and CNS-1518941. We are grateful to Ludvig Sinander and Leeat Yariv for comments on a previous draft, and to the editor and two anonymous referees for many suggestions that helped improve the paper.

1. Introduction

In seminal contributions, Arrow (1971; 1973) and Phelps (1972) postulated that discrimination along racial lines, or gender identities, can have a statistical explanation. In this note we study a version of statistical discrimination motivated by the ideas in Phelps (1972) and Aigner and Cain (1977). The main thrust is that two populations of workers, who are in essence identical, may have different economic remunerations for purely informational reasons.[1]

Phelps' theory connects workers' remuneration with the information available about their skills. His theory assumes a firm who observes a signal about the underlying skills of a worker. The worker is paid her expected contribution to the firm, conditional on the firm's observed signal about the worker. A competitive market ensures that workers are paid their contributions.

Phelps imagines two populations of workers: group X and group Y, who differ in the information available about their skills. Say that the signal about the skill of a member of group X is more informative than the signals for group Y. Aigner and Cain remark that, in this setting, the average payment to workers in X may be higher than that of workers in Y, even when their distribution of skills is identical. We term this phenomenon "Phelpsian" statistical discrimination.[2]

In Phelps' model, the signal may be the result of a test that has been designed with a population from group X in mind. The signal implemented by the test will then be more informative about the skills of an X worker than a Y worker. In support, Aigner and Cain (1977) cite evidence from the education literature to the effect that the SAT is less informative about the abilities of African-American students than it is for white students.

[1] Arrow's theory of statistical discrimination relies on a coordination failure, and is quite different from Phelps'. Statistical discrimination stands in contrast with taste-based discrimination, as in Becker (1957).

[2] Phelps actually compares the payment to a worker from X or Y conditional on getting a high or a low signal. Aigner and Cain (1977) point out that this does not lead to what is normally called discrimination in economics, and offer an "alternative model," which we follow here. See Section 4.4 below for a discussion.

We formulate the theory of statistical discrimination using the language of the recent literature on information design. A risk-neutral firm is characterized by a technology. We use the terms "firm" and "technology" interchangeably. The firm observes a signal about a worker's skills, and bases the payment to the worker on the revenue it expects to gain from how valuable her skills are for the firm's technology.

A population of workers comes with a distribution over signals: an information structure. The information structure of population X may be different from that of population Y. We say that statistical discrimination is present if the two populations of workers have the same overall distribution of skills, but receive different payments in expectation as a result of their different information structures.

We show that the focus on informativeness in Phelps and Aigner-Cain is misleading. There may be statistical discrimination even when the information structure of one population is not more informative than the other. We show that the relevant property is lack of identification (in the econometric sense) of signals from skills. Aigner and Cain trace statistical discrimination to pure informativeness. We argue that the situation is more general.

Our main result connects statistical discrimination with two seemingly distinct properties of the economic environment: one is identification of signals from skills, and the other is the linearity of firm revenue in "fair" skill-dependent payoffs. First, we prove that statistical discrimination is not possible if and only if every given distribution of skills results in a unique distribution of signals. By definition, when discrimination is possible, this identification property must be violated. Our contribution lies in establishing the converse: whenever identification is impossible, discrimination can arise. Specifically, if identification is impossible, there is a firm which will discriminate.

Second, we show that identification, and therefore the absence of discrimination, is equivalent to the existence of a skill-based remuneration for workers. Such remunerations amount to a payment for skill. Each list of skills must be associated with a payoff, which is independent of any signal, and every worker is paid the expectation according to the distribution of skills inherent in her realized signal. In other words, remuneration depends only on expected skill; which is, in a sense, fair.

When wages are determined in this way, a population's expected wage is linear in its distribution of skills. Therefore no statistical discrimination is possible. Again the contribution lies in establishing the converse: the emergence of statistical discrimination is equivalent to the absence of such fair remunerations.

As a corollary we provide a characterization of fair skill-based remunerations as a sort of "equilibrium" payments. With no discrimination, worker remunerations are the least that a worker can be paid, subject to the satisfaction of certain outside options. In particular the expected remuneration of a worker with a given signal equals the minimum expected skill-dependent payment the worker could receive, subject to achieving at least what she would get if she had any other signal value.

Finally, we show that the optimal information structure in the sense of Kamenica and Gentzkow (2011) achieves precisely the fair remuneration in our results. This connects back to Phelps' and Aigner-Cain's analysis of the informational content of workers' signal structures. In the absence of discrimination, workers are endowed with a maximally informative signaling structure.

2. Preliminary definitions

2.1. **Notation.** If A is a closed subset of a Euclidean space, we denote by $\Delta(A)$ the set of Borel probability measures on A.

2.2. **Signals, information, and Blackwell informativeness.** Let Θ be a finite set. Think of the elements of Θ as the possible states of the world: in the model below, the elements of Θ represent skills possessed by a worker, which are generally unknown by a firm. These states are relevant in that they will determine the payoffs of a given task to which a worker may be assigned. The set $\Delta(\Theta)$ of probability measures on Θ can be interpreted as the set of beliefs one may have about the elements of Θ.

We can also think of $\Delta(\Theta)$ as the set of possible values that a signal distribution may take. For example, imagine that the state of the world can only take two values, $\Theta = \{\theta_0, \theta_1\}$, and that these two states are equally likely. An agent receives information about the state of the world in the form of a signal s, which can take either the value $s = \spadesuit$ or $s = \clubsuit$. Suppose that, if the state is θ_0 she gets the $\spadesuit$ signal with probability 3/4, and $\clubsuit$ with probability 1/4. Conversely, if the state is θ_1 she gets the signal $\spadesuit$ with probability 1/4, and $\clubsuit$ with probability 3/4. If our agent observes $\spadesuit$, then she will hold a posterior belief of 3/4 for state θ_0 and 1/4 for θ_1. If, instead, she observes $\clubsuit$, then she will hold a posterior belief of 1/4 for state θ_0 and 3/4 for θ_1. So instead of keeping track of the labels $\spadesuit$ or $\clubsuit$ on signals, we can identify them with the posterior beliefs that they induce. In consequence, we think of the set $\Delta(\Theta)$ as the set of all possible values that a signal may take. This is without loss of generality.[3]

The $\clubsuit/\spadesuit$ signal example can be interpreted as a distribution over $\Delta(S)$: with probability 1/2 we have the signal realization $((3/4),(1/4))$ and with probability 1/2 we have the signal $((3/4),(1/4))$. Generally speaking, an *information structure* is a probability distribution $\pi \in \Delta(\Delta(\Theta))$.

