Opportunistic Discrimination*

Rick Harbaugh† Ted To‡

This Version: August 2009

Abstract

When can you cheat some people without damaging your reputation among others? In a trust game between a firm and a series of individuals from minority and majority groups, the firm has more incentive to cheat minority individuals because trade with the minority is less frequent and the long-term benefits of a reputation for fairness toward them are correspondingly smaller. If the majority is sufficiently large it gains nothing from a solidarity strategy of punishing opportunism against the minority, so the firm can continue doing business with the majority even if it cheats the minority. When a small fraction of firms have a preference-based bias against the minority, the interaction with reputation effects gives all firms a stronger incentive to cheat the minority, and discrimination is the unique renegotiation-proof equilibrium for firms of intermediate patience.

JEL Classification Categories: J71, J24, D63, L14.

Key Words: discrimination, trust, social capital, opportunism, reputation spillover

*For helpful comments we thank Mike Baye, George Evans, Dragan Filipovich, Sid Gordon, Anton Lowenberg, Barry Nalebuff, Jack Ochs, Debashis Pal, and Eric Rasmusen; conference participants at the American Law and Economics Association annual meeting, the C.O.R.E Conference on Heterogeneity in Organizations, the European meeting of the Econometric Society, the Mid-West Theory Conference, the Public Choice Society Conference, the Society of Labor Economists annual meeting, and the Stony Brook International Game Theory Conference; and seminar participants at the Bureau of Labor Statistics, the Claremont Colleges, the Federal Reserve Bank of Cleveland, Indiana University, and the University of Oregon.

†Indiana University, `riharbau@indiana.edu`

‡Bureau of Labor Statistics, `To_T@bls.gov`

> Where people seldom deal with one another, we find that they are somewhat disposed to cheat, because they can gain more by a smart trick than they can lose by the injury which it does their character.
>
> – Adam Smith, *Lectures on Jurisprudence*, 1766

1 Introduction

How do people react when other people are cheated? If a person's land is stolen during ethnic unrest, is the perpetrator viewed as generally opportunistic or as someone who can still be trusted in his own community? If a tourist is cheated in front of local customers, is the offender seen as untrustworthy in general or only toward outsiders? If a woman is unfairly denied partnership by a law firm, do her colleagues expect a similar fate and leave or do they see the firm as opportunistic only toward women? If a government expropriates foreign investors, is it treated as untrustworthy toward everyone or only toward foreigners? If an insurance company fails to pay one group of customers, do other customers expect similar opportunism or do they still expect to be treated fairly?

The incentive to engage in opportunism clearly depends on this question of how people are expected to react to it. As Smith (1766) noted, a trader must decide whether the short-run gains from cheating are worth the loss from a damaged reputation. But in evaluating this tradeoff it is not clear that all acts of opportunism affect reputation in the same way. If a person from one group is cheated, can people from another group choose to ignore it and proceed with business as usual? If so, then the incentive to cheat a person can depend on the person's group identity, so that it might be profitable to cheat members of one group but not another.

This idea that opportunism can be discriminatory in its target is surprisingly absent from economic theories of discrimination. Standard models including occupational segregation (Fawcett, 1892), non-competitive wage setting (Fawcett, 1918; Edgeworth, 1922), discriminatory preferences (Becker, 1957), statistical discrimination (Phelps, 1972; Arrow, 1973; Lundberg and Startz, 1983; Coate and Loury, 1993), and search costs (Mailath et al., 2000; Lundberg and Startz, 2007) do not capture the idea that people from some groups are more likely to be taken advantage of than others.[1] Unequal access to the legal system can be part of the problem (Douglass, 1879), but even an unbiased legal system offers incomplete

[1] For a recent survey see Donahue (2007).

protection from opportunism due to the difficulty of adjudicating disputes (Williamson, 1985). We examine how opportunism can be discriminatory in an environment where the primary constraint is reputational rather than contractual or legal.

To understand the interaction between opportunism and discrimination, we model a repeated trust game (Kreps, 1990) between a firm and a set of individuals. In each period one of the individuals trusts the firm by making an investment or other resource commitment, and the firm then either cheats the individual by taking most of the gains of the transaction, or lets the individual benefit also. Since only one player has the choice of whether to be opportunistic, this "one-sided prisoner's dilemma" is the simplest environment in which to analyze reputation. Versions of it have been used to capture relations between a firm and its contractors (Klein and Leffler, 1981), an owner and a series of managers (Radner, 1985), a salesperson and his customers (Dasgupta, 1990), a government and foreign merchants (Greif et al., 1994), etc. Consistent with Smith's early arguments, if trade is sufficiently frequent,[2] or equivalently if the firm is sufficiently patient, trust can be sustained by a trigger strategy where everyone initially trusts the firm but if the firm cheats anyone then no one trusts the firm again.[3]

We analyze this game when the set of individuals is divided into two identifiable groups that interact with the firm with different frequencies, i.e., one is the "majority" and the other the "minority". We first consider the standard trigger strategy, which we now call the "solidarity trigger strategy." Given such a strategy it is foolish for the firm to cheat anyone unless it plans to cheat everyone, so there is a reputation spillover and individuals are right to stop trusting the firm if it cheats a member of the other group. We then consider a "discriminatory trigger strategy" where individuals stop trusting the firm if it cheats a member of their own group, but continue to trust the firm if a member of the other group is cheated. Given such a strategy the firm recognizes that it can maintain part of its reputation even after cheating a member of the other group. Depending on how much the firm values its reputation, a discrimination equilibrium exists in which the firm is trustworthy toward one group but not toward the other group.

[2] Smith (1766) argued that opportunism decreases with the frequency of commercial exchange and was therefore highest in undeveloped regions like his native Scotland. He also argued that opportunism is more likely in political and diplomatic activities where transactions are less frequent than they are in commerce.

[3] However, in the trigger strategy everyone benefits ex post by forgiving the firm's cheating, implying that the trigger strategy might not be credible ex ante (Farrell and Weizsacker, 2001). We resolve this problem by introducing some uncertainty over the firm's type which induces a reputational concern to not cheat.

Since the minority group is smaller, transactions with the minority are rarer, and the value of maintaining a reputation for fairness toward the minority is correspondingly smaller. Therefore, even though majority and minority individuals are identical and the firm need not have any discriminatory preferences, we find that a discrimination equilibrium with discrimination against the minority is supported by a wider range of discount rates for the firm. Both the firm and the minority are better off ex ante if the firm can be trusted, but the minority is too small to sufficiently punish the firm for any opportunism so the firm has an incentive to cheat the minority ex post unless the majority switches to the solidarity trigger strategy. If the majority is sufficiently large to protect itself by punishing opportunism against its own members, then it gains nothing from switching strategies.

These results show that discrimination can arise even when the firm is only interested in maximizing its profits and does not have any preference-based biases as in Becker (1957). Nevertheless, there is evidence that such preference-based biases exist (e.g., Bertrand et al., 2006) and have important economic effects (Charles and Guryan, 2008). To examine the interaction of such biases with reputation effects, we allow for the possibility that the firm is a biased firm that prefers to cheat one group. Clearly if such bias is widespread then individuals will be afraid to trust the firm. We are interested in the case where bias is relatively rare so that any effect must be indirect through its interaction with reputation, and we allow for bias against either the majority or the minority.

We find that there is no impact if the potential bias is against the majority, but if the firm might be biased against the minority then such bias interacts with reputation effects to make discrimination the unique coalition-proof equilibrium for firms of intermediate patience. Even when majority individuals start with a solidarity strategy, if they believe that an act of cheating the minority is probably due to bias rather than opportunism, they have an incentive to renegotiate their punishment strategy and continue business as usual. Therefore a biased firm can reap both the short-term benefits from cheating and the long-term benefits from a good reputation with the majority. A non-biased firm of intermediate patience then has an incentive to pool with biased firms by also cheating the minority. For instance, a proprietor might literally add insult to injury after cheating a customer in order to suggest to other customers that his opportunism is limited to a particular group.

This paper emphasizes reputational constraints on opportunism, but contractual and legal constraints are typically also present. A large literature investigates when these different constraints are substitutes or complements (e.g., Ostrom, 2000; Poppo and Zenger, 2002;

Lazzarini et al., 2004). In the context of opportunistic discrimination, a related question is how contractual and legal constraints interact with reputational constraints across groups. In many countries differential access to the legal system allows some groups but not others to gain contractual protection against opportunism. We find that increased contractual protection for the majority reduces the dependence of the majority on reputational sanctions, and thereby weakens reputational constraints on opportunism against the minority, leaving the minority worse off than if both groups were forced to rely on reputation alone.

Regarding anti-discrimination policies, we find that targeting enforcement against opportunistic discrimination, i.e., penalizing the particular behavior of a firm being opportunistic toward one group and fair toward another, is more effective at reducing total opportunism than either general enforcement which penalizes any opportunism or one-sided enforcement which penalizes opportunism directed at one group regardless of its behavior toward the other group. Comparatively small amounts of such enforcement can break the discrimination equilibrium, leading to the solidarity equilibrium with its low levels of opportunism. In contrast, the other forms of enforcement can in some cases aggravate opportunistic discrimination or lead to a reverse discrimination equilibrium.

Applying the model to an employment environment, our results imply that minority workers in the discrimination equilibrium will be reluctant to invest in firm-specific human capital for fear of having their quasi-rents appropriated. This is similar to the argument from the statistical discrimination literature that discrimination can be a self-fulfilling equilibrium in that fear of discrimination leads to an underinvestment in human capital. Such models do not indicate why one group rather than another group suffers from discrimination. In contrast, we make the particular prediction that discrimination is directed against the minority. This prediction is consistent with the common perception that "minorities" in different societies are at a disadvantage. It is also consistent with survey data showing that both men and women are more likely to report gender discrimination in occupations in which their gender is in the minority (Antecol and Kuhn, 2000), with laboratory experiments showing that minorities are less trusting (e.g., Fershtman and Gneezy, 2001), and with field experiments showing that minorities are more likely to be taken advantage of in bargaining environments (Ayres and Siegelman, 1995; Ayres, 2001).

While absent from the discrimination literature, the idea that some people are more vulnerable to opportunism is inherent to the argument in the social capital literature that social networks facilitate communication and trust (e.g., Coleman, 1988; Dasgupta, 1990).[4] The ability to communicate information about opportunism is central to the Greif (1993) model of long-distance traders, the Greif et al. (1994) model of merchant guilds, the Dixit (2003) model of trade networks in which distant trade is both more valuable and harder to monitor, and to the Annen (2003) model of inclusive networks in which larger networks expand cooperation opportunities at the cost of weaker communication about who cooperates. This literature looks at outcomes that are efficient subject to differential access to information through networks about who is trustworthy.

