Stereotype Threat and Counter-Stereotypical Behavior.

Richard Chisik*

April 2009

ABSTRACT: We show that stereotype threat can occur in a labor market signaling model where no one believes that an agent played a strictly dominated strategy. This idea of self-fulfilling statistical discrimination is different from the one Spence posed in his original work on market signaling, which requires employers to believe that low-ability women would play a strictly dominated strategy. Our analysis builds instead on an endogenous quality choice. The existence of multiple un-dominated pooling equilibria, which can generate a stereotype threat effect, is shown to depend on the shape and variance of the distribution of types as well as the value of the signal. It is more likely if the variance is low (so that the types are more similar) or if the signal has more value to the firm. We also show that a very bad stereotype forces the high-quality agent with that bad stereotype to separate. This counter-stereotypical behavior increases that label's high-quality probability and education level. In this way the very discriminated against label overtakes the complacent good stereotype label, and the good stereotype label can suffer from a reputational Dutch disease.

JEL Classification: J70, D82, O15

---

*Department of Economics DM-309C, Florida International University, Miami, FL, 33199; Phone: (305) 348-3286; Fax: (305) 348-1524, E-mail: chisikr@fiu.edu.

I. Introduction

Perceptions can have an important self-fulfilling effect on outcomes. For example, simply reminding test takers of their race before a test can affect results on standardized tests (Steele and Aronson, 1995). This stereotype threat has also been demonstrated with respect to gender and entrepreneurship by Gupta et al. (2008), and gender and math performance by Good (2008) and Kiefer (2007). The self-fulfilling prophecy was also demonstrated by Rosenthal and Jacobson (1968) who analyzed how completely random information about student ability that was told to teachers at the beginning of the school-year soon became self-fulfilling. This Pygmalion effect can also be produced in the workplace as well as in the classroom (Eden, 1992, and Dvir et al., 1995).

We approach these psychology studies with the idea that people differ in several dimensions, some of which are un-observable. One thing that is observable is an unalterable marking that we call a label. The label refers to the agent's caste, ethnic origin, eye color, gender, or race. The experiments of Bodenhauser and Wyer (1985) and Bodenhauser and Lichtenstein (1987) on stereotypes suggests that when presented with a large amount of complex information, people use stereotypes as a heuristic in forming an impression of the labeled person, without considering their other attributes.[1]

We consider how label specific stereotypes interact with an agent's ability to signal their unknown type and chosen quality and we show that this heuristic process can become self-fulfilling. Although the label is not correlated with the unknown parameters in any fundamental way the incentive effects of the stereotype correctly correlates them in equilibrium. The label then provides statistical information to the firm who engages in statistical discrimination. Still, there are important facts about gender and raced based differences in education that stereotype threat cannot explain by itself. For example, 58% of bachelor's degrees in the U.S. are given to women. In addition, the return on a college education is much higher for African-Americans than for whites (Neil and Johnson, 1996) even though their graduation rate is lower. In this paper we show how this counter-stereotypical behavior can exist along with stereotype threat.

---

[1] If the label is, in fact, correlated with the complex information, then this heuristic process can be considered a form of bounded rationality (Simon, 1957).

We employ the following structure. An agent is endowed with a label and an ability. The ability is the agent's private information. At a young age, the agent can choose to become high quality. It is costly to become high quality and this cost depends on the agent's ability. The quality is also unobservable to the firm when they hire the worker, however, the agent may use an additional signal (such as education) to inform the firm of their chosen quality. The only reason the agent would choose high quality is because it lowers the cost of the additional signal.

In analyzing this framework we consider sequential equilibrium, however, as is well known, they do not restrict beliefs off of the equilibrium path and many sequential equilibrium are not reasonable. In particular, it is necessary to rule out certain equilibria which are equilibria only if some agents have an incredible out-of-equilibrium belief. We, therefore, define an equilibrium as a sequential equilibrium that satisfies a credibility requirement on beliefs. Although there is a vast literature on what constitutes a non-credible belief, the results in this paper require no more than the simplest, and least contentious, belief refinement. In particular, a credible posterior belief, at any information set, places zero probability on the event that an agent played a strictly dominated strategy. This dominance refinement does not rule out pooling equilibria that Pareto dominate the unique un-dominated separating equilibrium and we analyze both types of un-dominated equilibria.[2]

Any stereotype about the label is irrelevant in the unique separating equilibrium. If the distribution of types is the same for all labels, then, in a separating equilibrium, all labels have the same probability of a high-quality agent. This is not necessarily the case in pooling equilibria. If the distributions of types satisfies certain conditions (we provide a general analysis as well as solve a specific

---

[2] Some of the literature utilizes more powerful belief refinements, such as Cho and Kreps' (1987) iterated-intuitive criterion to rule out all pooling equilibria in games with two qualities or types and Banks and Sobel's (1987) universal divinity to rule out all pooling equilibria in games with any finite number of types. When quality is chosen endogenously, as in this paper, the pooling equilibria have some interesting properties. Mailath et al. (1993) also take issue with belief refinements that rule out Pareto dominating pooling equilibria. They introduce undefeated equilibrium, which preserves these pooling equilibria. The Mailath et al. equilibrium concept is motivated not only by the desire to preserve Pareto dominating pooling equilibria but also to avoid the Stiglitz critique of forward induction refinements. Our weaker dominance refinement is also not subject to this Stiglitz critique. The perfect sequential equilibria of Grossman and Perry (1986) can also be used to justify a deviation to a Pareto dominating pooling equilibrium. Unfortunately, their credible belief restriction also eliminates all pooling equilibria; therefore, even in simple signaling models, if the percentage of high quality types is large (so that pooling equilibria Pareto dominate the separating equilibria), then a perfect sequential equilibrium may fail to exist.

example that makes use of the Pareto distribution), then multiple un-dominated pooling equilibria exist. In this case as the required pooling signal is increased a higher measure of types will choose high-quality. Hence, for one label there exists a pooling equilibrium with a low stereotype and a low pooling expenditure and for another label there exists a pooling equilibrium with a high stereotype and a high pooling expenditure. The first contribution of our analysis is showing that these un-dominated, self-fulfilling, statistical discrimination equilibria are more likely to occur if there is less variance in the distribution of types, or if the signal is more valuable to the firm.

This form of statistical discrimination is different from the one Spence (1973) posed in his seminal work on market signaling. In his paper, if separating beliefs require a much higher level of the expenditure for women than for men, women may pool (at a low level) while men separate. Although provocative, his scenario requires employers to believe that low-quality women would play a strictly dominated strategy. Hence, when beliefs are subject to a reasonable refinement, statistical discrimination in a signaling model requires agents to have an endogenous quality (or participation) choice. This clarification on how statistical discrimination could occur in a labor market signaling model is our second contribution.

A very low pooling stereotype implies a very low pooling wage so that a high-quality agent would separate and break this low stereotype (and dominated) pooling equilibrium. Hence, a very negative stereotype cannot be part of an equilibrium. A label that suffers from a very low stereotype will, ex post, end up in the separating equilibrium which has a higher probability of a high-quality worker. The following situation is then possible. A high-quality agent with a moderate or high stereotype label will pool at a moderate or high level of the signal and a high quality agent with a low stereotype label will engage in counter-stereotypical behavior and separate themselves with an even higher level of the signal. This counter-stereotypical behavior may provide a partial explanation for why women have better education attainment than men and why the return to education is higher for African-Americans than for whites. Furthermore the good stereotype can generate complacency and a reputational Dutch disease effect, whereby a worker with a positive stereotype label has less reason to distinguish themselves and they remain content in the pooling equilibrium.[3] Identification of this counter-stereotypical behavior and the reputational Dutch disease is our third contribution.[4]

This paper follows a highly distinguished literature on the theory of discrimination. The pioneering work on discrimination in an economic environment is Becker (1957). He suggests that some agents may have a taste for discrimination and be willing to pay extra to satisfy their preferences. Arrow (1973) also explores this idea and notes that non-discriminators will thrive by hiring the equally talented discriminated-against labels at a lower wage. The resulting equilibrium entails segregation but wage disparity among labels is decreased.

Arrow (1973) recognizes this limitation and suggests an alternative model where employers harbor differing views about the productivity of differing labels. In this framework, imperfect observability about a worker's productivity leads employers to offer wages based in part on stereotypes about the worker's label. When the expected wage affects the productivity decision, negative or positive stereotypes can become self-fulfilling. A somewhat similar idea is explored by Phelps (1972), Aigner and Cain (1977), and Lundberg and Startz (1983); however, their framework requires that employers have less precise measurements of ability and productivity for discriminated against labels. Akerlof (1976) suggests that social taboos about caste interaction which are supported by strict punishments can lead to caste segregation, however, his model does not explain why this segregation persists when punishments for violating these social taboos are no longer in place.

Coate and Loury (1993) build on Arrow's model to study the effect of affirmative action policies in a statistical discrimination framework. Their paper differs from the above papers in that discrimination against certain labels can affect a worker's job assignment. Looking at affirmative action in a statistical discrimination framework allows for the analysis of this policy's joint effect on worker incentives and employer beliefs. Milgrom and Oster (1987) show that the desire to hide what they have learned about a

---

[3] The Dutch disease is often considered an international economic phenomenon, whereby the initial good luck of a natural resource find generates lower long-run growth potential.