Information structures were famously analyzed by Blackwell (1953), who discussed the idea that some signal structures are more informative than others. The definition of informativeness has to be independent of any individual agent. For example, in our model below, one information structure may be better than another for one firm, but a second firm may hold the opposite ranking. Blackwell's definition says that information structure π is *more informative* than π' if *every* agent derives more value from π than from π'. Blackwell proved that his informativeness property is equivalent to π' being a mean-preserving spread of π.

In our model, we shall see that different firms may value information structures differently; but when π is more informative than π' in the

[3] In our framework, the signal itself has no inherent value for the firm.

Blackwell sense, then all firms will prefer to base their decisions on π over basing them on π'. In essence, this is the basic source of the type of discrimination analyzed by Phelps, and by Aigner and Cain. More information about Θ allows a firm to more appropriately match a worker to a task.

2.3. **Identification.** In an economic model, something is *identified* if its values can be "backed out" from available data (Haavelmo, 1944). Mathematically, the property translates into an injective relation. Imagine a model that has a parameter π and an observable magnitude p_π. The parameter is *identified* if different values of π must always result in different p_π. Put differently, from knowing p_π one is able to determine, or back out, π. The map $\pi \mapsto p_\pi$ is one-to-one.

3. The model

We present a model of workers' remuneration: A firm hires a worker and expects the worker to generate revenue for the firm. We assume (following Phelps) that the labor market is competitive, and hence the firm pays the worker its expected revenue. Our theory of worker remuneration is therefore a theory of firm revenue.

The firm faces uncertainty over the revenue it obtains from hiring a worker: its revenue depends on the worker's skills, and how those skills match up with the firm's technology; but the firm does not observe skills. Instead, it gets a signal and forms a posterior belief distribution over skills. Our model has three components: A set of possible skills, the technology available to the firm, and a population of workers.

The first component is the set of possible skills; these are collected in a finite set Θ. Depending on the realized skill $\theta \in \Theta$, the worker could be good at computer programming, or driving long-haul trucks. The second component is the firm's technology, given as a nonempty finite set $A \subseteq \mathbf{R}^\Theta$. How do Θ and A match up? For each $a \in A$, and each skill $\theta \in \Theta$, the worker generates revenue $a(\theta) \in \mathbf{R}$ for the firm.[4]

[4]There is no loss of generality for any of our results in restricting $a(\theta) \geq 0$; in the interest of simplicity we allow it to be more general.

We could instead have written down a detailed model of the worker's role in the firm, for example an IT firm is going to value computer programming more than truck-driving, but ultimately what matters for the firm (and for our model) is the revenue generated. So we use the numbers $a(\theta)$, for $a \in A$, as a simple reduced-form model of the interaction between the worker with skill θ and the firm's technology A.

The third component of our model is the population π of workers, which captures the information available to the firm. Consider the ideas introduced in Section 2.2, and think of the possible skills Θ as the possible states of the world. The worker's skill θ is unobservable. Instead, the firm observes a signal $s \in \Delta(\Theta)$. Thus the expected value to a risk-neutral firm with technology A of a worker with signal s is

$$v_A(s) = \max_{a \in A} \int_{\Theta} a(\theta) ds(\theta)$$

The set Θ is finite, so we may identify s with its probability function and write $\int_{\Theta} a(\theta) ds(\theta)$ as $\sum_{\theta \in \Theta} a(\theta) s(\theta)$.

All possible signals are collected in a (Borel) set $\mathcal{S} \subseteq \Delta(\Theta)$. Observe that the revenue function v_A is the "value function" of A, as in Blackwell (1953), Machina (1984), or the profit function in the textbook theory of the firm when s is now interpreted as a "price." It is well known that such functions are always convex.

Finally, the population of workers is reflected in $\pi \in \Delta(\mathcal{S})$; π is both an information structure and a population.

To sum up, the model primitives consist of the triple

$$(\Theta, \mathbf{A}, \mathcal{S}),$$

where Θ is the finite set of types, $\mathbf{A}$ is the collection of finite technologies $A \subseteq \mathbf{R}^{\Theta}$, and $\mathcal{S} \subseteq \Delta(\Theta)$ is a set of possibles signals.[5]

[5]We take $\mathbf{A}$ to be the set of all possible firm technologies, but as we shall see it will be sufficient to restrict attention to binary sets: technologies with one or two elements.

Finally, each worker belongs to a population that generates signals according to an *information structure* $\pi \in \Delta(\mathcal{S})$. Here we are motivated by Phelps (1972) and Aigner and Cain (1977). Think of different populations as having different societal identities. They could for example be races, ethnicities, or genders. One population could be the set of all White workers, or all Hispanic workers. These differ in how firms learn about their skill: in the information structure that they are endowed with. A population with information structure π has an implied distribution of skills equal to

$$p_\pi(E) = \int_{\mathcal{S}} s(E) d\pi(s).$$

Think of $p_\pi(E)$ as the frequency of workers with skill $\theta \in E$ in the population. Then $p_\pi \in \Delta(\Theta)$ is the distribution of skills in a population that has information structure π. Thus, for our purposes, π is the same as a population.

A population's information structure π plays two roles. It describes both the overall skill distribution p_π in the population, and it also specifies the distribution of signals in the population. The latter describes how firms learn about the skills possessed by a worker who belongs to the population. In other words, it reflects something about the informativeness about the signal ascribed to the population.