We differ from the social capital literature in focusing on how inefficient discrimination equilibria can arise even with public information.[5] Since individuals are aware of opportunism against members of other groups, they must decide whether to punish opportunism against some people and not others. There are typically multiple equilibria so the individually optimal decisions of firms and individuals depend on the decisions of others. Since unequal outcomes in our model are not the unavoidable product of different social networks, but instead are the result of decisions by individuals who face strategic uncertainty about how to best pursue their own interests, our approach can leave a large role for cultural expectations and norms in helping determine what equilibria prevail. For instance, if a society has a history of extreme racial or ethnic divisions then equilibria based on such differences might be more focal.

A fundamental insight of the discrimination literature is that competitive markets undermine discrimination (Fawcett, 1918; Becker, 1957). Since there is "strength in numbers" in our model, the relevant question is whether in an environment with multiple firms the minority can concentrate its business so that transactions are frequent enough to ensure fair treatment. We examine this issue in an extension to the model where we allow for different allocations of individuals across firms. We find that the minority can sometimes benefit from "self-segregation," but that the minority does equally well and sometimes better by concentrating its business on a single firm that also does business with the majority. For instance, the minority might be treated more fairly by a large corporation than by a small firm that specializes in business with the minority.

[4] As analyzed in the early literature on social capital (Loury, 1977), members of different groups have differing costs and benefits of investing in human capital even without overt opportunism, e.g., if a social network has more skilled members historically then it is easier for new members to acquire skills.

[5] As shown in an earlier draft of this paper, if information about firm opportunism is not public the intuition that better access to information can help compensate for a smaller group size holds in our model.

In deciding whether to trust or not, individuals might not know beforehand how patient the firm is but instead learn about the firm from its actions. Examining this case in an extension to the model, we find that the minority faces the most danger that the firm will take advantage of misplaced trust. We also find that firms might "spend down" their reputation by first being fair to the majority, then cheating the minority, and then cheating the majority. If the majority fails to stand up to opportunism against the minority initially, it finds itself facing a firm which has already lost part of its reputation, and therefore has less to lose from also cheating the majority. If the firm is relatively impatient, it will use the opportunity to switch from being fair to the majority to cheating it, thereby doubling the short-run gains from opportunism.

The model offers insight into how impersonal reputation systems such as traditional credit bureaus and, more recently, internet websites (Resnick et al., 2000) can reduce discrimination. The economics and social capital literatures have noted that such systems make information about opportunism more public, thereby reducing differential access to information and reducing the need for informal networks to share information. From the perspective of this model, an additional benefit is that information about the group affiliation of victims of opportunism is typically not public, so users of reputation systems cannot condition their response to opportunism based on such information. Therefore the discrimination equilibria we consider cannot arise.

Our analysis is complementary to the literature on collective reputations of different groups in repeated games following Tirole (1996). This literature is related to the statistical discrimination literature in that it shows that when individuals within a group are not clearly differentiated, their reputations will depend in part on the reputation of their group, and that reputation differences between groups can be self-perpetuating. While this literature analyzes reputations for trustworthiness *by* different groups, this paper consider reputations for trustworthiness *toward* different groups.[6] To emphasize this focus of the paper, the firm in our model is not required to have a group identity.[7]

The following section provides our basic model of a trust game. Section 3 then ex-

[6]In a one-shot assurance game, Basu (2005) shows how cooperation can break down between groups even while it is maintained within groups, so the analysis incorporates an aspect of trustworthiness toward different groups.

[7]The identity of the firm owner or manager could serve as a focal point for helping choose between equilibria with and without discrimination, especially if concern for identity is part of the utility function as in Akerlof and Kranton (2000). The identity of the firm player is central to the analysis of Annen (2001) in which different groups might have different stereotypes about a firm's trustworthiness.

tends this model to capture added realism in a number of aspects, including the interaction between reputation and preference-based discrimination, limited communication about opportunism, the interaction between reputation and explicit enforcement against opportunism, the potential for individuals self-selecting into different firms, and uncertainty over the firm's discount factor. Section 4 then concludes the paper.

2 The Model

We consider an infinitely repeated trust (or "hold-up") game (Kreps, 1990) in which in each stage or period an "individual" decides whether to trust a "firm" and then the firm decides whether to cheat the individual or not. We assume that the individual is randomly chosen from a finite population of players who are of two observationally distinct groups.[8] The firm is the same player in each period and its group identity is not relevant for the analysis. Since the trust game only allows for opportunism in one direction, and since only the individuals are from distinct groups, this game is perhaps the simplest game that can capture opportunistic discrimination.

The stage game is depicted in Figure 1 where the individual trusts the firm or not and then, if given the opportunity, the firm cheats the individual or is fair. Trusting involves an up-front cost $c > 0$ paid by the individual. This cost could be a price for a good of unsure quality paid by a customer, a firm-specific human capital investment by an employee, a loan made by a creditor, a transaction-specific investment by a supplier, etc. If the individual trusts the firm (trades with it), then the firm is either fair and the individual receives α where $c < \alpha < 1$, or cheats and the individuals receives β where $0 < \beta < c$. The total gross value of the trade is normalized to 1 so the firm receives $1 - \alpha$ from being fair and $1 - \beta$ from cheating. In the no-trust case there is no trade and both sides earn 0.

In each period the firm is randomly matched with one of $n > 1$ individuals from one of the two groups $p \in \{x, y\}$ of size n_p where $n_x > n_y \geq 1$. The match is independent across individuals and time so the probability that a particular individual is matched in any period is $1/n$ and the probability that a matched individual is from group p is $\gamma_p = n_p/n$.[9] Firms discount the period between transactions by a common knowledge factor $\delta \in (0, 1)$. We

[8] The finiteness assumption facilitates description of the model but is not central to the results.

[9] This implies that the larger group does more business with the firm. Alternatively one could allow the smaller group to represent the bulk of the business, in which case it is effectively the majority from the perspective of our analysis.

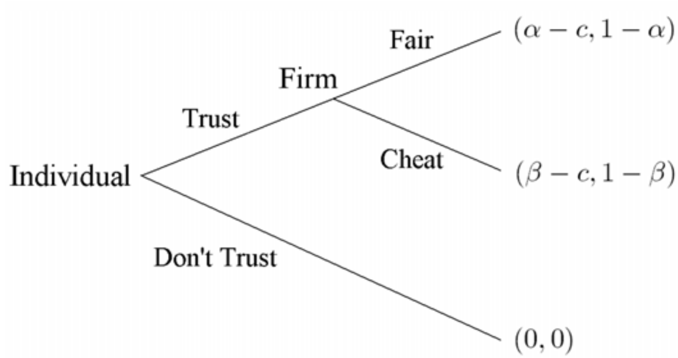

Figure 1: Game tree in each period: $0 < \beta < c < \alpha < 1$.

say a firm with a higher discount factor is more "patient" although this could also reflect a lower interest rate, a higher probability of survival to the next period, or a shorter interval between transactions. Individuals all share the same non-zero discount factor where the exact value does not affect the game.

To model reputation effects, in each period that the firm trades there is a small independent probability $\varepsilon > 0$ that the firm goes bad and transitions from being a "normal" firm which weighs the costs and benefits of cheating to being an "inept" firm which always cheats the individual.[10] For instance, the firm becomes very impatient due to financial problems, or cannot meet its obligations due to changes in ownership or the loss of key employees. We therefore follow the "separating" rather than "pooling" approach to reputation (Mailath and Samuelson, 2006), i.e., normal firms maintain their reputation by separating themselves from bad firms rather by pooling with good firms.[11] Since cheating is equilibrium behavior for some types in the separating approach, this assumption allows us to provide an intuitive analysis of how individuals respond to unexpected opportunism. All of our results are for the limiting case as ε goes to zero so we typically suppress explicit reference to ε.

Our main equilibrium concept is a *pure strategy perfect Bayesian [Nash] equilibrium*,

[10] The assumption that a firm can only go bad in a period in which it trades keeps the probability that the firm has gone bad from accumulating between trades. Since we are interested in small ε this technical assumption does not affect the intuition of the model. Alternatively, one could allow for a small chance in each period that the firm transitions back to being a normal firm.

[11] The pooling approach assumption that some types are "good" has its origins in finitely repeated games where, unlike in our infinitely repeated game, it is necessary to generate cooperation if the stage game has a unique equilibrium without cooperation (Kreps et al., 1982; Fudenberg and Levine, 1989). The assumption is the basis for the Cole and Kehoe (1998) analysis of reputation spillover.

i.e., in each continuation game the strategies for each player maximize payoffs given beliefs, and beliefs are consistent with strategies along the equilibrium path. We also consider a coalition-proofness refinement of such equilibria which allows for joint deviations by players that help those players at any point in the game. As discussed later, the probability ε of the firm going bad makes renegotiation of planned punishment strategies less attractive and thereby allows intuitive equilibria based on trigger strategies to survive coalition-proofness.

In the trust literature it is typical to concentrate on the no-trust strategy in which no individual ever trusts the firm and the (grim) trigger strategy in which trust stops if the firm ever cheats an individual. We refer to the standard trigger strategy as the solidarity trigger strategy and we define the discriminatory trigger strategy as the case where an individual trusts the firm if and only if the firm has never cheated anyone of her own type.

Definition 1 *Under the* no-trust strategy *an individual never trusts the firm.*

Definition 2 *Under the* solidarity trigger strategy *the individual trusts a firm if and only if the firm has never cheated anyone.*

Definition 3 *Under the* discriminatory trigger strategy *the individual trusts the firm if and only if the firm has never cheated anyone of the same type.*

We focus on equilibria that are *type-stationary* in that equilibrium strategies, while they might depend on the type of the individual, do not depend on other features of the game such as the period or sequence of play. Non-stationary equilibria can also exist, e.g., every third individual is cheated, but only on Tuesdays. In evaluating equilibria we will consider any possible deviations, but we will focus on equilibria that are type-stationary.