[4] Feltovich et al. (2002) consider counter-signaling. In their model, with three types and noisy additional information on quality, the middle type separates but the highest type's expected payoff is greater if they rely on the additional information and do not choose a separating signal level.

worker's ability will lead employers to hide some workers with certain labels in less visible job assignments. Coate and Tennyson (1992) make the extension from job assignment to sector choice. They show that discrimination in credit markets may lessen the likelihood of self employment among discriminated against labels. Fryer (2007) extends Arrow (1973) and Coate and Loury (1993) to a dynamic framework. His key insight is that the first period hiring decision truncates the distribution of workers differently for differing labels. Even though the groups are ex-ante identical, the hired workers from the label with a low self-fulfilling stereotype are of higher potential quality. If the unobservable potential quality that influences the worker investment necessary for being hired is sufficiently correlated with that necessary for being promoted, then there may be belief flipping so that the hired workers from the group with a low stereotype are more likely to receive a promotion in the second period.

In the next section we analyze a simple parametric example, and sketch the results in order to better explain the underlying intuition. In the third section we consider a more general example and provide a more formal analysis. Conclusions are in the fourth section.

II. Two Simple Examples

We start by considering two examples based on Kreps' (1990) version of Spence's (1973) education signaling model. In the first example the signal (education) is dissipative and has no value other than its ability to convey information. In the second example the signal is non-dissipative.

A. A Dissipative Signal Model

There is one worker and one firm. The worker may be low or high quality (q) and this quality is the worker's private information. High quality workers are more productive and have a value to the firm of $v^h = 2$. A low quality worker's value to the firm is $v^l = 1$. The firm's payoff function is a quadratic loss function that squares the difference between the workers expected value (v^E) and the wage (w) that they offer to the worker: $\pi = -(v^E - w)^2$. The firm, therefore, offers the worker their expected value: $w = v^E$.[5]

[5] We could also assume that there are at least two firms who bid for the worker and we would obtain all of the same results, however, we would need to modify the model so that there is more than one receiver of the signal. Either formulation ensures that the worker has monopoly power. Giving the worker monopoly power yields a straightforward division of the gains from trade. Other specifications are, of course, possible, however, as long as

Although the worker's quality is unobservable to the firm, the worker can choose a perfectly observable level of education (e) that may signal their quality. The wage would then be a function of education: $w(e)$. The cost of education depends on the unobservable quality and is higher for lower quality workers: $c(q, e) = e/v^q$. Although quality is unobservable, a high quality worker has a lower marginal cost of obtaining an education which may provide information about their chosen quality.

We add to this canonical model in the following way. The worker has an unobservable type θ that determines their additional cost of becoming high quality: $I(\theta,q) = (\theta - 1)(v^q - 1)$, where $(v^q - 1)$ is an indicator function that is equal to one only if the worker is high quality. The types are given by a Pareto distribution with location 1 and shape parameter k so that for $\hat{\theta} \geq 1$ the $\Pr(\theta \leq \hat{\theta}) = F(\hat{\theta}) = 1 - (1/\hat{\theta})^k$.[6] For this distribution $\theta_{min} = 1$ and $I(\theta_{min}, h) = 0$. The idea is that high quality is a lifetime decision and these costs are captured by the investment cost $I(\theta, q)$. For some types it is very easy to become high quality and for others it is quite costly. We can then define a marginal type θ^H that is just indifferent between choosing high and low quality. All θ greater than θ^H will choose low quality and all less than θ^H will become high quality. The probability of a high quality worker is, therefore, $F(\theta^H)$. Finally, the worker has a label: $i \in \{A, B, C, \ldots\}$. The label may correspond to caste, ethnicity, gender, race, religion, or some other publicly identifiable characteristic. It is common knowledge that the prior distribution of types is independent of the label and that the label has no direct effect on ability, the cost of signaling, or the value of the signal. Still we will see that any belief or stereotype, ρ_i, that is attached to the label may affect the probability of a high-quality worker $F(\theta^H)$. Of course the belief must be correct so that we require $\rho_i = F(\theta^H)$ in equilibrium.

The timing of the model is as follows. First the worker receives their label and their type. The

---

they allow the worker to retain some of the surplus they create by producing high quality, these alternative specifications do not change our results. Furthermore, by assuming that the firm bids for the worker we rule out the possibility of price or wage signaling by the worker. We do not make this assumption to detract from the importance of price signaling, but rather to limit our analysis to a one dimensional signal.

[6] The Pareto distribution was first used by Wilfredo Pareto in modeling income distribution where it is still considered to provide a good fit. We think that it may provide a good approximation of unknown worker types and we use it for this example. Cabral and Mata (2003) have shown that it provides a good fit for the size (and cost) distribution of firms. Since Melitz (2003) it has been used in international trade modeling where the assumption that firms unknown costs are given by a Pareto distribution provides a good fit to the data on the size of exporting and non-exporting firms.

worker then chooses their quality. Next, the worker chooses their education. The firm observes the worker's education and label and makes a wage offer: $w(e, i)$. Hence, the worker first chooses q and then chooses e to maximize their expected utility:

$$u(\theta, q, e, i, w) = w(e, i) - c(q, e) - I(\theta, q) = v^E - e/v^q - (\theta - 1)(v^q - 1). \qquad (2.1)$$

Under perfect information the worker chooses $e = 0$. A high-quality worker earns a wage of 2 and a low-quality worker earns 1. The marginal type θ^H is given by $u(\theta^H, h, 0, i, 2) = 2 - (\theta^H - 1) = 1 = u(\theta, l, 0, i, 1)$ or $\theta^H = 2$. Hence, with perfect information the prior probability of a high quality worker is $\rho_i = F(2) = 1 - (½)^k$.

With incomplete information there are many pooling and separating sequential equilibria (we will define them formally for the more general model in the next section). In a pooling equilibrium, a high and a low-quality agent choose the same education and the firms updated belief is the same as their prior belief. In a separating equilibrium a high and a low-quality agent choose a differing level of education and the firm's posterior beliefs correctly ascribe probability one to the higher education level coming from the higher quality.

A sequential equilibrium does not place adequate structure on beliefs off of the equilibrium path and many of the sequential equilibrium for this model are not reasonable. In particular, it is necessary to rule out certain equilibria which are equilibria only if the firm has an incredible out-of-equilibrium belief. We, therefore, define a reasonable equilibrium as a sequential equilibrium which satisfies a credibility requirement on beliefs. In particular, a credible posterior belief, at any information set, places zero probability on the event that the worker plays a strictly dominated strategy.

Any separating equilibria where the high quality worker chooses more than the lowest-cost separating level requires the firm to believe that a low quality worker would play a strictly dominated strategy. The only reasonable separating equilibrium is, therefore, given by the smallest e^s that satisfies $u(\theta, l, e^s, i, w(e^s)) = 2 - e^s \leq 1 = u(\theta, l, 0, i, w(0))$. Hence, $e^s = 1$, and $u(\theta, h, e^s, i, w(e^s)) = 2 - ½ - (\theta - 1)$, so that $\theta^H = 3/2$. In the unique reasonable separating equilibrium $\rho_i = 1 - (2/3)^k$. The probability of a high-quality worker is lower in the signaling equilibrium than in the full information outcome.

B. Reasonable Self-Fulfilling Statistical Discrimination Equilibria

In a pooling equilibrium the firm learns nothing from the agent's education choice. Still, consistency with the common prior distribution places a restriction on the education levels in a pooling equilibria. In particular, the marginal type, θ^H, is defined as a function of the pooling education level (e^p) by $u(\theta^H, h, e^p, w^p) = w^p - e^p/2 - (\theta^H - 1) = w^p - e^p = u(\theta, l, e^p, w^p)$ or

$$\theta^H = e^p/2 + 1. \qquad (2.2)$$

For each pooling education level there is a unique belief that is consistent with the common prior distribution. Consistency with the common prior distribution implies that the correct stereotype satisfies:

$$\rho_i = F(\theta^H) = 1 - (\tfrac{1}{\theta^H})^k = 1 - (\tfrac{2}{e^p+2})^k . \qquad (2.3)$$

Note that a higher pooling stereotype corresponds to a higher level of education. Using (2.3) the offered wage can be written as

$$w^p = \rho_i\, v^h + (1 - \rho_i)v^l = 2 - (\tfrac{2}{e^p+2})^k . \qquad (2.4)$$

Although there are a continuum of pooling sequential equilibrium, not all of them are reasonable. For a pooling equilibrium to be reasonable it is necessary that neither quality of worker prefers the proposed pooling equilibrium to the reasonable separating equilibrium.[7] For a high quality worker this implies $2 - (\tfrac{2}{e^p+2})^k - e^p/2 - (\theta - 1) \geq 3/2 - (\theta - 1)$ or

$$1/2 - (\tfrac{2}{e^p+2})^k - e^p/2 \geq 0. \qquad (2.5)$$

For a low quality worker this implies $2 - (\tfrac{2}{e^p+2})^k - e^p \geq 1$ or

$$1 - (\tfrac{2}{e^p+2})^k - e^p \geq 0. \qquad (2.6)$$

Note that if $e^p \geq 1$, then neither equation (2.5) or (2.6) can be satisfied, therefore, $e^p < 1 = e^s$. Furthermore, for $e^p < 1$ equation (2.6) is satisfied whenever equation (2.5) is satisfied. Hence, in the