Recall that we operate under the assumption of competitive labor markets, and therefore equate revenue and remuneration. The expected revenue can be computed by first calculating the expected revenue from each possible signal that the firm can observe, and then computing the expected revenue when signals obey the probability law in the relevant information structure. Specifically, a population that generates signals according to an information structure $\pi \in \Delta(\mathcal{S})$ will obtain a distribution of wages resulting from the interaction of π with v_A. In particular, the mean remuneration of this population is

$$\int_{\mathcal{S}} v_A(s) d\pi(s)$$

The idea of discrimination we study is that two populations $\pi, \pi' \in \Delta(\mathcal{S})$ may have the same distribution of skills (so that $p_\pi = p_{\pi'}$), but receive different payoffs because they have two different information structures. Formally, we say that the set of signals $\mathcal{S}$ is *non-discriminatory* if, for any information structures $\pi, \pi' \in \Delta(\mathcal{S})$, and any finite set $A \subseteq \mathbf{R}^\Theta$, if $p_\pi = p_{\pi'}$, then

$$\int_{\mathcal{S}} v_A(s) d\pi(s) = \int_{\mathcal{S}} v_A(s) d\pi'(s).$$

In words, the property of being non-discriminatory means that two populations that have no "genuine" difference in their distribution of skills, and only differ in how firms get to learn about their members, should not differ in mean remuneration. On the other hand, $\mathcal{S}$ is discriminatory if there exists at least one firm for which two populations with the same aggregate distribution of skills are remunerated differently. It does not require this to hold for every possible firm.[6]

3.1. **Motivation and a Phelpsian example.** We start with a simple example of statistical discrimination. It is a minimal example; the simplest we can think of that delivers the Phelpsian message. Let $\Theta = \{\theta_1, \theta_2, \theta_3\}$ be the set of skills, and $A = \{(1, 0, 0), (0, 1/2, 3)\}$ be the firm technology. Observe that, with this specification, workers are not "high" or "low" quality, but they simply have differing aptitudes for the different components of the firm technology. They "fit" in different parts of the firm's revenue-generating process.

Suppose that

$$\mathcal{S} = \{(1, 0, 0), (1/2, 1/2, 0), (0, 1/2, 1/2), (0, 0, 1)\}$$

[6]We want to talk about discrimination for fixed populations that differ in some observable trait (such as race), but not in their underlying distribution of skills. If we fix a particular firm, meaning a particular technology, then there may be no discrimination because the relation between the observable trait in question and the distribution of signals is unrelated to how this particular firm, with its particular technology, uses information. We wouldn't want to describe such a situation as the absence of statistical discrimination. Instead, we want to say that there is not statistical discrimination when no firm would use the informational differences between two population to provide higher remuneration to one over the other.

is the set of possible signals.

There are two populations of workers, say X and Y. The two populations differ in the information that the firm obtains about their skills. The workers might take a test, as in Phelps (1972), and the informational content of the test might be different for the two populations. So X members emit signals about their skills according to an information structure π, while Y members' information structure is π'.

The two information structures, π and π', are described in the table below, together with the revenue function v_A resulting from our assumed Θ and A. Observe that $p_\pi = p_{\pi'} = (1/3, 1/3, 1/3)$, reflecting that the populations overall have the same skills. Importantly, the two information structures are *not* ranked by Blackwell informativeness.[7]

	$s = (1,0,0)$	$s = (1/2,1/2,0)$	$s = (0,1/2,1/2)$	$s = (0,0,1)$
$\pi(s)$	1/3	0	2/3	0
$\pi'(s)$	0	2/3	0	1/3
$v_A(s)$	1	1/2	7/4	3

A worker from group X reveals that she is either good for $a_1 = (1,0,0)$ or $a_2 = (0,1/3,3)$. The Y worker reveals the same kind of information, but less usefully for the firm: a signal $s = (1/2,1/2,0)$ tells the firm that a_1 is the optimal choice given the information at hand, but leaves it with some doubts as to whether a_2 may have been optimal. In consequence, we have

$$\int_{\mathcal{S}} v_A(t) d\pi(t) = 1/3 + 7/6 > 1/3 + 1 = \int_{\mathcal{S}} v_A(t) d\pi'(t).$$

If workers are paid according to the revenues that they contribute to the firm, as would be the case in a competitive market, then X workers are paid more than Y workers in the aggregate. The differences in expected (or population) remuneration between the two is purely a consequence of the informational content in their corresponding signal structures.

[7]The firm itself is able to rank the two signal structures, but Blackwell informativeness implies a ranking that holds for all firms.

The example shows that statistical discrimination extends beyond a situation where information structures are ranked, or ordered, in terms of informativeness. Information structures π and π' are not ranked in terms of informativeness; neither is a mean-preserving spread of the other.

Moreover, in the example the two different information structures have the same mean. This is a necessary requirement for the existence of statistical discrimination. It is important to point out, however, that in this model, the aggregate skill level p_π can *always* be inferred from wages, even when there is discrimination. We present Proposition 1, which is obvious, to make this point.

Proposition 1. *Skill distributions are identified from revenue: if two populations $\pi, \pi' \in \Delta(\mathcal{S})$ differ in their skill distribution, so that $p_\pi \neq p_{\pi'}$, then there exists some firm $A \in \mathbf{A}$ at which the expected wage of the two populations differ.*[8]

The proposition results by finding $\theta \in \Theta$ for which $p_\pi(\theta) \neq p_{\pi'}(\theta)$, and setting $A = \{a\}$, where $a = \mathbf{1}_\theta$, the indicator function of θ.

4. Statistical discrimination and identification

The discussion in Section 3.1 suggested that discrimination is tied to identification, but skills distributions are always identified from revenue, even when there is discrimination (Proposition 1). The relevant identification problem is not about skills, but about the ability to back out the underlying information structure.

Formally, we say that $\mathcal{S}$ is *identified* if for any $\pi, \pi' \in \Delta(\mathcal{S})$, $p_\pi = p_{\pi'}$ implies that $\pi = \pi'$.

Our main result is that the absence of discrimination is equivalent to the ability to estimate a skills distribution from the information structure. Importantly, we show that this can *only* happen when worker remunerations are, in a sense, *fair.*

[8] This result is trivial. We prove a stronger statement, see Proposition 5 of Section 6.

4.1. **Fair remunerations.** Remunerations are fair if individual workers are paid according to their skills. Specifically, if there is a mapping from skills to payoffs so that a worker gets paid according to their signal about skills, not according to the information structure in the population that they come from. In consequence, we say that $\mathcal{S}$ *admits fair valuations* if for any $A \in \mathbf{A}$ there is $\alpha_A \in \mathbf{R}^\Theta$ for which

$$v_A(s) = \sum_\theta \alpha_A(\theta) s(\theta)$$

for all $s \in \mathcal{S}$.