First considering the no-trust strategy, in the corresponding *no-trust equilibrium* we define a normal firm's strategy as to cheat any individual for any history. Clearly this is an equilibrium since individuals will never trust if they expect to be cheated by the firm, and the firm loses nothing from always cheating if it is never trusted. Regarding the solidarity trigger strategy, in the corresponding *solidarity-trust equilibrium* we define a normal firm's strategy as to be fair to every individual if no individual has ever been cheated, and to cheat every individual if an individual has ever been cheated. Since an individual trusts the firm if and only if it has never cheated and since treating the individual fairly maintains the fair reputation, the value V_s of a reputation for being fair when individuals follow the solidarity trigger strategy is $V_s = 1-\alpha+\delta V_s$ or $V_s = (1-\alpha)/(1-\delta)$, where recall that ε is suppressed since we are considering the limit as ε goes to zero. Individuals will not trade when a firm has a reputation for cheating, so the value to the firm of such a reputation is 0. The firm receives $1-\beta$ from cheating, so the discount factor δ_s such that a (normal) firm is just indifferent between being fair to and cheating an individual is given by $1-\alpha+\delta_s V_s = 1-\beta$ or

$$\delta_s = \frac{\alpha-\beta}{1-\beta}. \tag{1}$$

Since V_s is increasing in δ, the solidarity equilibrium exists if and only if $\delta > \delta_s$.[12]

Now suppose type p follows the discriminatory trigger strategy and type q, expecting to be cheated, follows the no-trust strategy. For the corresponding *q-discrimination equilibrium* we define a normal firm's strategy as to cheat any member of group q for any history, to be fair to any member of group p if a member of group p has never been cheated, and otherwise (off the equilibrium path) to cheat any member of group p. Let V_p be the value of a reputation for treating members of group p fairly. Since in each round there is a γ_p and a γ_q chance of encountering a member of group p or q respectively, and since members of group q do not trust, $V_p = \gamma_p(1-\alpha) + \gamma_q \cdot 0 + \delta V_p$ or $V_p = \gamma_p V_s$. Given this reputation value, the discount factor δ_p such that the firm is indifferent between between being fair to and cheating a q individual is given by $1-\alpha+\delta_p V_p = 1-\beta$ or

$$\delta_p = \frac{\alpha-\beta}{\alpha-\beta+\gamma_p(1-\alpha)}. \tag{2}$$

Since V_p is increasing in δ, this discrimination equilibrium exists if and only if $\delta > \delta_p$.

Finally consider the case where both types of individuals follow the discriminatory trigger strategy and a normal firm's strategy is to be fair to every individual if no individual has ever been cheated, and to cheat every individual if an individual has ever been cheated. Following the above logic, if $\delta > \delta_p$ for $p = x, y$ then each group is sufficiently large to deter opportunism even though it receives no help from the other group. We refer to this as the *independent-trust equilibrium.*

In a type-stationary equilibrium an individual either invests and is treated fairly or does not trust at all. Therefore any type-stationary equilibrium must be equivalent to one of the above equilibria in the sense that the outcomes are the same even if the off-equilibrium-path strategies differ.[13] Noting that the solidarity-trust equilibrium has the same equilibrium

[12] We use the strict rather than weak inequality because for any $\varepsilon > 0$ the firm has an incentive to cheat for $\delta = \delta_s$. For the same reason we use strict inequalities in other cutoffs defined below.

[13] For instance, it is equivalent to the solidarity equilibrium for individuals to only penalize the firm for a

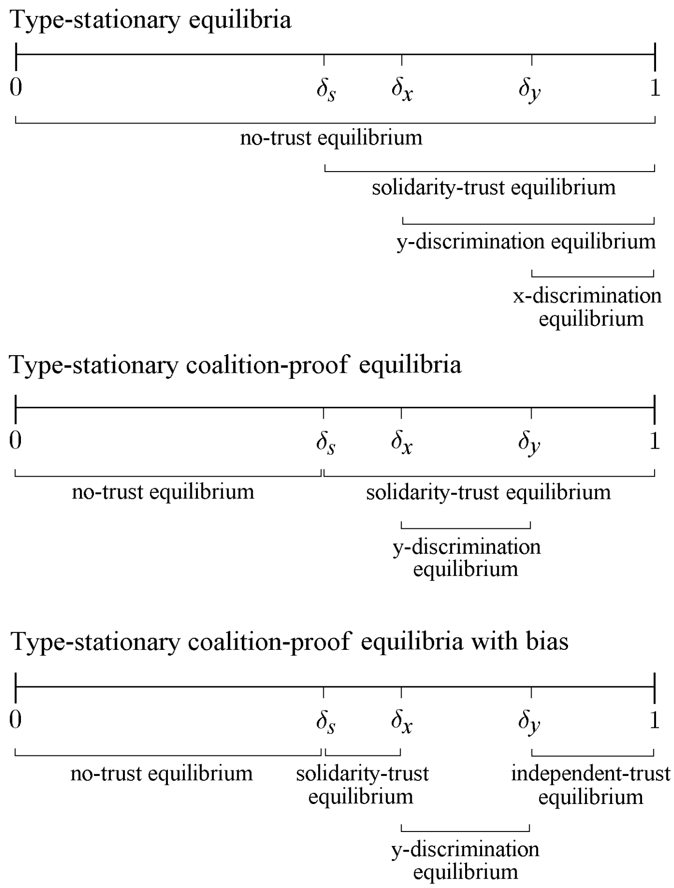

Figure 2: Equilibrium ranges for $\beta = \frac{1}{3}$, $c = \frac{1}{2}$, $\alpha = \frac{2}{3}$ and $\gamma_x = \frac{3}{4}$.

outcomes as the independent-trust equilibrium under our current assumptions, we find the following:

Proposition 1 *Any type-stationary equilibrium is equivalent to: i) the no-trust equilibrium if $\delta \in (0, \delta_s]$; ii) either the no-trust or solidarity-trust equilibrium if $\delta \in (\delta_s, \delta_x]$; iii) either the no-trust, solidarity-trust, or y-discrimination equilibrium if $\delta \in (\delta_x, \delta_y]$; and iv) either the no-trust, solidarity-trust, x-discrimination, or y-discrimination equilibrium if $\delta \in (\delta_y, 1)$.*

Proof: In the Appendix.

sufficiently long period that the firm does not cheat in equilibrium. In a noisy environment such strategies can outperform the trigger strategies we consider (e.g., Green and Porter (1984), but in our model there is no gain to limiting the punishment period.

Looking at the top section of Figure 2, in the range $\delta \in (\delta_s, \delta_x]$ the firm is relatively impatient so the two groups depend on each other to punish opportunism and trust by either group is only possible in the solidarity-trust equilibrium. However, for $\delta \in (\delta_x, \delta_y]$ the firm is patient enough that the majority is capable of dissuading opportunism against its members without the help of the minority, so discrimination becomes possible. For $\delta \in (\delta_y, 1)$ the firm is so patient that even the minority alone can dissuade opportunism against its members, so in this range an equilibrium also exists where the minority trusts the firm but the majority does not because its members do not coordinate on a strategy of punishing opportunism. Notice that δ_p is strictly decreasing in γ_p, so this pattern in the Figure that discrimination against the minority is supported by a wider range of discount factors must hold.

Corollary 1 *The range of δ supporting a p-discrimination equilibrium is larger than the range supporting a q-discrimination equilibrium iff $\gamma_p < \gamma_q$.*

As the minority population γ_y becomes smaller, δ_x falls toward δ_s while δ_y rises toward 1, so the range in which the y-discrimination equilibrium is the unique discrimination equilibrium increases to cover the entire range of the solidarity-trust, equilibrium. That is, as the population sizes become more different, the range $(\delta_s, \delta_x]$ under which the two groups depend on each other to achieve fairness and the range $(\delta_y, 1)$ under which they do not need each other at all both shrink, while the range $(\delta_x, \delta_y]$ under which only one group needs the other expands. Conversely, as the population sizes become more similar, the range $(\delta_x, \delta_y]$ shrinks and outside of this range either both groups have to rely on each other to dissuade opportunism or each group alone is large enough to dissuade opportunism.

While this result captures the basic insight of opportunistic discrimination, the Nash restriction to individual deviations in perfect Bayesian equilibria allows for some equilibria that seem less likely than others. For instance, for $\delta > \delta_s$ the solidarity equilibrium always coexists with the no-trust equilibrium and it offers strictly higher payoffs for every player. Since it is in the interest of everyone to collectively switch to the "good" equilibrium, it is often argued that the players should be able to talk their way to it. Similarly, for $\delta > \delta_y$, either group can stop opportunism against its members by following the discriminatory trigger strategy of only punishing opportunism against its own members. Since a discrimination equilibrium in this range arises only because the group following a no-trust strategy does not collectively switch to a trigger strategy even though they would all benefit, this equilibrium also seems less likely.

Such arguments cannot completely eliminate inefficient equilibria. Consider the y-discrimination equilibrium in the range $\delta \in (\delta_x, \delta_y]$. The solidarity equilibrium also exists in this range and both the y individuals and the firm are better off in it. However, the difference from the previous cases is that the y's and the firm alone cannot induce a change to this better equilibrium, but are dependent on the x's changing their strategy to the solidarity strategy. Since the x's are already doing as well as they can in the current equilibrium, it is unclear why they would switch.[14] This highlights a key difference between the position of the minority and majority. Since the minority interacts with the firm infrequently they can escape the discrimination equilibrium only if the majority adopts the solidarity strategy or if the firm is very patient. In contrast, the majority interacts with the firm frequently enough that it is in the interest of the firm to treat them fairly based on the punishment strategy of the majority alone.

This idea that an equilibrium should be ruled out if a coalition of players can attain higher payoffs through a joint deviation appears in different refinements of Nash equilibria, most notably in the different notions of coalition-proofness (e.g., Bernheim et al., 1987; Bernheim and Ray, 1989).[15] To capture the effect of joint deviations in the simplest way, we say a perfect Bayesian equilibrium (or just "equilibrium") is *coalition proof* if there does not exist in any period and any history another perfect Bayesian equilibrium for the continuation game attainable by a joint deviation of a subset of players such that every player in the subset expects to be strictly better off given their beliefs where these beliefs are consistent along the equilibrium path. Note that any deviation must be to an equilibrium, but this equilibrium need not itself be coalition proof.[16] For our game the relevant deviation in each case is to an equilibrium which is itself immune to any deviations by any coalition to another equilibrium, so "weaker" notions of coalition-proofness would also allow such

[14]Not only do they gain nothing from switching to the new equilibrium, they lose all gains to trade if there is a "miscommunication" and the firm does not change its strategy and cheats the minority, or if there is any uncertainty over whether or not the minority was really cheated.