[7] Although one pooling equilibria may Pareto dominate another pooling equilibria they are both reasonable if they both Pareto dominate the separating equilibrium. This occurs because when starting from one equilibrium, the firm may have beliefs about other pooling education levels so that neither quality (nor underlying type) is playing a strictly dominated strategy in the original Pareto dominated pooling equilibrium. On the other hand there is no belief which makes a deviation to the Riley level worthwhile for the low quality worker and such a deviation is strictly dominated for the low quality worker. If the high quality worker prefers the separating equilibrium to the proposed pooling equilibrium, then the firm's reasonable belief must ascribe probability one to the deviating education level coming from a high quality worker so that the high quality worker would make this deviation and break the low-level pooling equilibrium

dissipative signal case we only need to consider the high-quality agent's constraint in equation (2.5). From (2.5) we can define a minimum level of the shape parameter, $k(e)$, such that for all $k \geq k(e)$

$$k \geq k(e) = -\frac{\ln(\frac{1}{2} - \frac{e}{2})}{\ln(1 + \frac{e}{2})} \qquad (2.7)$$

reasonable pooling equilibria exist. The function $k(e)$ is minimized at $e = 0.57$ so that the minimum value of $k(e)$ such that reasonable pooling equilibria exist is $\underline{k} = 6.12$. When $k > \underline{k}$ but near this minimal value reasonable pooling equilibria only exist for a narrow range of e close to 0.57. As k increases the range of reasonable pooling equilibria is enlarged as well. This set of reasonable pooling equilibria education levels is denoted $e(k)$. Examination of equation (2.5) indicates that $e(k)$ is a compact interval whose length is increasing in k. Note from (2.7) that $e = 0$ or $e \geq 1$ could never be members of this set. We summarize the results from this section as result 1.[8]

Result 1: (*i.*) *If* $k > \underline{k}$, *then there exists a compact interval of reasonable pooling equilibria stereotypes and education levels,* e^p, *such that the for every* e^p *there exists a unique stereotype that is consistent with the common prior distribution.* (*ii.*) *The measure of this interval is strictly increasing in k.* (*iii.*) *The endogenously correct stereotype,* ρ_i, *is strictly increasing in* e^p. (*iv.*) *In all reasonable pooling equilibria* $0 < e^p < 1$. *The probability of a high quality agent and their corresponding education level is, therefore, less in all reasonable pooling equilibria than in the reasonable separating equilibrium.*

In figure 1 we graph the information from equations (2.5) and (2.6). On the vertical axis we have the costs and payoffs from each level of e that is on the horizontal axis. The three concave functions are the wage $w(e) = 2 - (\frac{2}{e^p+2})^k$ for values of $k = \{4, 7, 10\}$. The affine functions are the opportunity cost of the pooling equilibrium education level for low and high quality and they also includes the utility foregone in the separating equilibrium: $c(q, e) + u(\theta, q, e^s(q), i, w(e^s))$. As seen in figure 1 reasonable pooling equilibria exist for a compact interval of education levels, $e(k)$, and the length of this interval is increasing in k. If k is low then the only correct belief is the separating equilibrium. The pooling beliefs

[8] In the next section we will carefully describe our assumptions and provide proofs for our propositions. In this section we are interested in illustrating how the model works for a specific example. Still, it seems useful to point out when we feel that a certain result should be noticed and we call them results.

for the *A, B,* and *C* label worker are shown in figure 1 for pooling education levels of $e^{pA} = 2/3$, $e^{pB} = ½$, and $e^{pC} = ¼$ and $k = 7$. Note that for $k = 7$ the pooling belief for the C worker is not reasonable.

The following situation is possible. The firm could have a high reasonable pooling belief about an *A* label worker, and a lower (but still) reasonable pooling belief about a *B* label worker. In this case the *B* worker remains in a lower reasonable equilibrium than the initially identical *A* worker. The *A* label worker not only has a better stereotype, but also obtains more education so that the stereotype is, in fact, correct. This is an example of self-fulfilling statistical discrimination or stereotype threat.

As noted above the measure of reasonable pooling equilibria is increasing in *k.* For $k > 2$, the variance of the Pareto distribution is decreasing in *k*, therefore, multiple reasonable pooling equilibria (and the resulting stereotype threat) are more likely if the variance of the distribution of types is lower.[9] This corollary of result 1 is stated as result 2.

Result 2: *Stereotype threat is more possible, and the measure of endogenously correct stereotypes is larger, if there is less variance in the distribution of types.*

Result 2 indicates that self-fulfilling statistical discrimination is more likely if there is little difference between the types. In this case the label is a useful heuristic device. On the other hand, if there is more dispersion in the types, then the label is less meaningful and stereotype threat is less possible.

C. Counter-Stereotypical Behavior and the Reputational Dutch Disease

A more interesting possibility is that the firm has very low, and unreasonable, pooling beliefs about the *C* worker. In this case, there is also initially discriminatory beliefs against the *C* worker, however, these beliefs are not part of a reasonable equilibrium because the high quality worker would deviate to the separating equilibrium. In this case, the only reasonable equilibrium for the *C* worker is the separating equilibrium. The initially discriminatory beliefs against the *C* worker forced the high quality worker to separate, which because of the higher education requirement generated a higher probability of a

[9] The variance of the Pareto distribution is $[k(\theta_{min})^2]/[(k-2)(k-1)^2]$ which is infinite for $k = 1$ or 2 and is otherwise decreasing in *k*.

high quality *C* worker. Using the example in figure 1 the only reasonable belief for the *C* label worker is the separating equilibrium with $e^s = 1$, therefore, $\rho_C = 1 - (2/3)^7 > \rho_A = 1 - (3/4)^7 > \rho_B = 1 - (4/5)^7$.

Whereas moderately discriminatory beliefs can hurt a group, very discriminatory beliefs can serve as an impetus for the more capable in that group to distinguish themselves and separate. Ex post a higher percentage of the discriminated against group *C* attains a high level of education and they obtain more than the group that does not suffer any initial discrimination. The initially optimistic belief afforded to the *A* group, in a sense, causes a type of reputational "Dutch Disease" which generates complacency and lower education attainment. This discussion is stated as result 3.

Result 3: *All unreasonable pooling beliefs generate a reasonable separating equilibrium with an ex post probability of a high quality agent that is larger than in any reasonable pooling equilibrium. In this way non-reasonable pooling beliefs generate counter-stereotypical behavior. Optimistic reasonable pooling stereotypes can generate a reputational Dutch Disease.*

In Spence's (1977) original model, the discrimination against women requires the firm to believe that a woman is high quality only if she chooses an education level that is much greater than the lowest-cost separating level. This excessive required separating level is not chosen by either quality and both qualities of women pool at the lower level. The firm's unreasonable beliefs are not tested and are correct in a sequential equilibrium, however, they require the firm to believe that a low quality woman would play a strictly dominated strategy. Although Spence's concept of statistical discrimination in signaling models does not survive our dominance refinement, his idea that woman may be required to obtain more education also appears in our framework. The difference is that in the current framework, the more able women do receive this education. Empirically, women do exhibit much larger high school and college graduation rates than men, which fits well with our model (women receive 58% of U.S. bachelor's degrees). Of course, lower ability women, who separate with minimal education levels are worse off in the resulting equilibrium.

It is interesting to consider how the model would change if the high quality cost (or the distribution of types) differed across labels. For example, a group with a low stereotype may also face larger high quality costs because of exogenous factors (such as peer effects, reduced access to quality childhood education, or explicit forms of discrimination). In this case, a group with a low unreasonable pooling stereotype also results in a separating equilibrium, however, a smaller measure of types separate. The probability of a high quality worker could, therefore, be lower in the doubly disadvantaged separating label than in the favored pooling group. Still, the return to education is larger for the separating label than for the label that is in a reasonable pooling equilibrium. This result of our model is a potential explanation for the fact that the return on a college education is much higher for African-Americans than for whites (Neil and Johnson, 1996) even though the college graduation rate is lower. The important prediction that comes from the counter-stereotypical behavior in our model is that labels with a low stereotype would see greater disparities in income distribution and education attainment.[10]

D. Statistical Discrimination and the Reputational Dutch Disease with a Non-Dissipative Signal

In this section we briefly consider non-dissipative education. The firm values education apart from its information aspect and the agent's wage reflects this added value. The marginal value of education is assumed to be positive but diminishing. The square root function captures these properties. The worker's expected wage is then: $w(e, i) = v^E + e^{1/2}$. The worker chooses q, and then e to maximize:

$$u(\theta, q, e, i, w) = w(e, i) - c(q, e) - \mathrm{I}(\theta, q) = v^E + e^{1/2} - e/v^q - (\theta - 1)(v^q - 1). \qquad (2.8)$$

The full information education levels are $e^H = 1$ and $e^L = 1/4$ and the low quality payoff is 5/4. The reasonable separating equilibrium satisfies $2 + (e^s)^{1/2} - e^s = 5/4$ or $e^s = 9/4$. The resulting payoffs are $u(\theta, h, e^s, i, w(e^s)) = 19/8 - (\theta - 1)$ and $u(\theta, l, e^s, i, w(e^s)) = 5/4$, so that $\theta^H = 17/8$ and $\rho_i = 1 - (8/17)^k$.