In words, $\alpha_A(\theta)$ is the value to the firm with technology A of a worker with skills θ. When the firm observes signal $s \in \mathcal{S}$ it pays the worker the expected value of α_A according to s.

Importantly, if $\pi \in \Delta(\mathcal{S})$ and $\mathcal{S}$ admits fair valuations, then

$$\int v_A(s) d\pi(s) = \alpha_A \cdot \int s d\pi(s) = \alpha_A \cdot p_\pi.$$

Hence, under fair valuations, the expected payment to a population of workers with information structure π only depends on the distribution of skills in that population.

In particular, it should be clear that fair valuations imply non-discrimination. Our contribution will be to show that the converse holds, and how the property of fair valuations is connected to identification.

4.2. **Main result.**

Theorem 2. *A collection of signals $\mathcal{S}$ is non-discriminatory if and only if it is identified, which holds if and only if $\mathcal{S}$ admits fair valuations.*

For example, if $\Theta = \{\theta_1, \theta_2, \theta_3\}$, and

$$\mathcal{S} = \{(1/3, 1/3, 1/3), (0, 1/4, 3/4), (0, 3/4, 1/4)\},$$

then all of the equivalent conditions of Theorem 2 are satisfied. To see that identification is satisfied, observe that for $\pi \in \Delta(\mathcal{S})$, we can infer $\pi(\{1/3, 1/3, 1/3\}) = 3p_\pi(\theta_1)$. Similarly,

$$\pi(\{0, 1/4, 3/4\}) = (3/2)p_\pi(\theta_3) - (1/2)p_\pi(\theta_3) - p_\pi(\theta_1).$$

Finally,

$$\pi(\{0, 3/4, 1/4\}) = (3/2)p_\pi(\theta_2) - (1/2)p_\pi(\theta_3) - p_\pi(\theta_1).$$

An important consequence of Theorem 2 is the existence of fair valuations, which means that there exists a "fundamental" value α_A, for each firm A, so that workers are paid according to the expected value $\sum_\theta \alpha_A(\theta)s(\theta)$. Fair valuations come about because non-discrimination implies identification, which in turn means that the maximum expected revenue that can be achieved by any information structure with a given mean skill is an affine function of the mean skill.

It should be obvious that the existence of fair valuations implies non-discrimination. A population π would under fair valuations receive an expected remuneration that equals $\alpha_A \cdot p_\pi$: the value of the distribution of skills in the population. Such remuneration would be equal for any two populations that have the same skill distribution p. The thrust of the theorem lies in proving that non-discriminatory $\mathcal{S}$ admit fair valuations.

The equivalence of the three properties rests on a simple geometric argument. Suppose that $\mathcal{S}$ is finite. Observe that $\mathcal{S}$ being identified implies that, when each $s \in \mathcal{S}$ is viewed as a member of $\mathbf{R}^\Theta$, $\mathcal{S}$ is a linearly independent set.[9] Now consider what happens if $\mathcal{S}$ is not linearly independent: we can find $\beta \in \mathbf{R}^\mathcal{S} \setminus \{0\}$ for which $\sum_{s\in\mathcal{S}} \beta(s)s(\theta) = 0$. Since $\beta \neq 0$, we can find $\gamma \in \mathbf{R}^\mathcal{S}$ for which $\beta \cdot \gamma \neq 0$. The vector β can be scaled to become $\pi - \pi'$, and γ is v_A: the only difficulty here is that v_A cannot have an arbitrary shape, and must arise from a value function on $\mathbf{R}^\Theta$, but this poses no special difficulty. So lack of identification is equivalent to the existence of possible discrimination. Fair remuneration comes from the fact that on a linearly independent set,

[9] A violation of linear independence means that $s = \sum_{s'\in\mathcal{S}\setminus\{s\}} \beta(s')s'$ for some $\beta \neq 0$. By rearranging terms, then $s + \sum_{s':\beta(s')\leq 0}(-\beta(s'))s' = \sum_{s':\beta(s')\geq 0} \beta(s')s'$. Observe that the right hand side cannot be equal to 0, as all members of $\mathcal{S}$ are elements of $\Delta(\theta)$ and hence have nonnegative components. Observe that $\sum_\theta \sum_{s'\in\mathcal{S}\setminus\{s\}} \beta(s')s'(\theta) = 1$ as $\sum_\theta s(\theta) = 1$. Therefore, we can renormalize each side of the equation to find π, π' for which $\sum_\theta s(\theta)\pi(\theta) = \sum_\theta s(\theta)\pi'(\theta)$.

to find the average remuneration, it is necessary and sufficient to find the share of workers of each type.

The next claim is a simple duality result.

Proposition 3. *If $\mathcal{S}$ admits fair valuations, then for each firm $A \in \mathbf{A}$,*

$$\sum_\theta \alpha_A(\theta)s(\theta) = \inf\{\sum_\theta y(\theta)s(\theta) : y \in \mathbf{R}^\Theta \text{ and } v_A(s') \leq \sum_\theta y(\theta)s'(\theta) \forall s' \in \mathcal{S}\}.$$

To understand Proposition 3, suppose that a worker with skill θ can guarantee themselves a payment of $y(\theta)$ on the market. Then the expected remuneration for a worker with signal s', if she goes on the market, is $\sum_\theta y(\theta)s'(\theta)$. A firm would not pay the worker with signal s' more than her outside option, so that the remuneration to s' at a firm with technology A (which equals the worker's contribution to the firm, $v_A(s')$) cannot exceed $\sum_\theta y(\theta)s'(\theta)$. Finally the lowest payments must by definition be minimal among all those payments that satisfy that, for all s', $\sum_\theta y(\theta)s'(\theta) \geq v_A(s')$. Fair payoffs α_A can thus be viewed as the result of competitive forces. The workers earn the least possible that is compatible with their outside options.