[15]Early approaches to joint deviations, including strong equilibria (Aumann, 1959) and strong perfect equilibria (Rubinstein, 1978) did not require that deviations be immune to further deviations. The various definitions of coalition-proofness in the literature require that deviations be immune to further deviations to some degree, but there is no consensus on which definitions are most appropriate when. As discussed below, the incentive for further deviations does not arise in our context so these differences are not relevant for our analysis.

[16]Milgrom and Roberts (1996) refer to this concept as "strong coalition-proofness" since more deviations are potentially allowed, thereby potentially eliminating more equilibria. Since this definition of coalition proofness is not recursive it can be applied directly to our infinitely repeated game.

deviations.

As we show in the following proposition, coalition-proofness limits the multiplicity of equilibria in Proposition 1 in accordance with the above discussion. Note that since coalition-proofness allows a coalition to be formed at any period, it incorporates the possibility of renegotiating a planned punishment strategy following unexpected opportunism by the firm.[17] Allowing for a small probability ε that the firm becomes inept implies that individuals interpret unexpected cheating as a negative signal about the firm, thereby ensuring that trigger strategies are credible.[18]

Proposition 2 *Any coalition-proof, type-stationary equilibrium is payoff equivalent to: i) the no-trust equilibrium if $\delta \in (0, \delta_s]$; ii) the y-discrimination equilibrium or the solidarity-trust equilibrium if $\delta \in (\delta_x, \delta_y]$; and iii) the solidarity-trust equilibrium if $\delta \in (\delta_s, \delta_x] \cup (\delta_y, 1)$.*

Proof: In the Appendix.

Looking at the middle section of Figure 2, in the lower range $\delta \in (\delta_s, \delta_x]$ the firm is still relatively impatient, so any division between the individuals will make everyone worse off. Therefore the whole coalition of individuals can successfully adopt a punishment strategy that induces fairness by the firm, but not any one group alone. For the higher range $\delta \in (\delta_y, 1)$ the firm is sufficiently patient that if either group switches from the no-trust strategy to either the discriminatory trigger strategy or the solidarity-trust strategy the firm has an incentive to be fair to them regardless of what individuals in the other group do, so any equilibrium is payoff equivalent to the solidarity-trust equilibrium. Only in the range $\delta \in (\delta_x, \delta_y]$ is discrimination possible, and it must be directed against the minority.

The following corollary of Proposition 2 follows.

Corollary 2 *A coalition-proof p-discrimination equilibrium exists for some δ iff $\gamma_p < \gamma_q$.*

Whether coalition-proofness is an appropriate refinement presumably depends on the details of the situation such as the number of different players, the ease of communication, the

[17] The coalition-proofness literature generalizes the renegotiation literature following Farrell and Maskin (1989) which was for two players.

[18] Farrell and Weizsacker (2001) show that the standard trigger strategy in a trust game with complete information is not renegotiation-proof. Moreover, unlike the case of the repeated prisoner's dilemma (van Damme, 1989), there does not exist a more complicated equilibrium strategy that is payoff-equivalent or nearly so. Farrell and Weizsacker analyze a game between two players but the same issues arise in our game with multiple individuals.

willingness of players to change strategies, and the history of play in related contexts. Even when we require equilibria to be coalition-proof, the solidarity-trust and $y-$discrimination equilibria coexist in the range $\delta \in (\delta_x, \delta_y]$ because a switch to the solidarity-trust equilibrium means that x's have to agree to punish opportunism against the y's even though they themselves can only lose from lost trade with the firm. We now show how this problem of getting x's to punish opportunism against y's is exacerbated if firms are with some small probability biased.

2.1 Renegotiation and bias

Following Becker (1957), a large literature examines discrimination when firms have a preference-based bias. To integrate this approach with that of opportunistic discrimination, consider how equilibrium behavior is affected when there is some chance that the firm is not profit-maximizing but instead has a bias against members of one group. We model this as a small independent probability ϕ_p that the firm has a preference to cheat group p. We look at the case where the probability of bias is low enough that each group will still trust the firm if "normal" firms who are neither inept nor biased do not cheat, i.e., $(1-\varepsilon)(1-\phi_p)(\alpha - c) \geq (\phi_p + \varepsilon - \phi_p\varepsilon)(c-\beta)$. Since we are interested in the limiting case where ε goes to zero, we therefore assume $0 < \phi_p < (\alpha - c)/(\alpha - \beta)$ for $p = x, y$. Notice that we are assuming that there is some chance of the firm being biased against either group, not just against the minority.

The potential for such bias does not change the set of perfect Bayesian equilibria identified in Proposition 1, but it does effect the set of equilibria that satisfy coalition-proofness. After cheating a member of one group the firm might want to persuade members of the other group that they will not meet the same fate so they should still trust the firm. This is a problem for the minority in the range $\delta \in (\delta_x, \delta_y]$ because they are dependent on the majority to credibly threaten to punish the firm for opportunism against anyone, including the minority. If the majority believes that the firm is biased against the minority rather than inept, it has an incentive to the give the firm another chance. In contrast, the majority is only dependent on the minority in the range $\delta \in (\delta_s, \delta_x]$ where both groups depend on each other, so the minority does not benefit from forgiving opportunism against the majority. Therefore, even though we allow for bias against either group, the effect of bias is always against the minority.

Proposition 3 *With a potentially biased firm, any coalition-proof type-stationary equilibrium is equivalent to: i) the no-trust equilibrium if $\delta \in [0, \delta_s]$; ii) the solidarity-trust equilibrium if $\delta \in (\delta_s, \delta_x]$; iii) the y-discrimination equilibrium if $\delta \in (\delta_x, \delta_y]$; and iv) the independent-trust equilibrium if $\delta \in (\delta_y, 1)$.*

Proof: In the Appendix.

This proposition shows that preference-based bias among a small proportion of players can have a large impact in choosing between multiple equilibria.[19] If the majority starts with a solidarity strategy of punishing opportunism against the minority, not only will biased firms still cheat the minority, but unbiased firms have an incentive to pool with biased firms and thereby gain the short-term benefits of cheating while maintaining the long-term benefits of a good reputation with the majority.

This pooling incentive to mimic biased firms contrasts with the intuition of political correctness (Morris, 2001) in which unbiased advisors maintain their reputation by sometimes refraining from expressing opinions which happen to match those of biased advisors. Similar incentives might arise in this model by allowing for uncertainty over whether an act of opportunism occurred. For instance a very patient firm which is not discriminatory might try to avoid even the appearance of being opportunistic against the minority for fear of damaging its reputation.[20]

2.2 Contractual and legal constraints on opportunism

To see how contractual and legal constraints on opportunism can interact with reputation effects, and how anti-discrimination laws can affect opportunistic discrimination, we consider three enforcement strategies against opportunism. First, the government can pursue one-sided enforcement that selectively discourages opportunism against one group. Second, the government can more rigorously enforce contracts and laws against opportunism in general, narrowing the range for all opportunism, discriminatory or not. Third, the government can pursue anti-discrimination enforcement which penalizes opportunism against a member of any group if and only if the firm is also fair toward a member of another group.

[19] In an assurance game Basu (2005) similarly finds that a fraction of biased types who do not cooperate can induce non-biased types to also play the non-cooperation strategy, though the incentive is primarily defensive rather than opportunistic.

[20] As Smith (1766) stated, "...a prudent dealer, who is sensible of his real interest, would rather choose to lose what he has a right to than give any ground for suspicion."

To analyze this problem we use the result from Proposition 2 on the set of coalition-proof equilibria, though the general insights still hold if we think about the potential for discrimination against the minority but not the majority in the range $\delta \in (\delta_x, \delta_y]$ from Proposition 1.

Let π_p be the penalty imposed on a firm if it engages in opportunism against group p. This penalty could be for breaking a private contract or for breaking laws against opportunism. We assume that $\pi_p < \alpha - \beta$ so the penalty does not simply eliminate all opportunism against p but rather allows for some interaction between the penalty and reputation effects. Noting that V_s is unchanged from the base model, and that the marginal firm will cheat an individual for which the penalty is lowest, the cutoff for the solidarity equilibrium is δ_s^π such that $1 - \alpha + \delta_s^\pi V_s = 1 - \beta - \min\{\pi_x, \pi_y\}$, or

$$\delta_s^\pi = \frac{\alpha - \beta - \min\{\pi_x, \pi_y\}}{1 - \beta - \min\{\pi_x, \pi_y\}}. \tag{3}$$

Regarding discrimination equilibria, V_p is also unchanged from the base model, so when p individuals follow the discriminatory trigger strategy and q individuals follow the no-trust strategy, the cutoff discount factor for cheating p is δ_p^π such that $1 - \alpha + \delta_p^\pi V_p = 1 - \beta - \pi_p$, or

$$\delta_p^\pi = \frac{\alpha - \beta - \pi_p}{\alpha - \beta + \gamma_p(1 - \alpha) - \pi_p} \tag{4}$$

We can now use these cutoffs just as in the previous analysis. In particular, it is straightforward to show that the ranges for coalition-proof equilibria will be the same as given in Proposition 2, except with these penalty-adjusted cutoffs.

When enforcement is selective it is often aimed at protecting the majority rather than minority ($\pi_x > 0$, $\pi_y = 0$), e.g., foreigners in many countries have limited access to the legal system to enforce contracts, and in some countries women are still unable to sign binding contracts. Such enforcement would seem to only benefit the majority, but in fact it can hurt the minority by reducing the dependence of the majority on reputational sanctions against opportunism. To see this, note that an increase in π_x decreases δ_x^π but does not have an impact on δ_s^π or δ_y^π so the lower solidarity equilibrium region $(\delta_s^\pi, \delta_x^\pi]$ shrinks and the y-discrimination region $(\delta_x^\pi, \delta_y^\pi]$ increases. Since the majority is better able to protect itself without relying on a solidarity strategy with the minority, the minority is made more vulnerable to opportunism. Therefore enforcement is not just a substitute for reputation, but undermines reputation so much that there is a net loss in trade. Only if π_x is higher than $\pi_x^* = (\alpha - \beta)\gamma_y$, which is the point in Figure 3 where $\delta_x^\pi(\pi_x) = \delta_s^\pi(0)$, does δ_x^π become smaller than δ_s^π, in which case the majority is better off and the minority is not hurt.