In a reasonable pooling equilibria the relationship between the marginal type and the education level is again given by $\theta^H = e^p/2 + 1$ and $\rho_i = 1 - (\frac{2}{e^p+2})^k$. The offered wage is now:

$$w^p = 2 + (e^p)^{1/2} - (\tfrac{1}{\theta^H})^k = 2 + (e^p)^{1/2} - (\tfrac{2}{e^p+2})^k . \qquad (2.9)$$

In figure 2 we graph the offered wage for $k = \{5, 7\}$ as well as the opportunity costs of the pooling equilibrium education level for low and high quality:

[10] A further extension of our model is to consider how alternative policies and institutional factors affect the signaling costs for differing labels. We leave these extensions for future research.

$$c(h, e^p) + u(\theta, h, e^s(h), i, w(e^s)) = e^p/2 + 19/8 \quad (2.10)$$

$$c(l, e^p) + u(\theta, l, e^s(l), i, w(e^s)) = e^p + 5/4. \quad (2.11)$$

Note that equation (2.10) is greater than (2.11) for any $e^p < 9/4 = e^s$, therefore, as in the dissipative case, we only need to consider the high quality constraint. The results in the non-dissipative case are very similar to the dissipative case. The main difference is that in the non-dissipative case $\underline{k} = \min k(e) = \min \{-\frac{\ln(-\frac{3}{8} + \sqrt{e} - \frac{e}{2})}{\ln(1 + \frac{e}{2})}\} = 4.1$ is lower than in the dissipative case and that for any $k > \underline{k}$ the measure of the compact interval of correct reasonable pooling beliefs and education levels is larger. This example suggests that multiple reasonable pooling equilibria are more likely if the signal has more value to the firm (so that there is more benefit for being high quality), however, it depends on the functional form.

III. A More General Model

A. The Economic Environment

The worker begins by choosing their unobservable quality: $q \in \{l, h\}$. High quality requires an additional lifetime expenditure that depends on the worker's type: $\theta \geq \underline{\theta} \geq 0$. The worker has private information about their type, which is drawn from a continuously differentiable and commonly known log-concave distribution function $F(\theta)$.[11] The worker also has an unalterable label, $i \in \{A, B, C, \ldots\}$, which corresponds to their race, or sex, or caste, etc. It is common knowledge that the label contains no direct information, in the sense that the distribution of types is independent of i: $F(\theta|i) = F(\theta)$ for all i. The label may, however, provide information about the worker's unobservable quality choice. The common stereotype about the worker's label is denoted as $\Pr(h|i) = \rho_i$. After observing their type and label, the worker chooses their unobservable quality, q and an additional expenditure or signal, $s \in \Re^+$, which may enhance their productivity as well as indicate their chosen quality.

Our idea is the following. The worker's desired signal level, s, is related to the stereotype, ρ_i, which is based on the label i. This relationship is captured by the expected wage function. The

[11] Examples of log-concave distribution functions with a non-negative support are the Pareto, uniform, exponential, gamma, chi-squared, power, log normal, and Weibull. The logistic and normal distributions are log-concave, however, their support includes negative numbers.

relationship between the desired signal and the quality choice is given by the worker's disutility of effort or cost function. We show in this section that the key to our main result on multiple self-fulfilling stereotypes is a series of monotonicity assumptions, which ensure that the signal, s, is correlated with the chosen quality, q, and the worker's type, θ.

Consider first the cost function. To simplify the analysis we find it useful to assume that the worker's type does not directly affect their signal cost. We then write the cost function as additively separable $C(\theta, q, s) = I(\theta, q) + c(q, s)$. Whereas the function $I(\theta, q)$ describes the relationship between the worker's type and their cost of becoming high-quality, the function $c(q, s)$ isolates the effect of the quality choice on the signal cost. The unobservable θ indexes the cost of being high quality, an activity that is more costly for higher types: $I_\theta(\cdot, h) > I_\theta(\cdot, l) \geq 0$.[12] These costs naturally vary and a given worker has superior information as to their true cost of quality investments. Hence, we make the assumption that the worker has private information about their specific type. In order to concentrate on the moral hazard in quality choice and avoid issues of the adverse selection of the worker out of the market, we assume that $I(\theta, l) = 0$ for all θ and that low-quality production is profitable under certainty.[13] The additional expenditure, s, admits a variety of interpretations. The main requirement is that it is easier for a high-quality worker to undertake the expenditure, in that a higher q reduces the cost of s. The complementarities between s and q imply that $c_S(l, \cdot) > c_S(h, \cdot) > 0$. Other than its effect on quality choice, type has no additional effect on the cost of the additional expenditure.[14] These restrictions on the cost function are assumption 1:

$$I(\theta, h) \geq I(\theta, l) = 0 \text{ for all } \theta, I_\theta(\cdot, h) > 0; c_S(l, \cdot) > c_S(h, \cdot) > 0 > c_{SS}(q, \cdot) \qquad (3.1)$$

---

[12] We follow the convention that a subscript refers to the partial derivative of the function with respect to the subscript: $\partial I(\cdot, q)/\partial\theta \equiv I_\theta(\cdot, q)$

[13] Weaker restrictions can be made, however, they require assumptions based on endogenously determined values. As long as the expected wage for a low quality worker is greater than the cost for the highest type worker, then our results hold under the more general ordering: $I(\underline{\theta}, h) > I(\underline{\theta}, l)$ and $I_\theta(\cdot, h) > I_\theta(\cdot, l) \geq 0$.

[14] An agent's type only affects the cost of s indirectly through its effect on quality choice. Allowing for θ to directly affect the cost of s would mildly complicate the analysis without altering the results.

Assumption 1 captures the following idea. A worker with a lower type may choose higher quality since it lowers the cost of signaling expenditures. The firm may thus rationally infer that a higher expenditure is associated with higher unobservable quality indicating that the worker is a low type.

Next consider the firm's objective. The firm's revenue function is $v(q, s)$. The firm values higher quality ($v_q(\cdot, s) > 0$) and weakly prefer greater expenditures ($v_s(q, \cdot) \geq 0$). We assume that the marginal profit of s is no lower with a high-quality worker ($v_{qs}(\cdot, \cdot) \geq 0$) and that the firm experiences diminishing gains in revenue as the expenditure is further increased ($v_{ss}(q, \cdot) \leq 0$). If the firm hires the worker, then their expected revenue is: $v^E(s, \rho_i) = E_q[v(q, s)] = b(s, i)\cdot v(h, s) + (1 - b(s, i))\cdot v(l, s)$. The variable $b(s, i) = Pr(h|i, s)$ is the firm's posterior belief that the worker is high quality, conditional on their observation of the agent's label and additional expenditure. The firm's offered wage is chosen to maximize their expected profit function is $\pi = -(v^E - w)^2$. These profit function restrictions are assumption 2.

$$v_q(\cdot, s) > 0, v_s(q, \cdot) \geq 0, v_{qs}(\cdot, \cdot) \geq 0 \geq v_{ss}(q, \cdot), \pi = -(v^E - w)^2. \tag{3.2}$$

The firm observes the label and the signal, but does not observe the worker's type or their unobservable quality choice. Hence, the wage is a function $w(s, i) = v^E(s, i)$. Given their type and label the worker chooses q and then s to maximize their expected utility.

$$Eu(\theta, i) = Eu(\theta, q(\theta, i), s(\theta, i), i) = E[w(s, i)] - C(\theta, q, s). \tag{3.3}$$

To reiterate, the timing of the game is as follows. First the agent receives θ and i. The agent then chooses $q(\theta, i)$ and then $s(\theta, i)$. The firm observes s and i and makes an offer $w(s, i)$. The following assumption 3, along with assumptions 1 and 2 ensures that the agent is hired.

$$v(l, 0) > c(l, 0) \geq 0. \tag{3.4}$$

B. Description of the Equilibrium

A sequential equilibrium (Kreps and Wilson, 1982) is a collection of strategies and beliefs $\{s(\theta, i), q(\theta, i), w(s, i), b(s, i)\}$ which satisfy the following three conditions on actions and beliefs. Sequential rationality for the worker requires:

$$\{q(\theta, i), s(\theta, i)\} \in \text{argmax}_{q,s} E[w - C(\theta, q, s)]. \tag{E1a}$$

Sequential rationality for the firm requires that

$w(s, i) \in \text{argmax}_w - (v^E(s, i) - w)^2$. (E1f)

In defining Bayesian beliefs note that s only depends on θ through the worker's choice of $q(\theta, i)$ and only depends on i through the stereotype, therefore, we can write $s(\theta, i) = s(q(\theta), \rho_i) = s(q, \rho_i)$. Hence, $s(q, \rho_i) = \text{argmax}_s Eu(\theta, q, s, \rho_i)$ so that $s(h, \rho_i)$ is the expenditure chosen by a high-quality worker when the equilibrium stereotype is ρ_i and $s(l, \rho_i)$ is defined similarly. Bayes-consistency of beliefs requires that:

If $\{s(h, \rho_i), \rho_i\} = \{s(l, \rho_i), \rho_i\}$ then $b[s(l, \rho_i), \rho_i] = \rho_i$; (E2bp)

If $\{s(h, \rho_i), \rho_i\} \neq \{s(l, \rho_i), \rho_i\}$ then $b[s(l, \rho_i), \rho_i] = 0$ and $b[s(h, \rho_i), \rho_i] = 1$. (E2bs)

Pooling equilibrium beliefs are described by E2bp, where a high and a low-quality worker chooses the same expenditure and the firm learns nothing from their choice. In a separating equilibrium (E2bs) a high and a low-quality worker choose a differing level of the expenditure and the posterior beliefs recognize these separating expenditure levels.