Proposition 3 is an illustration of a standard duality result, heavily used in the information design literature. This duality result says nothing more than that the concave envelope of v_A (the "inf" in the equation) is the same as the maximal average remuneration given to an arbitrary population π. Under fair remuneration, that maximal average remuneration is given by the linear specification on the left hand side of the equation. We now proceed to flesh out this connection to information design.

4.3. **Connection to Bayesian persuasion.** The recent literature on Bayesian persuasion (Kamenica and Gentzkow (2011)) deals with the optimal design of information structures. In our model, populations of workers are identified with an information structure, and we have in particular traced Phelpsian statistical discrimination to the informative content of a population of workers. Statistical discrimination is connected to Bayesian persuasion if we consider the populations that are maximally informative.

For each technology A, consider the optimal information structure for firm A when facing a population of workers with skill distribution $p \in \Delta(\Theta)$. The firm is uncertain about the skill $\theta \in \Theta$ of a particular worker. She learns about θ from observing a signal s that is drawn according to an information structure π. Since the skill distribution is known to be p, we must have $p = p_\pi$. Thus the problem of optimal information design (Kamenica and Gentzkow, 2011) for firm A is:

$$\begin{array}{ll} \max & \displaystyle\int_T v_A(\tilde{s}) d\pi(\tilde{s}) \\ \text{s.t} & \begin{cases} \pi \in \Delta(T) \\ p = \int_T \tilde{s} d\pi(\tilde{s}), \end{cases} \end{array}$$

with T being the closed convex hull of $\mathcal{S}$ and $p \in T$.

The solutions to this problem are the most profitable information structures for firm A. Let $W_A : T \to \mathbf{R}$ be the value function of this problem: the value to the firm of an optimal information structure. Our next result connects optimal information design with statistical discrimination. We need one piece of notation, let ∂T denote the extreme points of T, so that T is the closed convex hull of ∂T. Note that ∂T is not necessarily equal to $\mathcal{S}$, as the extreme points of the convex hull of a set need not equal that set.

Corollary 4. *For any $\mathcal{S}$, ∂T is non-discriminatory if and only if, for every A, W_A is affine (linear).*[10]

To interpret Corollary 4, recall that the "concavification" of a function is a crucial idea in Bayesian persuasion. The idea is, roughly speaking, that optimal information structures will place probability zero on signals that can be obtained as the mean of other signals giving a higher expected payoff.[11] In our environment, an optimal information structure can be chosen to have support on the extreme points of the

[10]Because the domain of W_A is a set of probability measures, W_A is linear if it is affine. In fact, in this case we have $W_A(s) = \sum_{\theta \in \Theta} \alpha_A(\theta) s(\theta)$, where α_A is as in Proposition 3.

[11]At this point, we can further formalize the discussion in Section 2.2. Following Blackwell (1953), for $\pi, \pi' \in \Delta(\mathcal{S})$, we say that π is *more informative* than π' if, for every A, $\int v_A(t) d\pi(t) \geq \int v_A(t) d\pi'(t)$.

convex hull of $\mathcal{S}$ due to the convexity of v_A. Moreover, W_A is always weakly concave (the concavification in question).

Strict concavity of W_A reflects a situation where the firm strictly prefers the optimal value for a skill distribution rather than the expected optimal value for a probability over skill distributions with the same expectation (Kamenica and Gentzkow, 2011). This is a firm who prefers no uncertainty about the skill distribution (supposing they can observe the skill distribution before choosing an information structure). Hence, it reflects a kind of "risk aversion" of the firm in the overall skill distribution. Affinity therefore reflects risk-neutrality of the firm in the skill distribution.

Going back to Corollary 4, it says that discrimination is possible exactly when W_A exhibits strict concavities. In other words, discrimination is impossible exactly when the firm is always risk-neutral in the skill distribution.

Finally, recall our motivating Phelpsian example. There, discrimination was present even though $\mathcal{S}$ consisted of the extreme points of its convex hull T, and thus $\mathcal{S}$ was maximally informative. Phelps' original point can thus be refined: discrimination obtains because an employer has "different" information about two classes of individuals, rather than better information.

Let us see how this manifests itself in the choice of optimal information structure. In this case, for each $s \in \mathcal{S}$, we have $W_A(s) = v_A(s)$, as each s is extreme in the convex hull of T. We therefore obtain: $(2/3)W_A(1/2, 1/2, 0) + (1/3)W_A(0, 0, 1) = \frac{4}{3} < \frac{3}{2} = (1/3)v_A(1, 0, 0) + (2/3)v_A(0, 1/2, 1/2) \leq W_A(1/3, 1/3, 1/3)$. Hence, in our example, W_A is nonlinear. This is a general artifact of non-identification and discrimination, as is evidenced by Corollary 4.

4.4. **On Phelps, Agner and Cain.** We read Phelps through the eyes of Aigner and Cain (1977). What we term Phelpsian statistical discrimination is really how Aigner and Cain interpret the term; in particular in their "alternative model." They adopt Phelps' model of normally distributed signals, and his notion of informativeness captured by the variance of a noise term in the signaling technology. Aigner and Cain then consider a risk-averse firm, and, roughly, make the point that a less informative signaling structure will expose the firm to more risk than a more informative signaling structure. Hence a population that is associated with a less informative signaling structure will be worth less to a risk averse firm than one who is associated with a more informative signals — even when the underlying skill distribution is the same.

In our model, risk aversion is not the reason informativeness matters. Rather, it matters because the firm uses information to match a worker with its technology. So the channel is different from Aigner and Cain's model, but follows from similar mathematical ideas. Our model is also quite general, and not restricted to the normal-linear model in Phelps. Finally, the connection to identification and Bayesian persuasion is completely new.

5. Conclusion

We have formulated Phelps' theory of statistical discrimination using the modern language of information design. Our results shed new light on the nature of discrimination, and on some of the empirical approaches one might take to establish the existence of statistical discrimination.

Statistical discrimination turns out to be equivalent to the absence of econometric identification of signals from skills. While the identification of skills from salaries is always possible, even in the presence of discrimination, we show that the crucial identification property is that of signals from skills.

In second place, we connect discrimination with the source of worker remunerations. We show that identification is possible if and only if remunerations are linear in fair skill-dependent, signal-independent, payoffs. In this situation, fair payoffs admit an interpretation as the lowest expected payments that respects certain workers outside options in the market. The relevant outside options turn out to be the counterfactual wages they would have earned with different signals.