In some cases the policy response to discrimination might involve selective enforcement that is targeted at opportunism against the minority ($\pi_x = 0$, $\pi_y > 0$). This can eliminate the y-discrimination equilibrium if $\delta \in (\delta_y^\pi, \delta_y]$ and it cannot cause a switch into the y-discrimination region since it does not affect δ_x^π. However, it too can be counterproductive. If π_y is higher than

$$\pi_y^* = (\alpha - \beta)\frac{\gamma_x - \gamma_y}{\gamma_x}, \tag{5}$$

which is the point in Figure 3 where $\delta_y^\pi(\pi_y) = \delta_x^\pi(0)$, then $\delta_y^\pi < \delta_x$ so a reverse discrimination equilibrium becomes possible if $\delta \in (\delta_y^\pi, \delta_x]$. Since the solidarity equilibrium would only survive in the region $\delta \in (\delta_s, \delta_x]$, this is a net loss for $\delta \in (\max\{\delta_s, \delta_y^\pi\}, \delta_x]$. In Figure 3 this is the point π_y^* where $\delta_y^\pi(\pi_y) = \delta_x^\pi(0)$.

Now consider general enforcement against opportunism where the penalties are the same, $\pi_x = \pi_y = \pi > 0$. Since δ_s^π and δ_p^π are decreasing in π, general enforcement decreases all the cutoffs as seen in Figure 3. If $\delta \in (\delta_s^\pi, \delta_s]$ then general enforcement induces a switch from the no-trust region to a region where only the solidarity equilibrium survives. And if $\delta \in (\delta_y^\pi, \delta_y]$ then it induces a switch from the y-discrimination region to a region where only the solidarity equilibrium survives. But if $\delta \in (\delta_x^\pi, \delta_x]$ then general enforcement perversely induces a switch from a region where only the solidarity equilibrium survives to the region where the y-discrimination equilibrium survives. Without any enforcement the x individuals would have to follow the solidarity strategy to avoid the no-trust equilibrium, but the combination of reputation and explicit sanctions make it possible for the x individuals to follow the discriminatory trigger strategy and not be cheated.

The final option is to penalize opportunism against one group if and only if the firm is also fair towards the other group. This "anti-discrimination enforcement" has the same effect as enforcement targeted at opportunism against the minority at decreasing δ_y^π,[21] except that reverse discrimination cannot result even if the penalty is higher than π_y^* because discrimination against y's will also be penalized. Therefore the original y-discrimination region disappears, an x-discrimination region is not created, and only the solidarity equilibrium survives for $\delta > \delta_s$. Comparing these different enforcement strategies, we have the following result.

[21] In the case where the firm first cheats an individual from one group and is subsequently fair to an individual from another group, the penalty should be imposed with interest to main the exact correspondence.

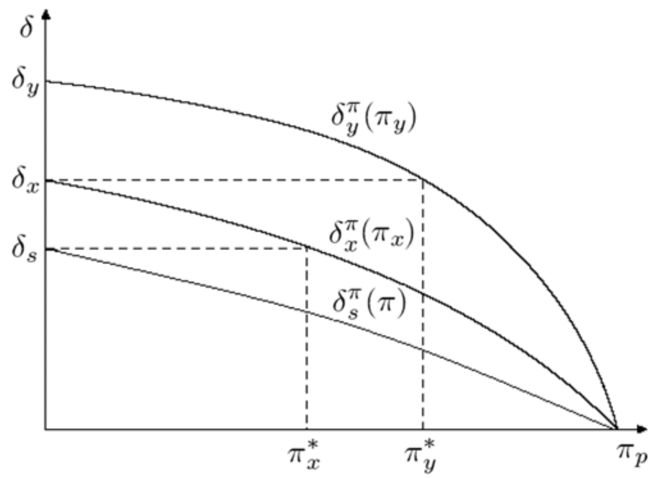

Figure 3: General, one-sided, and anti-discriminatory enforcement.

Proposition 4 *Anti-discrimination enforcement is the only enforcement strategy that never allows for increased opportunism in a coalition-proof type-stationary equilibrium.*

This analysis adds to the long-standing debate on whether legal and reputational sanctions against opportunism are substitutes or complements (e.g. Ostrom, 2000; Poppo and Zenger, 2002; Lazzarini et al., 2004). Anti-discrimination enforcement is a complement to reputation since it makes punishment strategies more effective at stopping opportunism over a larger set of discount factors, but the other forms of enforcement can make reputation more or less effective depending on the parameter values. Of particular relevance for understanding discrimination, stronger enforcement that protects the majority never helps the minority and often makes it worse off, so from the perspective of the minority such enforcement is worse than just being a substitute for reputation.

2.3 Self-segregation and self-integration

The model assume that individuals appear before a single firm according to a stochastic mechanism beyond their control, but if there are multiple firms in a market then individuals that anticipate being cheated at one firm might choose to go to another firm. In particular, individuals who are in the minority at different firms might find it worthwhile to concentrate their business in one firm where their numbers are sufficient to avoid opportunism. This could represent "self-segregation" where each group concentrates its business on firms which do not do business with the other group, or it could represent "self-integration" where both groups concentrate their business on firms which do business with both groups.[22]

To investigate this issue we assume that there are $m > 1$ firms with discount factors δ^i, $i = 1, 2, \ldots, m$. Each firm i does business with $n_p^i \leq n_p$ members of each group where $\sum_{i=1}^m n_p^i = n_p$ and $\sum_{p=x,y} \sum_{i=1}^m n_p^i = n$. Following the same approach as the single firm case, in any period the probability that firm i does business with a member of group p is $\gamma_p^i = n_p^i / n$. To simplify the analysis we again focus on the result from Proposition 3 on coalition-proof equilibria when the firm is potentially biased. Following the same calculations for the cutoffs as in (1) and (2), the new cutoffs for firm i are

$$\delta_s^i = \frac{\alpha - \beta}{\alpha - \beta + (\gamma_x^i + \gamma_y^i)(1 - \alpha)}$$

and

$$\delta_p^i = \frac{\alpha - \beta}{\alpha - \beta + \gamma_p^i (1 - \alpha)}.$$

The following proposition shows that, unless a firm is known to be biased, individuals are never better off from self-segregation than from "full self-integration" where all individuals patronize the same firm. If the minority is distributed across different firms and their numbers at each firm make the cutoff δ_p^i too high to ensure ensure fair treatment, then minority individuals can sometimes benefit from joining other minority members at the same firm. However, in this case the minority does just as well by patronizing a firm which is also patronized by the majority, including the case of full self-integration. And if the minority is insufficient to ensure fair treatment even when it is fully concentrated on one firm, then it is no worse off, and sometimes better off, under full self-integration since the cutoff δ_s^i might be sufficiently low to allow for the solidarity equilibrium.[23] Therefore, as proven in the Appendix, the following proposition follows.

Proposition 5 *Full self-integration offers equal or better expected outcomes for all individuals than self-segregation in a coalition-proof, type-stationary equilibrium.*

[22] For simplicity this paper assumes that the gains from trade are fixed so we do not investigate the role of competition between firms. Competition increases the costs of a bad reputation (Horner, 2002), but also reduces the rents from a good reputation, so its effects on trust are ambiguous (Bar-Isaac, 2005). The effect on opportunistic discrimination is therefore likely to be sensitive to the exact environment.

[23] Sometimes partial integration is better for the minority than full integration. For instance, suppose there are two firms and the combined size of the majority and minority puts firm 1 in the solidarity equilibrium range, $\delta \in (\delta_s^1, \delta_x^1]$. If firm 2 has any majority members and the firms are combined then $\delta_x < \delta_x^1$, so if $\delta \in (\delta_x, \delta_x^1]$, then full integration newly exposes the minority in firm 1 to the discrimination equilibrium.

Preference-based models of discrimination following Becker (1957) predict that workers self-segregate into different firms. In our model of opportunistic discrimination workers benefit from avoiding firms that are known to be biased, but otherwise there is strength in numbers, even when the numbers involve members of another group. The gains from self-integration may offer insight into the potential for fairer treatment by large corporate employers and retailers than by smaller proprietorships and partnerships.[24] Because of more frequent business, a large firm might be fair to all individuals when a small firm is not fair to even a subpopulation which it specializes in serving. For instance, a large retailer like Wal-Mart may be more likely than a small local store to replace a defective product from a minority customer, even if that store's customers are primarily minorities.

2.4 Reputation unravelling with uncertain discount factors

So far we have considered the standard repeated trust game where the firm's discount factor δ is common knowledge, except for the small probability that the firm is so inept or myopic that it always cheats. But in many cases individuals have to decide whether to trust the firm while facing substantial uncertainty over how patient the firm is. This can be a problem for the minority in particular since even relatively patient firms might still discriminate against them. To gain insight into this issue we assume that it is common knowledge that the firm's discount factor is distributed according to the distribution $F(\delta)$ but the exact value is the firm's private information. We are interested in the case where this distribution is relatively favorable so that it is an equilibrium for individuals to trade with the firm until they learn from the firm's actions that the firm is too impatient to be trusted. For simplicity we assume $\varepsilon = 0$ and $\phi_p = 0$ in this final section.

Regarding the no-trust and solidarity-trust equilibria, the situation is little different from the standard case. Clearly if $F(\delta_s)$ is high enough the no-trust equilibrium will be the only equilibrium, and if $F(\delta_s)$ is low enough the solidarity-trust equilibrium will exist. In particular it is an equilibrium for individuals to follow the solidarity trigger strategy if the expected gain from the firm being fair when $\delta \geq \delta_s$ is enough to compensate for the expected loss from being cheated when $\delta < \delta_s$, i.e., if $(1 - F(\delta_s))(\alpha - c) \geq F(\delta_s)(\beta - c)$. In this equilibrium firms with $\delta < \delta_s$ will immediately reveal themselves by cheating, while firms with $\delta \geq \delta_s$ will be fair.

[24] The idea that integration facilitates trust by creating a higher cost of cheating is related to the idea that multi-market contact between firms can sometimes facilitate collusion (Bernheim and Whinston, 1990).

The interesting case arises regarding discrimination. Our previous calculation of δ_p still reflects the differential incentive to cheat the minority. Since individuals do not know δ, the minority will trade with the firm that has no history of interaction with the minority if the expected payoff is positive, i.e., if $(1 - F(\delta_y))\,(\alpha - c) \geq F(\delta_y)(\beta - c)$. The majority makes a similar calculation but there is an additional concern. Once a firm has cheated a y individual the value of its reputation has been diminished because y individuals will no longer trade with it. Therefore a firm might first be fair to an x, then cheat the first y and then cheat the next x that is encountered. By spending down its reputation in this way, the firm can cheat twice rather than only once.