Because a sequential equilibrium does not place adequate structure on beliefs off of the equilibrium path, an equilibrium shall be defined as a sequential equilibrium which satisfies a credibility requirement on beliefs. We rule out equilibria the firm has a non-credible out-of-equilibrium belief and we label the remaining set as reasonable equilibria.

Although there is a vast literature on what constitutes a non-credible belief, the results in this paper require no more than the simplest, and least contentious, belief refinement. In particular, a credible posterior belief, at any information set, places zero probability on the event that worker played a strictly dominated strategy. For the model in this section this dominance refinement on beliefs requires that:

Define $\xi(s) = \{q(\theta, i) \mid \sim\exists\, s'$ satisfying $\text{Min}_w Eu(\theta, q, s', \rho_i, w') > \text{Max}_w\, Eu(\theta, q, s, \rho_i, w)\}$

Then, $Pr[q(\theta, i)|s''] > 0$ if and only if $s'' \in \xi(s)$ (E2bd)

Note that we include the posterior belief (through its effect on the firm's action – the wage) in condition (E2bd). When considering separating equilibrium beliefs it is useful to include the firm's posterior belief (or a function of this belief) in the worker's expected utility function.

The following observation is helpful in defining the third equilibrium condition. Since the wage $w(s(q, \rho_i), \rho_i) = v^E(s(q, \rho_i), \rho_i)$ is a function of the stereotype, the quality decision is also a function of ρ_i.

Defining θ^H as the solution to

$$Eu[\theta^H, h, s(h, \rho_i), \rho_i] - Eu[\theta, l, s(l, \rho_i), \rho_i] = 0, \tag{3.5}$$

it is seen that $\theta^H(\rho_i)$ is indifferent between high and low quality and is, therefore, the highest θ choosing high quality and $F(\theta^H(\rho_i))$ is the probability that a worker with label i chooses high quality. The third equilibrium condition, consistency with a common prior distribution, then requires:

$$\rho_i = Pr(q{=}h|i) = Pr(Eu[\theta, h, s(h, \rho_i), \rho_i] \geq Eu[\theta, l, s(l, \rho_i), \rho_i]) =$$

$$Pr(Eu[\theta, h, s(h, \rho_i), \rho_i] \geq Eu[\theta^H, h, s(h, \rho_i), \rho_i]) =$$

$$Pr[C(\theta, h, s(h, \rho_i)) \leq C(\theta^H(\rho_i), h, s(h, \rho_i))] = Pr(\theta \leq \theta^H(\rho_i)) = F(\theta^H(\rho_i)), \tag{E3c}$$

so that stereotypes about label i induces the same measure of workers with label i to choose high quality. Consistency, in this context, is essentially a fixed point argument. That is, the stereotype is an interim belief on quality choice that is self-fulfilling and is also consistent with the commonly known prior distribution of types.

Given the consistency requirement, the three equilibrium conditions can be stated as follows: a collection of strategies $\{q, s, w\}$ that are best responses given $\{\theta, i\}$ and the firm's beliefs, $b(s, i)$; posterior beliefs are derived by Bayes' Rule where applicable and are subject to the dominance refinement at all information sets; $\rho_i = F(\theta^H(\rho_i))$.

C. Separating Equilibrium.

In this section we demonstrate that a reasonable separating equilibrium always exists for this economic environment, that it is unique, and that it renders labels meaningless in determining quality decisions. In describing a separating equilibrium we note that the worker's expected utility can be written as a function of the posterior belief: $Eu(\theta, q, s, \rho_i, b(s, \rho_i)) = Eu(\theta, q, s, \rho_i)$. Furthermore, if the posterior belief is held constant, then assumptions 1 and 2 provide for the strict concavity of the worker's expected utility function in s. The full information signals for the high, s^{h*}, and low-quality, s^{l*} are, therefore, positive and finite.

In a separating equilibrium a low-quality worker does not mimic the high-quality worker's signal. The highest signal that the low-quality worker would copy, even if they were believed to be high quality, is denoted as s^s and is given by:

$$Eu[\theta, l, s^s, \rho_i, 1] = Eu[\theta, l, s^{l*}, \rho_i, 0]. \tag{3.6}$$

Given the strict concavity of the expected utility, the low-quality would prefer to reveal their quality and choose the full information signal, s^{l*}, instead of any signal $s > s^s$. In the typical case $s^s > s^{h*}$, so that the high-quality worker must distort their expenditure to signal their quality, however, if the low-quality cost is much steeper than the high-quality cost, then no distortion is necessary and $s^{h*} \geq s^s$. Hence, a candidate for a separating sequential equilibrium is where the high-quality agent chooses $s^R = \max\{s^{h*}, s^s\}$.

Still, a separating sequential equilibrium may require a high-quality signal that is even larger than s^R (with posterior beliefs of zero for all s that are less than this high separating signal). Such equilibria, however, are not reasonable. In particular, in any equilibrium with $s > s^R$ the high-quality worker would prefer to reduce their signaling expenditure towards s^R. Such a deviation could not have come from the low-quality worker because for them any $s > s^R$ is strictly dominated by choosing s^{l*} and admitting to being low quality. By equilibrium condition (E2bd) the firm would have to ascribe probability one to any $s > s^R$ coming from a high-quality worker. Hence, in any reasonable separating equilibrium the high-quality worker chooses $s^R = \max\{s^{h*}, s^s\}$. Assumption 4 below insures that high quality is chosen in a separating equilibrium:

$$Eu[\underline{\theta}, h, s^s, \rho_i, 1] > Eu[\theta, l, s^s, \rho_i, 1] < Eu[\theta, l, s^{h*}, \rho_i, 1]. \tag{3.7}$$

The second part of assumption 4 says that the low quality would mimic s^{h*}, therefore, $s^s > s^{h*}$ and we can restrict attention to distortionary signaling equilibria: $s^R = s^s$. The first part of assumption 4 says that a positive measure of types choose high quality at the least costly distortionary separating equilibrium. We now show the following (the proof is in the appendix).

Proposition 1: *If assumptions 1 – 4 are satisfied, then there exists a unique reasonable separating equilibrium where* $\rho_i = F(\theta^H) \in (0, 1]$ *is determined uniquely by the workers abilities and independently of their labels. All* $\theta \leq \theta^H$ *choose* $\{h, s^s\}$ *and all* $\theta > \theta^H$ *choose* $\{l, s^{l*}\}$.

In interpreting proposition 1, note that s^R is the solution to the constrained maximization problem $Eu[\theta, h, s, \rho_i, 1]$ subject to $Eu[\theta, l, s, \rho_i, 1] \leq Eu[\theta, l, s^{l*}, \rho_i, 0]$ and it is the Riley (1979) or lowest cost separating equilibrium of this signaling game. The worker becomes high quality only if it is more profitable than low quality and this profitability is determined by the worker's type, but not their label. A worker with a low cost of being high quality ($\theta \leq \theta^H$) chooses high quality and those with higher costs choose low quality.

D. The Main Results

We now consider pooling equilibria. There is no a priori reason that the equilibrium correspondence given by E3C should have a unique solution. In fact, as we saw for the example in section 2, it may have an infinite number of solutions. We now examine that idea here. First note that in a pooling equilibrium both qualities choose the same pooling expenditure s^p, therefore, from (E2bp) the posterior beliefs are the same as the prior beliefs and the expected wage is the same for either quality. Using these observations and rewriting equation (3.5) for a pooling equilibrium then yields the following:

Proposition 2: *If assumptions 1 – 4 are satisfied, then:* (*i*) *there exists a continuum of pooling sequential equilibria;* (*ii*) *in each equilibrium the pooling expenditure level is determined uniquely from the stereotype;* (*iii*) *pooling sequential equilibria with higher stereotypes require larger pooling expenditures. (iv.) The probability of a high-quality worker is larger in the separating equilibrium, than in any pooling equilibrium with* $s^p \leq s^s$.

Proof: For pooling equilibria equation (3.5) takes the following form: $Eu[\theta^H, h, s^p, \rho_i] = Eu[\theta, l, s^p, \rho_i]$. The expected wage is the same for any quality choosing s^p, therefore, this equation can be rewritten as:

$$I(\theta^H, h) = c(l, s^p) - c(h, s^p). \tag{3.8}$$

Solving this equation for ρ_i yields:

$$\rho_i = F(\theta^H) = F(I^{-1}(I(\theta^H, h))) = F(I^{-1}(c(l, s^p) - c(h, s^p))). \tag{3.9}$$

Noting that *F, I,* and $c(l, s^p) - c(h, s^p)$ are continuous, monotonic functions establishes parts *i* and *ii.* To establish part (*iii.*) totally differentiate (3.9), and note that *F, I,* and therefore I^{-1} are strictly increasing:

$$\frac{\partial \rho_i}{\partial s^p} = F_\theta(\cdot) I^{-1}{}_\theta(\cdot)[c_s(l,\cdot) - c_s(h,\cdot)] > 0 \,. \tag{3.10}$$

To establish part (*iv.*) first denote the indifferent type in a separating equilibrium as $\theta^{Hs}(s)$ and in a pooling equilibrium as $\theta^{Hp}(s)$. For any s, we have that $I(\theta^{Hs}(s)) = v(h, s) - c(h, s) - v(l, s) + c(l, s) > c(l, s) - c(h, s) = I(\theta^{Hp}(s))$. Hence, for $s^p \le s^s$, we have that $F(\theta^{Hs}(s^s)) > F(\theta^{Hp}(s^p))$.□

Proposition 2 shows that when quality choice is endogenous, pooling equilibria can generate self-fulfilling statistical discrimination. In particular, there are multiple stereotypes that are correct in a pooling equilibrium. Equilibria with higher stereotypes are correct because they require a higher pooling expenditure which, in turn, increase the benefit of becoming high quality.