Our results have immediate consequences for empirical research on discrimination. They imply that discrimination is absent if and only if empirical approaches to linearly estimating fair skills-based payoffs are viable.

Finally, our model is static. The dynamic implications of information being revealed over time may have implications for our results and should be the focus of future studies.

6. Proofs

Proposition 5 is a formal statement of Proposition 1 (in fact a substantially stronger statement).

For a technology $A = \{a_1, \ldots, a_n\}$, and $k \in \mathbf{R}^{\Theta}$, let $A + k = \{a_1 + k, \ldots, a_n + k\}$ be the technology shifted by k.

Proposition 5. *For any two populations any $\mathcal{S}$ and any finite technology A, if $\pi, \pi' \in \Delta(\mathcal{S})$ are such that $p_\pi \neq p_{\pi'}$, then there is k for which*

$$\int_T v_{A+k}(s) d\pi(s) \neq \int_T v_{A+k}(s) d\pi'(s).$$

Proof. Observe that for any A and any element $l \in A$, we have $v_{A+l}(s) = v_A(s) + l \cdot s$. Now, since $p_\pi \neq p_{\pi'}$, there is l for which $l \cdot p_\pi \neq l \cdot p_{\pi'}$. Consequently, there is α for which:

$$\alpha l \cdot (p_\pi - p_{\pi'}) \neq \int_{\mathcal{S}} v_A(s) d\pi'(s) - \int_{\mathcal{S}} v_A(s) d\pi(s).$$

Let $k = \alpha l$, and conclude that:

$$\int_{\mathcal{S}} v_{A+k}(s) d\pi(s) = k \cdot p_\pi + \int_{\mathcal{S}} v_A(s) d\pi(s) \neq k \cdot p_{\pi'} + \int_{\mathcal{S}} v_A(s) d\pi'(s) = \int_{\mathcal{S}} v_{A+k}(s) d\pi'(s).$$

□

6.1. **A general result.** Our main result follows from Theorem 6 below. We need two additional definitions involving technologies A that are binary.[12] Say that $\mathcal{S}$

[12] A set is *binary* if it has one or two elements.

- *admits fair valuations for binary sets* if for any binary subset $A \subseteq \mathbf{R}^\Theta$, there is $\alpha_A \in \mathbf{R}^\Theta$ for which for all $s \in \mathcal{S}$, $v_A(s) = \sum_\theta \alpha_A(\theta)s(\theta)$, and
- is *non-discriminatory for binary sets* if for any $\pi, \pi' \in \Delta(\mathcal{S})$ and any binary $A \subseteq \mathbf{R}^\Theta$, if $p_\pi = p_{\pi'}$, then

$$\int_{\mathcal{S}} v_A(s) d\pi(s) = \int_{\mathcal{S}} v_A(s) d\pi'(s).$$

Theorem 6. *The following are equivalent.*

(1) $\mathcal{S}$ *is non-discriminatory.*
(2) $\mathcal{S}$ *is non-discriminatory for binary sets.*
(3) $\mathcal{S}$ *is identified.*
(4) $\mathcal{S}$ *admits fair valuations.*
(5) $\mathcal{S}$ *admits fair valuations for binary sets.*

The proof of Theorem 6 is in 6.3. First, we discuss some preparatory results.

6.2. **Additional notation.** Let T be the closed convex hull of $\mathcal{S}$. Recall that ∂T denotes the extreme points of T. The definition of v_A extends to T. Let $Y_A : T \to \mathbf{R}$ be the concave envelope of v_A, defined as the pointwise infimum of the affine functions that dominate v_A. So if $\mathcal{A}(T)$ denotes the space of all affine functions on T, then $Y_A(t) = \inf\{l(t) : l \in \mathcal{A}(T) \text{ and } v_A \le l\}$. Recall the definition of W_A from Section 4.3.

In the Bayesian persuasion literature, Y_A is the concavification of v_A, and our next result essentially establishes the Bayesian persuasion result.

Lemma 7. $Y_A = W_A$

Proof. Let $l : T \to \mathbf{R}$ be an affine function and $v_A \le l$. For any $\pi \in \Delta(T)$ with $\int_T q d\pi(q) = p$,

$$\int_T v_A(q) d\pi(q) \le \int_T l(q) d\pi(q) = l\left(\int_T q d\pi(q)\right) = l(p),$$

as l is affine. Thus $W_A \le l$, as π was arbitrary. This implies that $W_A \le Y_A$, as l was arbitrary.

Now suppose that $W_A(p) < Y_A(p)$. Recall that W_A is concave. Then the set $D = \{(q, y) \in T \times \mathbf{R} : y \leq W_A(q)\}$ is closed and convex, so there exists α with $(q, y) \cdot \alpha \leq (p, W_A(p)) \cdot \alpha < (p, y') \cdot \alpha$ for all $(q, y) \in D$ and all $y' \geq Y_A(p)$. Write $\alpha = (\alpha^1, \alpha^2) \in \mathbf{R}^{\Theta} \times \mathbf{R}$. Clearly we cannot have $\alpha^2 = 0$ as $(p, W_A(p) \in D$. Consider the affine function $l : T \to \mathbf{R}$ defined by

$$q \mapsto (1/\alpha^2)((p, W_A(p)) \cdot \alpha - \alpha^1 \cdot q).$$

This means that $l(p) = W_A(p) < Y_A(p)$. Moreover, for any $q \in T$, $\alpha \cdot (q, W_A(q)) \leq \alpha \cdot (p, W_A(p))$; hence,

$$l(q) = (1/\alpha^2)\alpha^1 \cdot p + W_A(p) - (1/\alpha^2)\alpha^1 \cdot q \geq W_A(q) \geq v_A(q),$$

where the last inequality follows from the definition of W_A. Then $l \in \mathcal{A}(T)$, $v_A \leq l$, and $l(p) < Y_A(p)$; a contradiction. □

6.3. **Proof of Theorem 6.** By the Choquet-Meyer Theorem (Theorem II.3.7 in Alfsen (2012) or p. 56-57 in Phelps (2000)), T is a simplex iff ∂T is identified.