The following proposition shows that this strategy of "double-crossing" the x individuals can be an equilibrium strategy for all firms $\delta \in [\delta_d, \delta_x)$ for some $0 < \delta_d < \delta_s$ where the exact value of δ_d is derived as part of the proof in the Appendix. The only requirements are that the distribution of firms is sufficiently favorable that x and y individuals are willing to trade initially and that x individuals are still willing to trade after observing a y individual being cheated.

Proposition 6 *If $F(\delta_y), F(\delta_x)/F(\delta_y) \leq (\alpha - c)\,/\,(\alpha - \beta)$ then there exists an equilibrium such that x's follow the discriminatory trigger strategy, y's follow the solidarity trigger strategy, and i) the firm cheats everyone for $\delta \in (0, \delta_d)$, ii) the firm treats all x's fairly until it encounters a y whereupon it cheats everyone for $\delta \in [\delta_d, \delta_x)$, iii) the firm treats all x's fairly and cheats all y's for $\delta \in [\delta_x, \delta_y)$, and iv) the firm treats everyone fairly for $\delta \in [\delta_y, 1)$.*

Regarding the effect of group sizes, since δ_x converges to δ_y as $\gamma_x - \gamma_y$ goes to zero, and since $(\alpha - c)\,/\,(\alpha - \beta) < 1$, the last condition implies that the difference in group sizes must be sufficiently large for the equilibrium to exist. Also since $\delta_x < \delta_y$ and hence $F(\delta_y)/F(\delta_x) > 1$, the condition implies that the equivalent equilibrium in which the majority is discriminated against does not exist. Hence the general pattern is consistent with that of Proposition 2 even though we have not restricted our analysis to coalition-proof equilibria.

Recall that in the standard model all firms are better off in the solidarity equilibrium than in the no-trust or discrimination equilibria. Although firms in the discriminatory region $[\delta_x, \delta_y)$ are better off ex post from discrimination if a y individual trusts them and x individuals continue to trade with them, in equilibrium y individuals do not trust the firm. In this extension where there is real uncertainty over the firm's discount factor, y individuals will initially trust the firm in the discrimination equilibrium, so firms which cheat them benefit both ex ante and ex post from the discrimination equilibrium relative to the solidarity equilibrium. Firms in the region $[\delta_d, \delta_x)$ benefit even more since they can cheat twice, so it is only firms in the region $[\delta_y, 1)$ that do not benefit from the discrimination equilibrium.

This model offers insight into how failure to stand up to opportunism against the minority can eventually endanger even the majority.[25] Once the firm has cheated the minority and lost γ_y share of its business, the value of its reputation is devalued because cheating the majority will now result in only a loss of the remaining γ_x share of business, rather than of all business if it had cheated the majority from the beginning. Once opportunism against anyone begins, it becomes more likely against everyone.

3 Conclusion

In a standard repeated trust game, we find that the minority is more susceptible to opportunism than the majority. Such opportunistic discrimination does not require discriminatory preferences nor differences in individual attributes. Rather it follows from the simple fact that the minority is by definition smaller so trade with the minority is correspondingly less frequent. Long-standing theories about trust and reputation dating back to Smith (1766) then imply that there is less value in a reputation for honesty toward the minority, so there is correspondingly less incentive to forego the short-term gains of cheating them. Reputation spillover can protect the minority if the majority is expected to follow a solidarity strategy of punishing opportunism against anyone, but the majority does not gain from such a strategy unless it is also so small that both groups depend on each other to dissuade opportunism.

Any population can be divided in a myriad of ways, such as gender, ethnicity, race, language, caste, religion, etc. From the perspective of our analysis, whether a particular division affects trust depends on how people are expected to react to opportunism, so the question is what divisions are "focal" for historical or other reasons. Clearly one possibility is that preference-based biases among some players might make particular divisions focal. When a firm might have a discriminatory preference to cheat the minority based on some division, we find the stronger result that such biases interact with reputational concerns

[25] The majority would like to commit to a solidarity strategy, but once a firm has cheated the minority the majority can still be better off in expectation from renegotiating with the firm than from stopping all trade. We do not formally model renegotiation in this section.

to make discrimination the unique coalition-proof equilibrium for firms of intermediate patience. Hence the existence of such bias in a society, even if not widespread, may help explain why some divisions have a large effect on trust and others do not.

Appendix

Proof of Proposition 1: The text shows that any type-stationary perfect Bayesian equilibria [PBE] is outcome-equivalent to one of the indicated equilibria for the given ranges of δ. We now confirm that the given equilibria are PBE by examining whether any player benefits from changing their strategies in reaction to an observed deviation.

We begin with the no-trust equilibrium. Individuals follow the no-trust strategy. Given the opportunity, the firm always cheats. Take any possible history. Since individuals believe that the firm will always cheat, trust yields a payoff of $\beta - c < 0$ but a payoff of 0 if they do not trust. Thus individuals cannot gain by trusting the firm. On the other hand, since individuals never trust, it is a best response for the firm to cheat the individual.

Next consider the solidarity-trust equilibrium. All individuals follow the solidarity strategy. The firm always treats the individual fairly as long as the firm has never cheated. Take any possible history. If the history involves only trust and no cheating, the individual expects the firm to treat her fairly so the short-run payoff to an individual who trusts is $\alpha - c > 0$ and the payoff to an individual who does not trust is 0. Thus the individual prefers to trust the firm. For the firm, treating the individual fairly yields payoff $1-\alpha+\delta V_s$ and cheating the individual yields payoff $1 - \beta$. Whenever $\delta > \delta_s$, $1 - \alpha + \delta V_s > 1 - \beta$ so the firm is better of treating the individual fairly. Whenever $\delta \leq \delta_s$, $1 - \alpha + \delta V_s < 1 - \beta$ so the firm is better off cheating the individual. If instead, the firm has cheated then the individuals and the firm play the "no-trust equilibrium" and the above argument holds.

Finally consider the p-discrimination equilibrium. When the firm faces an individual from population p, the argument is the same as under the no-trust equilibrium – the firm has no incentive to treat a p type individual fairly as it expects no trust in the future and a p type individual never trusts the firm because she expects to be cheated. When the firm faces an individual from population q, if the history involves no cheating between the firm and individuals from population q, for the individual, the argument is precisely the same as for the solidarity equilibrium. For the firm, treating the type q individual fairly yields payoff V_q and cheating yields $1 - \beta$. Whenever $\delta > \delta_q$ the firm prefers to treat the q individual fairly, but if $\delta \leq \delta_q$ then $1 - \alpha + \delta V_q < 1 - \beta$ so the firm is better off cheating the individual. If the history involves cheating, then the firm and all individuals switch to the "no-trust equilibrium" and the same argument holds. ■

Proof of Proposition 2: Proposition 1 gives the set of type-stationary PBE. We now show which of these equilibria are also coalition proof.

i) For $\delta \in (0, \delta_s]$ from Proposition 1 the only PBE is the no-trust equilibrium. Similarly for any history the only PBE of any subgame is also the no-trust equilibrium, so there is no deviation to another PBE which can ever improve on the no-trust equilibrium for any set of players. Therefore the no-trust equilibrium is coalition proof.

ii) For $\delta \in (\delta_x, \delta_y]$ from Proposition 1 the no-trust, solidarity, and y-discrimination equilibria are the only type-stationary PBE. The no-trust equilibrium is not coalition-proof since all players are better off switching to the solidarity equilibrium. Regarding the solidarity equilibrium, the only deviation in any subgame that can benefit any player is if the firm cheats. But if a firm ever deviates and cheats, then the only consistent belief for individuals is that the firm has become inept, so individuals have no incentive to change their equilibrium strategy of not trusting the firm after any cheating, and the firm is worse off. Therefore there does not exist in any subgame a deviation to a PBE that improves on the solidarity equilibrium for any player, implying the solidarity equilibrium is coalition proof. Now consider the y-discrimination equilibrium. First, just as in the solidarity equilibrium, cheating an x will lead individuals to infer the firm has become inept so based on cheating of an x there does not exist in any subgame a deviation to a PBE that improves on the discrimination equilibrium for any of players. Now consider deviations that involve trust between the firm and some or all of the y's. Since x's can never improve on their payoffs by deviating from the discrimination strategy, this coalition does not involve any x's. But if x's stick with the discrimination strategy, we know that for $\delta \leq \delta_y$ the firm always benefits from cheating a y individual in any subgame. Therefore there is no deviation to a PBE that improves on the discrimination equilibrium for those players who deviate, so the equilibrium is coalition proof.

iii-a) For $\delta \in (\delta_s, \delta_x]$ from Proposition 1 the no-trust and solidarity equilibria are the only type-stationary PBE. By the same arguments as in ii) the former is not coalition proof and the latter is.

iii-b) For $\delta \in (\delta_y, 1)$ from Proposition 1 the no-trust, solidarity, and the two discrimination equilibria are the only type-stationary PBE. By the same arguments as above the no-trust equilibrium is not coalition proof and the solidarity equilibrium is. Both discrimination equilibria also exist but they are not coalition proof since the discriminated group and the benefit from a switch to the independent-trust equilibrium. In this equilibrium the individuals do the best they can and the firm does better than from cheating as long as $\delta > \delta_x, \delta_y$. ■

Proof of Proposition 3: Under the assumption $0 < \phi_p < (\alpha - c)/(\alpha - \beta)$ for $p = x, y$, individuals of type p will initially trust a firm if a normal firm's strategy is not to cheat. If a firm unexpectedly cheats a member of group p, then it could be that the firm is a biased firm or an inept firm. We are interested in how this affects the set of coalition-proof PBE identified in Proposition 2. Since we are still interested in the limiting case as ε approaches zero, the different cutoffs are the same, and the only difference is with respect to beliefs after histories that are off-the-equilibrium-path for normal firms.

i) For $\delta \in (0, \delta_s]$, the equilibrium set is unaffected since for any history it is still a dominant strategy for a normal firm to cheat any individual.

ii) For $\delta \in (\delta_x, \delta_y]$, first consider the solidarity equilibrium. If the firm cheats the first y then, since $\phi_y(\alpha - \beta) > \varepsilon(c - \beta)$ as ε goes to zero, the probability that the firm is biased against y's rather than inept is high enough that the x's still have an incentive to renegotiate away from the solidarity equilibrium to the y-discrimination equilibrium. This always holds as ε goes to zero. Anticipating this, a firm that is neither inept nor biased against y's gets a payoff of $1 - \beta + \delta V_x$ from such a deviation. The firm will cheat the y whenever this is greater than $1 - \alpha + \delta V_s$. But after rearranging, this inequality is equivalent to $1 - \beta > 1 - \alpha + \delta\gamma_y V_s = 1 - \alpha + \delta V_y$, which holds whenever $\delta < \delta_y$ (equation (2)). Since all firms will therefore pool and cheat the y, the solidarity equilibrium is not coalition proof.