There are two related observation about this set of equilibria. The first is that equilibria with a higher stereotype do not necessarily Pareto dominate those with a lower stereotype. Whereas the worker benefits from a better stereotype, the higher stereotype also requires a larger expenditure. For example, the further restrictions placed on the distribution function considered in section 2 provide for a unique optimal stereotype and corresponding signal for each quality. The worker's utility is decreasing for stereotypes that are higher than this optimal stereotype.

The second observation is that although there are many correct equilibrium stereotypes, not all of these beliefs are necessarily reasonable. For example, if $s^p = 0$, then no type of worker would become high quality and, therefore, $\rho_i = 0$. If high quality is important to the firm, and the investment cost is not expensive for a sufficiently large measure of the types, then at least one type would choose to deviate to the separating equilibrium, which would break this pessimistic pooling equilibrium. More generally, for any proposed pooling belief, if either quality prefers the separating equilibrium to their payoff at the proposed pooling equilibrium then at least one type would choose to break the pooling equilibria. Only pooling equilibria that are preferred by both qualities are reasonable. Whereas a reasonable separating equilibrium always exists, the same is not necessarily true of reasonable pooling equilibrium. If, however, the distribution function satisfies a certain condition, then reasonable pooling equilibria exist.

To explain this condition, we define

$$\sigma(\theta) \equiv (\ln(F'))' = F''/F' \tag{3.11}$$

as the degree of concavity of the distribution. In particular, for two distributions F_1 and F_2 of the same type, with corresponding σ_1 and σ_2, if $\sigma_1(\theta) < \sigma_2(\theta)$ for each θ, then we say that distribution 1 is relatively more concave. Assumption 5 states that $\sigma(\theta)$ is less than zero and that $F(\theta)$ is log concave.

$$\sigma(\theta) < 0 \tag{3.12}$$

Assumption 5 is a strong assumption (it is only a sufficient condition, therefore, our results would obtain with a weaker restriction); however, it is illustrative. In particular, assumption 5 is that the density function is decreasing. Note that the density (F') must be positive, therefore, if it is decreasing, then $(\ln(F))'' = \frac{FF''-F'F'}{F^2} < 0$, so that a decreasing density function is sufficient for the distribution function to be log concave.[15] We can now show the following.

Proposition 3: *If assumptions 1 – 5 are satisfied, then there exists a* $k^* > 0$ *such that if* $\sigma(\underline{\theta}) < -k^*$*, then reasonable pooling equilibria exist. The most pessimistic stereotype,* $s^p = \rho_i = 0$*, is never part of a reasonable pooling equilibrium.*

The proof of proposition 3 is in the appendix. The idea is straightforward. For the Pareto distribution $\sigma(\theta) = -(k+1)\frac{\underline{\theta}}{\theta}$ and as we saw in section 2 reasonable pooling equilibria exist for the Pareto distribution when the shape parameter, k, is greater than some specific level that depends on the chosen parameters of the model. For the exponential distribution $\sigma(\theta) = -k$ and reasonable pooling equilibria exist if the rate k is sufficiently large. The Zipf distribution is a special case of the Pareto distribution with $k = 1$; however, if the signal is valuable enough to the firm (apart from its information aspect), then reasonable pooling equilibria exist for the Zipf distribution as well. The uniform distribution, like the exponential distribution is log concave with a weakly log-convex density, however, the uniform density is non-decreasing so that $\sigma(\theta) = 0$ and reasonable pooling equilibria do not exist for the uniform distribution.

An immediate corollary of the first part of proposition 3 is that self-fulfilling statistical discrimination can exist in labor market signaling models even when the firm does not believe that the

---

[15] These conditions are satisfied for the Pareto, exponential, and Zipf, as well as restricted versions of the Weibull, power, beta, gamma, chi-squared, truncated normal, and log-normal distributions. For more on log-concave distributions see Bagnoli and Bergstrom (2005). For more on the Pareto and Zipf distributions see Axtell (2001).

worker would play a strictly dominated strategy.

Corollary 1: *Reasonable self-fulfilling statistical discrimination equilibria (where the worker is not believed to play a strictly dominated strategy) exist in signaling models where quality choice is endogenous and the distribution of types is sufficiently concave.*

We now consider some properties of the set of reasonable pooling equilibria. It is easier to describe the more interesting of these properties if we can be certain that the agent's expected utility function is strictly concave in s^p. Whereas assumptions 1 and 2 guarantee strict concavity when the posterior belief is fixed (the full information and separating equilibria), we need to place more structure on the model to ensure strict concavity in a pooling equilibrium. The following assumption 6 is weakly sufficient for the strict concavity of the stereotype function given in equation (3.9) and of the worker's expected wage (and, therefore, their expected utility)

$$I_{\theta\theta}(\cdot, h) \geq 0 = c_{ssq}(\cdot, \cdot) = v_{qss}(\cdot, \cdot) = v_{qs}(\cdot, \cdot). \tag{3.13}$$

In interpreting assumption 6 note that the weak convexity of the investment function indicates that its inverse is weakly concave. Differentiation of equation (3.10) with respect to s^p reveals that weakly sufficient conditions for the stereotype to be strictly concave in s^p are weak concavity of $I^{-1}(\theta, h)$ along with the assumption that the convexity of the cost function in s does not depend on quality. The last part of assumption 6 states that even when the signal is non-dissipative its value is the same in a high and a low quality agent. It is straightforward to verify that these assumptions (along with assumptions 1, 2 and 5) are weakly sufficient for the strict concavity of the worker's expected utility and that they were satisfied for the examples in section 2. We can now show the following. The proof is in the appendix.

Proposition 4: *If assumptions 1 – 6 are satisfied then the following hold. (i.) The set of reasonable pooling equilibria stereotypes is compact.* (*ii.*) *Any type that would break a low non-reasonable low-stereotype pooling equilibrium would separate as a high quality worker. (iii.) For any non-reasonable pooling equilibrium, there is a continuum of the lowest types starting at* $\underline{\theta}$ *that would deviate to the high quality outcome in the separating equilibrium. (iv.) All reasonable pooling equilibria generate a weakly* *lower probability of a high quality worker than does the separating equilibrium.*

Parts (*i.*) and (*iii.*) of proposition 4 use the assumed concavity to establish that the set of reasonable pooling equilibria is a compact set and adjacent to this set there is another compact set of non-reasonable pooling equilibria that begins with the most pessimistic stereotype. (There may also exist another set of very optimistic non-reasonable pooling stereotypes that are less interesting.) Dividing the set of pooling equilibria is convenient and it allows us to concentrate on the more important parts (*ii*.) and (*iv*.) of proposition 4. In particular, low types are the types that would break a pessimistic non-reasonable pooling equilibrium by choosing high quality and the separating expenditure. Part (*ii*.) shows that all of these low types separate to high quality and part (*iii*.) shows that these types comprise a compact interval beginning with the lowest type. Part (*iv*.) shows that although reasonable pooling equilibria may be preferred by all types, they still generate a lower probability of a high-quality agent than does the separating equilibrium.

An interesting implication of part (*iv*.) of Proposition 4 is that a reasonable (high or low) stereotype can generate complacency. In particular, an agent with that label can share their groups reputation and choose not to differentiate themselves. On the other hand an agent from a group that has a low stereotype will choose to overcome the pessimistic stereotype and separate if they can. Of course, this separation will leave a low quality agent worse off than in the pooling equilibrium. Even the separating high-quality agent would have preferred a reasonable pooling equilibrium if the option was open to them (but it was not available). We state this point as corollary 2.

Corollary 2: *Low types from a label to which is attached a pessimistic non-reasonable pooling stereotype will engage in counter-stereotypical behavior and their label will have a higher probability of a high quality agent than a label that has a more favorable reasonable stereotype.*

Corollary 2 suggests that a good stereotype generates a "Dutch Disease" effect whereby a (capable) agent with that label invests in less education and is less likely to choose to become high quality, than if they had a label with a lower stereotype.

IV. Conclusion

We have introduced a model of stereotype threat in a signaling model that does not require employers to believe that anyone would play a strictly dominated strategy. The existence of multiple equilibria, which can generate a stereotype threat effect, is shown to depend on the shape and variance of the distribution of types as well as the value of the signal. They are more likely if there is less variance in the types or if the signal has more value to the firm. We also show that a very bad stereotype forces the high-quality agent with that bad stereotype to separate. This counter-stereotypical behavior increases that label's high-quality probability and education level. In this way the very discriminated against label overtakes the complacent good stereotype label, and the good stereotype label may suffer from a reputational Dutch disease.