Now, to prove the theorem: it is obvious that $3 \implies 1 \implies 2$. We shall prove that $2 \implies 3$. To this end, let $\mathcal{S}$ be non-discriminatory for binary menus. The proof that $2 \implies 3$ has two parts. The first is to show that $\mathcal{S} = \partial T$. The second is that T must be a simplex.

First, it is obvious by definition of T that $\partial T \subseteq \mathcal{S}$. So we prove that $\mathcal{S} \subseteq \partial T$. To this end, suppose by means of contradiction that there is $s^* \in \mathcal{S}$ for which there are $t, t' \in T$, $t \neq t'$, and $\gamma \in (0, 1)$ for which $s^* = \gamma t + (1 - \gamma)t'$. Let $f = (s^* - t) + [t \cdot s^* - s^* \cdot s^*]\mathbf{1}$ and $g = -f$. Observe that $f \cdot s^* = 0$, $g \cdot t = -t \cdot (s^* - t) - s^* \cdot (t - s^*) > 0$ and $f \cdot t' = (s^* - t) \cdot (t' - s^*) = \gamma(1 - \gamma)(t' - s^*) \cdot (t' - s^*) > 0$.

Let $A \equiv \{f, g\}$. Then we obtain that $v_A(t) \geq g \cdot t > 0$, $v_A(t') \geq f \cdot t' > 0$, while $v_A(s^*) = 0$ (as $f \cdot s^* = g \cdot s^* = 0$).

Now, for each of t, t', there are finitely supported (by Caratheodory's theorem) π_t and $\pi_{t'}$ on ∂T (so in particular on $\mathcal{S}$) for which $t = \int_{\mathcal{S}} s d\pi_t(s)$ and $t' = \int_{\mathcal{S}} s d\pi_{t'}(s)$. This means that $\int_{\mathcal{S}} v_A(s) d\pi_t(s) \geq$

$v_A(t) > 0$ and $\int_{\mathcal{S}} v_A(s) d\pi_{t'}(s) \geq v_A(t') > 0$, as v_A is convex. Then

$$\int_{\mathcal{S}} v_A(s) d(\gamma\pi_t + (1-\gamma)\pi_{t'})(s) > 0.$$

But this contradicts 2 as $\int_{\mathcal{S}} sd(\gamma\pi_t + (1-\gamma)\pi_{t'})(s) = \gamma t + (1-\gamma)t' = s^*$, and $v_A(s^*) = 0$.

So we have shown that $\mathcal{S} = \partial T$, and we turn to the proof that T is a simplex (and thus $\mathcal{S}$ identified). By Alfsen (2012) Theorem II.4.1, since T is convex and compact, T is a simplex if and only if $\mathcal{A}(T)$ forms a lattice in the usual (pointwise) ordering on functions. So, suppose by means of contradiction that $\mathcal{A}(T)$ does not form a lattice. Then, there are $f, g \in \mathcal{A}(T)$ which possess no supremum in $\mathcal{A}(T)$.

Lemma 8. *Let $f, g \in \mathcal{A}(T)$. For any $z \in \partial T$, if $f(z) \geq g(z)$, then there is $h \in \mathcal{A}(T)$ for which $h \geq f, g$ and $h(z) = f(z)$.*

Proof. Let M be the subgraph of the concave envelope of $v_{\{f,g\}}$. Observe by definition that it is the comprehensive, convex hull of the points $\{(z, v_{\{f,g\}}) : z \in \mathcal{S}\}$, so that it is polyhedral (Corollary 19.I.2 of Rockafellar (1970)). Therefore, by definition of polyhedral concave function, there is h supporting it at $(z, f(z))$. □

From Lemma 8, and the fact that f and g possess no supremum in $\mathcal{A}(T)$, it follows that there is no affine function h for which for all $z \in \partial T$, $h(z) = \max\{f(z), g(z)\}$. Consequently, if $A \equiv \{f, g\}$, then Y_A is not affine, since for all $z \in \partial T$, it follows that $Y_A(z) = \max\{f(z), g(z)\} = v_A(z)$. Now, Y_A being concave and not affine means that there is $\hat{\pi} \in \Delta(T)$ with $\int_T Y_A(q) d\hat{\pi}(q) < Y_A(p_{\hat{\pi}})$. Since $\mathcal{S} = \partial T$, and Y_A is concave, we can in fact find (by Lemma 4.1 in Phelps (2000)) $\pi \in \Delta(\mathcal{S})$ with $p_{\hat{\pi}} = p_\pi$ and

$$\int_{\mathcal{S}} v_A(q) d\pi(q) = \int_{\mathcal{S}} Y_A(q) d\pi(q) \leq \int_T Y_A(q) d\hat{\pi}(q) < Y_A(p_{\hat{\pi}}) = Y_A(p_\pi),$$

where the first equality follows from $v_A(q) = Y_A(q)$ for $q \in \mathcal{S}$, and the second inequality from the choice of π.

Now, by Lemma 7, $Y_A(p_\pi) = \sup\{\int v_A(q) d\tilde{\pi}(q) : \tilde{\pi} \in \Delta(T) \text{ and } p_{\tilde{\pi}} = p_\pi\}$. Then there is $\pi' \in \Delta(\mathcal{S})$ (as the sup is achieved for a measure with support in $\partial T = \mathcal{S}$) with $p_\pi = p_{\pi'}$ and $\int_{\mathcal{S}} v_A(q)d\pi(q) < \int_{\mathcal{S}} v_A(q)d\pi'(q)$, contradicting the fact that $\mathcal{S}$ is non-discriminatory for binary menus.

Now, we have shown that 1,2, and 3 are equivalent, and that they are all equivalent to $\mathcal{S}$ forming the vertices of a simplex. To see that these are equivalent to 4 and 5, recall from above that T is a simplex if and only if $\mathcal{A}(T)$ forms a lattice. In particular, it is enough to check that for any $f, g \in \mathcal{A}(T)$, $f \vee g \in \mathcal{A}(T)$ exists.[13]

So, $\mathcal{A}(T)$ forms a lattice iff for any $f, g \in \mathcal{A}(T)$, the smallest concave function dominating both f and g is affine.