Regarding the y-discrimination equilibrium, suppose the firm unexpectedly cheats an x. Members of population x can only conclude that the firm has become inept so it is a dominant strategy for each x to not trust the firm in the future, so there is no improving joint deviation to a PBE, and the y-discrimination equilibrium is coalition proof.

iii-a) For $\delta \in (\delta_s, \delta_x)$, first consider the solidarity equilibrium. If a firm cheats the first y then , since $\phi_y(\alpha - \beta) > \varepsilon(c - \beta)$ as ε goes to zero, x individuals have an incentive to form a coalition with the firm and continue trading. However, y types have a dominant strategy of not trusting, and as shown in Proposition 1, if $\delta < \delta_x$ then the firm will cheat any x individual in any subgame if x individuals alone can punish the firm, so there is no improving joint deviation to a PBE. Since $\delta > \delta_s$, the firm will never deviate in any other fashion (cheat a subsequent y or cheat an x) because all individuals will conclude that the firm is inept and never trust again. The no-trust equilibrium, in any history with no cheating, can be improved on by a joint deviation of all players to the solidarity equilibrium and is therefore not coalition-proof.

iii-b) For $\delta \in (\delta_y, 1)$, first consider the solidarity equilibrium. If the firm cheats the first y it encounters then, since $\phi_y(\alpha - \beta) > \varepsilon(c - \beta)$ as ε goes to zero, x individuals have an incentive to form a coalition with the firm and continue trading. However, while the firm benefits relative to no trade in the subgame, the firm is better off for $\delta > \delta_p$ by not deviating in the first place so the deviation cannot be part of an improving joint deviation to a PBE.

Now consider a p-discrimination equilibrium. In either case, since $\delta > \delta_y > \delta_x$ there is an improving joint deviation by the p's and the firm to a PBE where the p's adopt either the solidarity strategy or the discriminatory trigger strategy, either of which is payoff equivalent to the solidarity equilibrium. Finally, the no-trust equilibrium can be improved on by a joint deviation of all players to the solidarity equilibrium. ■

Proof of Proposition 5: Under full integration all individuals patronize the same firm, in which case the equilibrium ranges are as given by Proposition 3. A necessary condition for another allocation to be better for some individual for some δ is there exists a firm i such that either (i) $\delta_s^i < \delta_s$, (ii) $\delta_x^i > \delta_x$, or (iii) $\delta_y^i < \delta_y$. In the first case the solidarity region expands downward, in the second case the discrimination region contracts from below, and in the third case the discrimination region contracts from above. Checking these possibilities:

i) Since $\delta_s = (\alpha - \beta) / (\alpha - \beta + (1 - \alpha))$, $\delta_s^i = (\alpha - \beta) / \left(\alpha - \beta + \gamma^i(1 - \alpha)\right)$, and $\gamma^i \leq 1$, this is impossible.

ii) This holds if $\gamma_x^i < \gamma_x$. For the equilibrium to be self-segregating it must be that $\gamma_x^i = 0$. But in this case $\delta_s^i \geq \delta_y$, implying that even though the region of discrimination against y has contracted (in fact, disappeared) the solidarity region has contracted so much that, even if $\delta \in (\delta_x, \delta_y)$ so that there is room for a better outcome than under full integration, it must be that $\delta < \delta_s^i$, implying the no-trust equilibrium is the unique equilibrium for firm i and no one benefits relative to full integration.

iii) Since $\delta_y = (\alpha - \beta) / \left(\alpha - \beta + \gamma_y(1 - \alpha)\right)$, $\delta_y^i = (\alpha - \beta) / \left(\alpha - \beta + \gamma_y^i(1 - \alpha)\right)$, and $\gamma_y^i \leq \gamma_y$, this is impossible. ■

Proof of Proposition 6: To fully specify strategies assume that following any deviation from the equilibrium path the firm always cheats and no individual trusts and the value to the firm of going off the equilibrium path is 0. We will show that under any possible history, it is an equilibrium for the firm and for each individual to play according to the specified strategies. Every history can be categorized as either on-the-equilibrium-path or off-the-equilibrium-path. On-the-equilibrium-path histories induce particular beliefs. Given the strategies of the candidate equilibrium, individuals can have four terminal beliefs and four interim beliefs over the firm's discount factor: 1) $\delta < \delta_d$, 2) $\delta \in [\delta_d, \delta_x)$, 3) $\delta \in [\delta_x, \delta_y)$, 4) $\delta \geq \delta_y$and 5) $\delta \in (\delta_d, \delta_y]$, 6) $\delta \geq \delta_d$, 7) $\delta < \delta_y$and 8) $\delta \in (0, 1]$.

We begin by showing that it is a best response for players to play according to equilibrium in each of these states.

1)–2) If individuals believe that $\delta < \delta_x$ then it must be that the firm has cheated an x. Since the firm is expected to cheat all subsequent individuals, no individual trusts the firm. Since no individual trusts the firm, it is a best response for the firm to cheat at every opportunity. The value of such histories to the firm is 0.

3) If individuals believe that $\delta \in [\delta_x, \delta_y)$ then it must be that a y has been cheated and every x has been treated fairly, including at least one x subsequent to the cheated y. The equilibrium value to the firm of such beliefs is given by $V_3 = \gamma_x(1-\alpha) + \delta V_3$ or $V_3 = V_x$. Thus as long at $\delta \geq \delta_x$, it is a best response for the firm to treat x's fairly and to cheat every y.

4) If individuals believe that $\delta \geq \delta_y$ then in must be that every individual has been treated fairly including at least one y. The equilibrium value to the firm of such beliefs is V_s and as long as $\delta \geq \delta_y$, it is a best response for the firm to treat all individuals fairly.

5) If individuals believe that $\delta \in [\delta_d, \delta_y)$ then it must be that x's have been encountered and never been cheated, a y has been cheated and no x's have been encountered subsequent to the y being cheated. Since y individuals anticipate being cheated, they never trust again and since y individuals never trust the firm, it is a best response for the firm to cheat every y. An x individual will trust the firm if and only if she believes that the expected gain from being treated fairly, $[F(\delta_y) - F(\delta_x)](\alpha - c)$, outweighs the expected loss from being cheated, $[F(\delta_x) - F(\delta_d)](c - \beta)$. If the firm encounters an x and the x trusts the firm then the firm's payoff from treating the x fairly is $(1-\alpha) + \delta V_x$ and the payoff from cheating the x is $1 - \beta$. If $\delta \geq \delta_x$ then the firm's best response is to treat the x fairly and play proceeds to 3). If $\delta < \delta_x$ then the firm's best response is to cheat the x and play proceeds to 2).

6) If individuals believe that $\delta \geq \delta_d$ then it must be that x's have been encountered and treated fairly and no y's have been encountered. There are three separate cases to be analyzed:

1. $\delta \in [\delta_d, \delta_x)$: Suppose the firm encounters a y. According to equilibrium the firm cheats the y and its payoff is $(1-\beta) + \delta V_{da}$ where V_{da} is the continuation payoff from cheating the next x it encounters. V_{da} must satisfy $V_{da} = \gamma_x(1-\beta) + \gamma_y \delta V_{da}$ or $V_{da} = \gamma_x(1-\beta)/(1-\gamma_y\delta)$. If instead the firm treats the y fairly, individuals believe $\delta \geq \delta_y$ so that the firm can only continue to play as if it is fair to all or it can cheat the next individual and go off the equilibrium path (there's no point to waiting to go off the equilibrium path). When $\delta \geq \delta_s$ it prefers the former and gets payoff $(1-\alpha)+\delta V_s$. When $\delta < \delta_s$ it strictly prefers the latter and gets payoff $(1-\alpha)+\delta(1-\beta)$. First note that for $\gamma_x \geq \gamma_y$, V_{da} is minimized at $\gamma_x = \gamma_y = 1/2$ so that the equilibrium payoff, $(1-\beta) + \delta V!_{da}$, is therefore bounded below by $2(1-\beta)/(2-\delta)$. When $\delta \geq \delta_s$, some straightforward manipulation reveals that $2(1-\beta)/(2-\delta) \geq (1-\alpha)/(1-\delta)$. When $\delta < \delta_s$, it is similarly straightforward to show that $2(1-\beta)/(2-\delta) > (1-\alpha) + \delta(1-\beta)$. Therefore in either case, it is a best response for the firm to play according to the equilibrium and cheat the y.

   Now suppose that the firm encounters an x. According to equilibrium the firm treats the x fairly in order to cheat the next y whereupon it then will cheat the next x. The firm's payoff to being fair to the x is $(1-\alpha) + \delta V_{db}$ where $V_{db} = \gamma_x((1-\alpha)+\delta V_{db}) + \gamma_y((1-\beta)+\delta V_{da})$ or $V_{db} = (1-\beta)(1-\delta_s\gamma_x + \frac{\delta\gamma_x\gamma_y}{1-\delta\gamma_y})/(1-\delta\gamma_x)$. On the other hand, if the firm cheats the x, play moves off the equilibrium path and the firm gets $1-\beta$. Therefore the firm is fair whenever $(1-\alpha)+\delta V_{db} \geq 1-\beta$. Solving at equality yields:

$$\delta_d = \frac{1 + \delta_s\gamma_y - \sqrt{(1+\delta_s\gamma_y)^2 - 4\delta_s{\gamma_y}^2}}{2{\gamma_y}^2} \tag{6}$$

   Since V_{db} is increasing in δ, it follows that if $\delta \geq \delta_d$ then $(1-\alpha) + \delta V_{db} \geq 1-\beta$.

2. $\delta \in [\delta_x, \delta_y)$: According to equilibrium the firm treats the x's fairly and cheats

the first y it encounters. Its equilibrium payoff is $V' = \gamma_x[(1-\alpha)+\delta V'] + \gamma_y[(1-\beta) + \delta V_x] > V_x$. Since $\delta \geq \delta_x$ the firm strictly prefers to treat x's fairly.