Appendix

Proof of Proposition 1:

By assumptions 1 and 2 the agent's expected utility function is continuous and strictly concave in s, therefore, s^{l*} and s^{h*} are both positive and finite and $s^{l*} < s^{h*}$. Furthermore, $Eu[\theta, l, s, \rho_i, 1]$ attains a maximum at s^{lh*} where $s^{l*} < s^{hl*}$, $s^{lh*} < s^{h*}$, and $Eu[\theta, l, s, \rho_i, 1]$ is strictly decreasing for $s > s^{lh*}$. Now, $Eu[\theta, l, s^{lh*}, \rho_i, 1] > Eu[\theta, l, s^{l*}, \rho_i, 1] > Eu[\theta, l, s^{l*}, \rho_i, 0]$, therefore, by continuity of $Eu[\theta, l, s, \rho_i]$ in s, there exists a $s^s > s^{lh*}$ such that $Eu[\theta, l, s^s, \rho_i, 1] = Eu[\theta, l, s^{l*}, \rho_i, 0]$ and for all $s > s^s$ it is the case that $Eu[\theta, l, s, \rho_i, 1] < Eu[\theta, l, s^{l*}, \rho_i, 0]$. Hence all $s > s^s$, are dominated strategies for the low-quality agent for any posterior beliefs. If there are posterior beliefs such that s^R is not a dominated strategy for the high-quality agent then there is an equilibrium where credible posterior beliefs are $\Pr(h|s < s^s) = 0$ and the low-quality agent's best response is s^{l*}.

We now show that if the posterior beliefs are $\Pr(h|s < s^s) = 0$, $\Pr(h|s \geq s^s) = 1$, then a high-quality agent's unique dominant strategy is s^R. First note that $Eu[\theta, h, s, \rho_i, 0]$ attains a maximum at s^{hl*} where $s^{l*} < s^{hl*} < s^{h*}$. We want to show that $Eu[\theta, h, s^R, \rho_i, 1] > Eu[\theta, h, s^{hl*}, \rho_i, 0]$, which implies that

$$Eu[\theta, h, s^R, \rho_i, 1] - Eu[\theta, h, s^{hl*}, \rho_i, 0] > 0 = Eu[\theta, l, s^s, \rho_i, 1] - Eu[\theta, l, s^{l*}, \rho_i, 0] \quad \text{(A.1)}$$

$$\Leftrightarrow v^E(h, s^R) - C(\theta, h, s^R) - v^E(l, s^{hl*}) + C(\theta, h, s^{hl*}) > v^E(h, s^R) - C(\theta, l, s^R) - v^E(l, s^{l*}) + C(\theta, l, s^{l*})$$

$$\Leftrightarrow C(\theta, l, s^R) - C(\theta, h, s^R) > v^E(l, s^{hl*}) - C(\theta, h, s^{hl*}) - v^E(l, s^{l*}) + C(\theta, l, s^{l*}) \quad \text{(A.2)}$$

Now, the right-hand-side of (A.2) is greater than

$$v^E(l, s^{l*}) - C(\theta, h, s^{l*}) - v^E(l, s^{l*}) + C(\theta, l, s^{l*}) = C(\theta, l, s^{l*}) - C(\theta, h, s^{l*}). \quad \text{(A.3)}$$

Hence, for high quality, s^R dominates all lower s if

$$C(\theta, l, s^R) - C(\theta, h, s^R) > C(\theta, l, s^{l*}) - C(\theta, h, s^{l*}) \quad \text{(A.4)}$$

$$\Leftrightarrow C(\theta, l, s^R) - C(\theta, l, s^{l*}) > C(\theta, h, s^R) - C(\theta, h, s^{l*}), \quad \text{(A.5)}$$

which is true by assumption 1.

To see that $\Pr(h|s < s^s) = 0$, $\Pr(h|s \geq s^s) = 1$ are the unique reasonable separating equilibrium beliefs, note that the belief $\Pr(h|s \geq s^s) < 1$ implies that a low-quality agent played a dominated strategy however, these beliefs do not satisfy (E2bD). To see this point note that s^s is a weak best response for the low quality agent only when $\Pr(h|s \geq s^s) = 1$, for any lower posterior belief s^s is a dominated response for the low-quality agent. Furthermore, a posterior belief $\Pr(h|s') = 1$ for some $s' < s^s$ causes the low-quality agent to deviate to s'.

Finally note that that $Eu[\theta^H, h, s^R, \rho_i, 1] = Eu[\theta, l, s^{l*}, \rho_i, 0]$ and for all $\theta < \theta^H$, $Eu[\theta, h, s^R, \rho_i, 1] > Eu[\theta, l, s^{l*}, \rho_i, 0]$; therefore, all $\theta \in [\underline{\theta}, \theta^H]$ choose h. For $\theta < \theta^H$, $Eu[\theta, h, s^R, \rho_i, 1] < Eu[\theta, l, s^{l*}, \rho_i, 0]$; therefore, all $\theta \in (\theta^H, \overline{\theta}]$, choose l.□

Proof of Proposition 3: In a reasonable pooling equilibria no type would deviate to the full-information separating equilibrium. For type θ this implies that

$$Eu[\theta, q(\theta, i), s^p, \rho_i] \geq \text{Max } \{Eu[\theta, h, s^s, \rho_i, 1], Eu[\theta, l, s^{l*}, \rho_i, 0]\} \quad . \qquad \text{(A.6)}$$

Using equation 3.9 to rewrite the left hand side of the above inequality yields:

$$v^E[s^p, F(I^{-1}[c(l, s^p) - c(h, s^p)])] - C(\theta, q(\theta, i), s^p) \geq \text{Max } \{Eu[\theta, h, s^s, \rho_i, 1], Eu[\theta, l, s^{l*}, \rho_i, 0]\}. \qquad \text{(A.7)}$$

From assumption 4 we know that if s^p is close enough to s^{h*} (which is less than s^s) and if ρ_i is close enough to one, then at least $\underline{\theta}$ would prefer the pooling equilibrium. If, in addition, $\sigma(\theta)$ is sufficiently negative and approaches negative infinity, then the entire distribution would converge to a mass point at $\underline{\theta}$. The correct stereotype would approach one and all types would prefer the pooling equilibrium. Now, for any two distributions, if $\sigma_1(\theta) < \sigma_2(\theta)$ for all θ, then $F_1(\cdot)$ first-order stochastically dominates $F_2(\cdot)$, therefore, for any s^p we have that $\rho_i = F(I^{-1}[c(l, s^p) - c(h, s^p)])$ is increasing in the degree of concavity of the distribution. Finally, note that v^E is increasing in ρ_i, therefore, it is increasing in s^p and it is increasing faster in s^p when $\sigma(\theta)$ is lower. Hence, there exists a $k^* > 0$ such that if $\sigma(\underline{\theta}) < -k^*$, then $F(I^{-1}(\cdot)$ is sufficiently large and equation A.2 holds as a strict inequality for some $\theta^H > \underline{\theta}$. To see the second part of proposition 3, note from assumption 4 that $Eu[\underline{\theta}, h, s^s, \rho_i, 1] > Eu[\underline{\theta}, l, s^s, \rho_i, 1] = Eu[\underline{\theta}, l, 0, 0, 0]$ so that at least $\underline{\theta}$ would break the most pessimistic pooling equilibria.□

Proof of Proposition 4: The right hand side of equation (A7) is $Eu[\theta, h, s^s, \rho_i, 1]$ for all $\theta \leq \theta^{Hs}(s^s)$ and it is $Eu[\theta, l, s^{l*}, \rho_i, 0] = Eu[\theta^{Hs}(s^s), h, s^s, \rho_i, 1]$ for all $\theta > \theta^H(s^s)$. Hence, for any θ the right hand side of equation (A7) is constant. We now rewrite equation (A7) as:

$$v^E[s^p, F(I^{-1}[c(l, s^p) - c(h, s^p)])] \geq C(\theta, q(\theta, i), s^p) + \text{Max } \{Eu[\theta, h, s^s, \rho_i, 1], Eu[\theta, l, s^{l*}, \rho_i, 0]\}. \qquad \text{(A8)}$$

Note that although $c(q, s)$ is strictly convex in s, $C(\theta, q(\theta, i), s^p)$ is not continuously differentiable in s^p. In particular, as s^p increases a given θ will switch from low to high quality. From proposition 3 we know that this critical $s < s^s$. Hence, we only need to consider the upper contour of the right hands side. There, therefore, exists a $\hat{s}$ such that, for $s \leq \hat{s}$ the right hand side of (A8) equals $v(h, s^s) - c(h, s^s) + c(h, s)$ and for $s > \hat{s}$ it equals $v(l, s^s) - c(l, s^s) + c(l, s)$. Both of these functions are strictly convex in s and the latter crosses the former exactly once and it crosses from below at $\hat{s}$.