We show that the property of $\mathcal{A}(T)$ being a lattice and $\mathcal{S} = \partial T$ is equivalent to the property that for all A binary (or for all A finite) $\int_{\mathcal{S}} v_A(s)d\pi(s) = \alpha_A \cdot p_\pi$ for some α_A. So suppose $\mathcal{A}(T)$ is a lattice; and let A be arbitrary. Observe that $\int_{\mathcal{S}} v_A(s)d\pi(s) = \int_{\mathcal{S}} W_A(s)d\pi(s) = \int_{\mathcal{S}} Y_A(s)d\pi(s)$, where the first equality occurs as $\mathcal{S} = \partial T$ and the second by Lemma 7. But by definition of Y_A and the fact that $\mathcal{A}(T)$ is a lattice, we therefore have $Y_A \in \mathcal{A}(T)$, which is enough to prove fair valuation for binary acts. It also establishes fair valuations for arbitrary sets of acts.

Conversely suppose fair valuations for binary sets of acts. First, we claim that $\mathcal{S} = \partial T$ where T is the convex hull of $\mathcal{S}$. If not, then by Caratheodory, there is $s^* \in \mathcal{S}$ and finitely supported $\pi \in \Delta(\mathcal{S})$ for which $s^* = \int_{\mathcal{S}\setminus\{s^*\}} s d\pi(s)$. Let T^* be the support of π, and let f be affine so that $f(s^*) = 0$, while $f(s) \neq 0$ for all $s \in T^*$. Then for $A = \{f, -f\}$, the fair valuations property is not satisfied, since $v_A(s^*) = 0$ whereas for all $s \in T^*$, $v_A(s) > 0$, so that $\int_{\mathcal{S}} v_A(s)d\pi(s) > 0$.

Next, we claim that $\mathcal{A}(T)$ is a lattice. To see this, let $f, g \in \mathcal{A}(T)$. Observe that each of f, g can be induced by a member of $\mathbf{R}^\Theta$, so without loss, assume that $f, g \in \mathbf{R}^\Theta$, and let $A = \{f, g\}$; let α_A be the member of $\mathbf{R}^\Theta$ guaranteed by definition of fair valuations for binary acts. Now, the smallest concave function dominating f, g on T is, by definition, $Y_A(t)$, and by Lemma 7, we have $Y_A(t) = W_A(t)$. Further, $t \in T$ is induced by a probability distribution $\pi \in \Delta(\mathcal{S})$, so that $t = \int_{\mathcal{S}} s d\pi(s)$.

[13] Because then $f \wedge g$ will also necessarily exist; as $f \wedge g = -((-f) \vee (-g))$.

Let us assume that π is chosen so that $W_A(t) = \int_{\mathcal{S}} v_A(s)d\pi(s)$. Now, we know that for each $s \in \mathcal{S}$, $v_A(s) = W_A(s)$, because $\mathcal{S} = \partial T$. So, we have $W_A(t) = \int_{\mathcal{S}} v_A(s)d\pi(s) = \int_{\mathcal{S}} \sum_\theta \alpha_A(\theta)s(\theta)d\pi(s)$. And $\int_{\mathcal{S}} \sum_\theta \alpha_A(\theta)s(\theta)d\pi(s) = \sum_\theta \alpha_A(\theta)t(\theta)$. So, we conclude $Y_A(t) = W_A(t) = \sum_\theta \alpha_A(\theta)t(\theta)$, which is what we wanted to show.

6.4. **Proof of Proposition 3.** The Lagrangian for the maximization problem in the definition of W_A is

$$\begin{aligned} L(\pi, \lambda) &= \int_T v_A(t)d\pi(t) + \lambda \cdot \left[p - \int_T q d\pi(q)\right] \\ &= \lambda \cdot p + \int_T (v_A(t) - \lambda \cdot p)d\pi(t) \end{aligned}$$

and apply the maximin theorem (see for example Theorem 6.2.7 in Aubin and Ekeland (2006), which applies here because $\Delta(T)$ is compact).

6.5. **Proof of Corollary 4.** By the Choquet-Meyers Theorem (Theorem II.3.7 in Alfsen (2012)) T is a simplex iff the concave envelope of every lower semicontinuous and convex function is affine. Clearly, when $\mathcal{S}$ is identified, T is a simplex, and since v_A is convex and lower semicontinuous, we obtain that $W_A = Y_A$, the concave envelope. So W_A is affine.

Conversely, suppose that W_A is affine for each finite A. We will show that T is a simplex (so that ∂T forms the vertices of a simplex, and is identified). But this again follows from the fact that W_A is the smallest concave function on T dominating each $a \in A$. Since it is affine, it follows that $\mathcal{A}(T)$ is a lattice, and hence T is a simplex.

References

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

Alfsen, E. M. (2012): *Compact convex sets and boundary integrals*, vol. 57, Springer Science & Business Media.

ARROW, K. J. (1971): "Some models of Racial Discrimination in the Labor Market," Tech. Rep. RM-6253-RC, RAND.

——— (1973): *The Theory of Discrimination. S. 3–33 in: O. Ashenfelter/A. Rees (Hrsg.), Discrimination in Labor Markets*, Princeton University Press.

AUBIN, J.-P. AND I. EKELAND (2006): *Applied nonlinear analysis*, Courier Corporation.

BECKER, G. S. (1957): *The Theory of Discrimination*, University of Chicago Press.

BLACKWELL, D. (1953): "Equivalent comparisons of experiments," *The annals of mathematical statistics*, 265–272.

HAAVELMO, T. (1944): "The probability approach in econometrics," *Econometrica*, 12, iii–115.

KAMENICA, E. AND M. GENTZKOW (2011): "Bayesian persuasion," *American Economic Review*, 101, 2590–2615.

MACHINA, M. J. (1984): "Temporal risk and the nature of induced preferences," *Journal of Economic Theory*, 33, 199–231.

PHELPS, E. S. (1972): "The Statistical Theory of Racism and Sexism," *American Economic Review*, 62, 659–661.

PHELPS, R. R. (2000): *Lectures on Choquet's theorem, second edition*, Springer Science & Business Media.

ROCKAFELLAR, R. T. (1970): *Convex analysis*, Princeton university press.