3. $\delta \geq \delta_y$: According to equilibrium the firm treats all individuals fairly and its equilibrium payoff is $V_s = (1-\alpha)/(1-\delta)$. Suppose that the firm encounters an x. If it cheats the x, it gets $1-\beta$. Since $\delta > \delta_s$, the firm strictly prefers to treat the x fairly.

   Suppose that the firm encounters a y. If it cheats the y, then individuals will believe that $\delta \in [\delta_d, \delta_y)$ and provided that case 5)'s conditions on F are satisfied, the firm's best response is to treat future x's fairly. Therefore the firm's maximal continuation payoff from cheating the y is $1-\beta+\delta V_x$. Since $\delta > \delta_s$, $V_s > 1-\beta+\delta V_x$ and the firm strictly prefers to treat the y fairly.

7) If individuals believe that $\delta < \delta_y$ then it must be that no x's have been encountered and the firm cheated the first y. Type y individuals never trust because they expect to be cheated and since y individuals never trust, it is a best response for the firm to cheat y's.

For the x's if the firm cheats the x, it gets $1-\beta$. If it treats the x fairly, it gets V_x. Thus when $\delta < \delta_x$, the firm strictly prefers to cheat the x and when $\delta \geq \delta_x$, the firm prefers to treat the x fairly.

8) If $\delta \in [0, 1)$ then the firm has not yet encountered any individuals. A y individual will only trust the firm if the expected gain from being treated fairly, $[1-F(\delta_y)](\alpha - c)$, outweighs the expected loss from being cheated, $F(\delta_y)(c-\beta)$. If y's trust the firm then clearly x's will also be willing to trust the firm.

Suppose that the firm encounters a y. If the firm treats the y fairly, the game proceeds to case 4) with an equilibrium payoff of V_s. If the firm cheats the y, the game proceeds to case 7). If $\delta \geq \delta_x$ then $V_x \geq 1-\beta$ and the firm treats x's fairly. If $\delta < \delta_x$ then $V_x < 1-\beta$ and the firm cheats the x. Thus the firm's continuation payoff from cheating the y is $1-\beta+\delta V_{da}$ if $\delta < \delta_x$ and $1-\beta+\delta V_x$ if $\delta \geq \delta_x$. We know that $1-\beta+\delta V_x > V_s$ for $\delta < \delta_y$. Moreover, when $\delta < \delta_x$, $V_{da} > V_x$ so that $1-\beta+\delta V_{da} > 1-\beta+\delta V_x > V_s$ and the firm strictly prefers to cheat the y. ■

References

Akerlof, G. and Kranton, R., Economics and Identity, *Quarterly Journal of Economics*, 115:715–753, 2000.

Annen, K., Stereotypical Knowledge and Trust Among Strangers, 2001, working paper.

—, Social Capital, Inclusive Networks, and Economic Performance, *Journal of Economic Behavior and Organization*, 50:449–463, 2003.

Antecol, H. and Kuhn, P., Gender as an Impediment to Labor Market Success: Why do Young Women Report Greater Harm?, *Journal of Labor Economics*, 18:702–728, 2000.

Arrow, K. J., The Theory of Discrimination, in A. Ashenfelter, Orley and Rees, ed., *Discrimination in Labor Markets*, pp. 3–33, Princeton University Press, Princeton, NJ, 1973.

Aumann, R. J., Acceptal Points in General Cooperative \$n\$-person Game, *Annals of Mathematical Studies*, 40:287–324, 1959.

Ayres, I., *Pervasive Prejudice: Unconventional Evidence of Gender Discrimination*, University of Chicago Press, Chicago, IL, 2001.

Ayres, I. and Siegelman, P., Race and Gender Discrimination in Negotiation for the Purchase of a New Car, *American Economic Review*, 85:304–321, 1995.

Bar-Isaac, H., Imperfect Competition and Reputational Commitment, *Economics Letters*, 89(2):167 – 173, 2005.

Basu, K., Racial Conflict and the Malignancy of Identity, *Journal of Economic Inequality*, 3:221–241, 2005.

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

Bernheim, B. D., Peleg, B. and Whinston, M., Coalition Proof Nash Equilibria: I Concepts, *Journal of Economic Theory*, 42(1):1–12, 1987.

Bernheim, B. D. and Ray, D., Collective Dynamic Consistency in Repeated Games, *Games and Economic Behavior*, 1:295–326, 1989.

Bernheim, B. D. and Whinston, M. D., Multimarket Contact and Collusive Behavior, *The RAND Journal of Economics*, 21(1):1–26, 1990.

Bertrand, M., Chugh, D. and Mullainathan, S., Implicit Discrimination, *American Economic Review, Papers and Proceedings*, 95:94–98, 2006.

Charles, K. K. and Guryan, J., Prejudice and the Economics of Discrimination, *Journal of Political Economy*, 116:773–809, 2008.

Coate, S. and Loury, G. C., Will Affirmative-action Policies Eliminate Negative Stereotypes?, *American Economic Review*, 83:1220–1240, 1993.

Cole, H. L. and Kehoe, P. J., Models of Sovereign Debt: Partial Versus General Reputations, *International Economic Review*, 39:55–70, 1998.

Coleman, J. S., Social Capital in the Creation of Human Capital, *American Journal of Sociology*, 94:S95–S120, 1988.

Dasgupta, P., Trust as a Commodity, in D. Gambetta, ed., *Trust: Making and Breaking Cooperative Relations*, Basil Blackwell, Oxford, 1990.

Dixit, A. K., Trade Expansion and Contract Enforcement, *Journal of Political Economy*, 111(6):1293–1317, 2003.

Donahue, J. J., The Law and Economics of Antidiscrimination Law, in A. M. Polinsky and S. Shavell, eds., *Handbook of Law and Economics*, North Holland, 2007.

Douglass, F., Negro Exodus from the Gulf States, 1879, address before American Social Science Association meetings in Saratoga, New York.

Edgeworth, F., Equal Pay to Men and Women for Equal Work, *Economic Journal*, 32:431–457, 1922.

Farrell, J. and Maskin, E., Renegotiation in Repeated Games, *Games and Economic Behavior*, 1:327–360, 1989.

Farrell, J. and Weizsacker, G., Renegotiation in the Repeated Amnesty Dilemma, with Economic Applications, *International Series in Operations Research and Management Science*, 35:213–246, 2001.

Fawcett, M. G., Mr. Sidney Webb's Article on Women's Wages, *Economic Journal*, 2:173–176, 1892.

—, Equal Pay for Equal Work, *Economic Journal*, 28:1–6, 1918.

Fershtman, C. and Gneezy, U., Discrimination in a Segmented Society, *Quarterly Journal of Economics*, 116:351–377, 2001.

Fudenberg, D. and Levine, D. K., Reputation and Equilibrium Selection in Games with a Patient Player, *Econometrica*, 57(4):759–778, 1989.

Green, E. J. and Porter, R. H., Noncooperative Collusion Under Imperfect Price Information, *Econometrica*, 52:87–100, 1984.

Greif, A., Contract Enforceability and Economic institutions in Early Trade: the Maghribi Trader's Coalition, *American Economic Review*, 83:523–548, 1993.

Greif, A., Milgrom, P. and Weingast, B. R., Coordination, Commitment, and Enforcement: the Case of the Merchant Guild, *American Economic Review*, 84:745–776, 1994.

Horner, J., Reputation and Competition, *American Economic Review*, 92(3):644–663, 2002.

Klein, B. and Leffler, K. B., The Role of Market Forces in Assuring Contractual Performance, *Journal of Political Economy*, 89:615–641, 1981.

Kreps, D. M., Corporate Culture and Economic Theory, in J. Alt and K. Shepsle, eds., *Perspectives on Positive Political Economy*, Harvard University Press, Cambridge, MA, 1990.

Kreps, D. M., Milgrom, P., Roberts, J. and Wilson, R., Rational Cooperation in the Finitely Repeated Prisoner's Dilemma, *Journal of Economic Theory*, 27:245–252, 1982.

Lazzarini, S., Miller, G. and Zenger, T., Order with Some Law: Complementarity vs. Substitution of Formal and Informal Arrangements, *Journal of Law, Economics and Organization*, 20(2):261–298, 2004.

Loury, G. C., A Dynamic Theory of Racial Income Differences, in P. A. Wallace and A. M. LaMond, eds., *Women, Minorities, and Employment Discrimination*, Lexington Books, 1977.

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

—, Information and Racial Exclusion, *Journal of Population Economics*, 20:641–642, 2007.

Mailath, G. J. and Samuelson, L., *Repeated Games and Reputations: Long-Run Relationships*, Oxford University Press, 2006.

Mailath, G. J., Samuelson, L. and Shaked, A., Endogenous Inequality in Integrated Labor Markets with Two-sided Search, *American Economic Review*, 90:46–72, 2000.

Milgrom, P. and Roberts, J., Coalition-proofness and Correlation with Arbitrary Communication Possibilities, *Games and Economic Behavior*, 17:113–128, 1996.

Morris, S., Political Correctness, *Journal of Political Economy*, 109:231–265, 2001.

Ostrom, E., Collective Action and the Evolution of Social Norms, *Journal of Economic Perspectives*, 14:37–158, 2000.

Phelps, E. S., The Statistical Theory of Racism and Sexism, *American Economic Review*, 62:650–651, 1972.

Poppo, L. and Zenger, T., Do Formal Contracts and Relational Governance Function as Substitutes or Complements?, *Strategic Management Journal*, 23:707–725, 2002.

Radner, R., Repeated Principal-agent Games with Discounting, *Econometrica*, 53:1173–1198, 1985.

Resnick, P., Zeckhauser, R., Friedman, E. and Kuwabara, K., Reputation Systems, *Communications of the ACM*, 43:45–48, 2000.

Rubinstein, A., Strong Perfect Equilibria in Supergames, *International Game Theory Review*, 9:1–12, 1978.

Smith, A., *Lectures on Jurisprudence*, 1766.

Tirole, J., A Theory of Collective Reputations, *Review of Economic Studies*, 63:1–22, 1996.

van Damme, E., Renegotiation-proof Equilibria in Repeated Prisoners' Dilemma, *Journal of Economic Theory*, 47(1):206–217, 1989.

Williamson, O., *The Economic Institutions of Capitalism*, Free Press, 1985.