We now show that the left hand side of (A8) is strictly concave in s, which implies that it is greater than the right hand side over a compact interval. First differentiate the left hand side twice with respect to s^p to yield:

$$v^E_{ss} + 2[v_s(h,\cdot) - v_s(l,\cdot)]\frac{\partial \rho_i}{\partial s^p} + \frac{\partial[v_s(h,\cdot) - v_s(l,\cdot)]}{\partial \rho_i}\left(\frac{\partial \rho_i}{\partial s^p}\right)^2 + [v(h,\cdot) - v(l,\cdot)]\frac{\partial^2 \rho_i}{\partial (s^p)^2}. \quad \text{(A9)}$$

The first term is negative, the second and third terms are zero and the fourth term depends on the sign of

$$\frac{\partial^2 \rho_i}{\partial (s^p)^2} = F_{\theta\theta}(\cdot)I^{-1}{}_{\theta}(\cdot)[c_s(l,\cdot) - c_s(h,\cdot)] + F_{\theta}(\cdot)I^{-1}{}_{\theta\theta}(\cdot)[c_s(l,\cdot) - c_s(h,\cdot)] + F_{\theta}(\cdot)I^{-1}{}_{\theta}(\cdot)[c_{ss}(l,\cdot) - c_{ss}(h,\cdot)] < 0.$$

To establish part (*ii.*) note that the strict concavity of the left hand side of (A8) and the increasing convexity of the right hand side after $\hat{s}$ implies that $\hat{s}$ would never be less than the minimum member of this compact set. Hence, for any low non-reasonable pooling stereotype with corresponding $s < \hat{s}$, any θ that prefers to separate will choose high quality. To establish part (*iii.*) note that this set is also compact and begins at $\underline{\theta}$. To establish part (*iv.*) note that s^s with corresponding $\rho_i = 1$ is the maximal element of the set of reasonable pooling equilibria. Hence, proposition 3 along with the fact that a reasonable $s^p \leq s^s$ implies that for all reasonable pooling equilibria $F(\theta^{Hs}(s^s)) \geq F(\theta^{Hp}(s^p))$.□

REFERENCES

Aigner, D. and G. Cain. 1977. "Statistical Theories of Discrimination in the Labor Market," *Industrial and labor Relations Review,* 30: 175-187.

Akerlof, G. 1976. "The Economics of Caste and the Rat Race and Other Woeful Tales," *Quarterly Journal of Economics,* 90: 599-617.

Arrow, K. J. 1973. "The Theory of Discrimination," in O. Ashenfelter and A. Rees, eds., *Discrimination in Labor Markets*. Princeton, NJ: Princeton University Press.

Axtell, R. 2001. "Zipf Distribution of US Firm Sizes," *Science*, 293: 7, September 2001, 1818-1820.

Bagnoli, M. and T. Bergstrom. 2005. "Log-concave Probability and its Applications," *Economic Theory,* 26, 445-469.

Banks, J. and J. Sobel. 1987. "Equilibrium Selection in Signaling Games," *Econometrica,* 55: 647-662.

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

Bodenhausen, G. and M. Lichtenstein. 1987. "Social Stereotypes and Information-Processing Strategies: The Impact of Task Complexity," *Journal of Personality and Social Psychology*, 52: 871-880.

Bodenhausen, G. and R. Wyer. 1985. "Effects of Stereotypes on Decision Making and Information-Processing Strategies," *Journal of Personality and Social Psychology*, 48: 267-282.

Cabral, L. and J. Mata. 2003. "On the Evolution of the Firm Size Distribution: Facts and Theory," *American Economic Review,* 93(4): 1075-1090.

Cho, I. and D. Kreps. 1987. "Signaling Games and Stable Equilibria," *Quarterly Journal of Economics,* 102: 179-221.

Coate, S. and G. Loury. 1993. "Will Affirmative Action Policies Eliminate Negative Stereotypes," *American Economic Review,* 83: 1220-1240.

Coate, S. and S. Tennyson. 1992. "Labor Market Discrimination, Imperfect Information and Self Employment," *Oxford Economic Papers,* 44: 272-288.

Dvir, T., D. Eden and M. Banjo. 1995. "Self-fulfilling Prophecy and Gender: Can Women be Pygmalion and Galatea?" *Journal of Applied Psychology*, 80: 253-270.

Eden, D. 1992. "Leadership and Expectations: Pygmalion Effects and Other Self-fulfilling Prophecies in Organizations," *Leadership Quarterly*, 3: 271-305.

Feltovich, N., R. Harbaugh and T. To. 2002. "Too Cool for School? Signaling and Countersignaling," *The Rand Journal of Economics*, 33(4): 630-649.

Fryer, R. 2007. "Belief Flipping in a Dynamic Model of Statistical Discrimination," *Journal of Public Economics,* 91(5-6): 823-1230

Good, C. 2008. “Problems in the Pipeline: Stereotype Threat and Women’s Achievement in High-level Math Courses,” *Journal of Applied Developmental Psychology*, 29(1): 17-28.

Gupta, V., D. Turban and N. Bhawe. 2008. “The Effect of Gender Stereotype Activation on Entrepreneurial Intentions,” *Journal of Applied Psychology*, 93:5.

Grossman, S. and M. Perry. 1986. “Perfect Sequential Equilibrium,” *Journal of Economic Theory,* 39: 97-119.

Kreps, D. and R. Wilson. 1982. “Sequential Equilibria,” *Econometrica,* 50: 863-894.

Kiefer, A. 2007. “Implicit Stereotypes and Women’s Math Performance: How Implicit Gender-math Stereotypes Influence Women’s Susceptibility to Stereotype Threat,” *Journal of Experimental Social Psychology,* 43(5): 825-832.

Lundberg, S. and R. Startz. 1983. “Private Discrimination and Social Intervention in Competitive Labor Markets,” *American Economic Review,* 73: 340-347.

Mailath, G., M. Okuno-Fujiwara and A. Postlewaite. 1993. “Belief Based Refinements in Signaling Games,” *Journal of Economic Theory,* 60: 241-276.

Milgrom, P. and S. Oster. 1987. “Job Discrimination, Market Forces and the Invisibility Hypothesis,” *Quarterly Journal of Economics,* 102: 453-476.

Melitz, M. 2003. “The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity,” *Econometrica*, 71: 1695-1725.

Neal, D. and W. Johnson. 1996. “The Role of Premarket Factors in Black-White Wage Differences,” *Journal of Political Economy,* 104: 869–95.

Phelps, E. 1972. “The Statistical Theory of Racism and Sexism,” *American Economic Review,* 62: 659-661.

Riley, J. 1979. “Informational Equilibria,” *Econometrica,* 47: 331-360.

Rosenthal, R. and L. Jacobson. 1968. Pygmalion in the Classroom: Teacher Expectation and Pupil's Intellectual Development. Holt, Rinehart and Winston.

Simon, H. 1957 *Models of Man.* New York: Wiley.

Spence, M. 1973. “Job Market Signaling,” *Quarterly Journal of Economics*, 87: 355-374.

Steele C. and J. Aronson. 1995. “Stereotype Threat and the Intellectual Test Performance of African Americans,” *Journal of Personal and Social Psychology,* 69(5): 797–81

Figure 1

Reasonable Pooling Equilibria with a Dissipative Signal

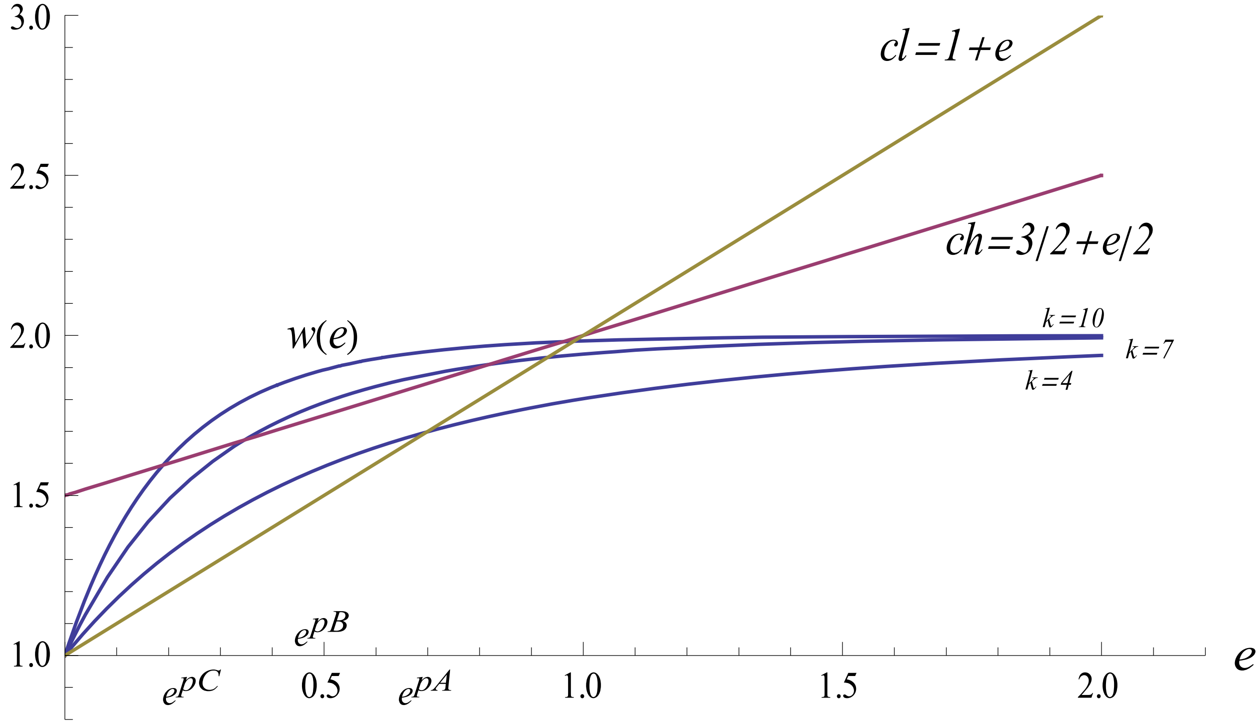

Figure 2

Reasonable Pooling Equilibria with a Non−Dissipative Signal