Opportunistic Discrimination*

Rick Harbaugh† Ted To‡

July 2013

Abstract

Are minorities more vulnerable to opportunism? We find that individuals from a minority group face greater danger of being cheated by an opportunistic firm because trade with the minority is less frequent and the value of a reputation for fairness toward the minority is correspondingly smaller. If the majority is sufficiently large it has no reason to fear opportunism by the firm, so the firm can continue business as usual with the majority even after cheating the minority. And if there is a small chance that the firm might have an implicit or preference bias against either group, then the interaction with reputational incentives gives unbiased firms an incentive to cheat the minority but not the majority. The prediction that smaller groups are more susceptible to discrimination distinguishes the model from most other discrimination models.

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

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

*For helpful comments we thank Mike Baye, George Evans, Dragan Filipovich, Sid Gordon, Nandini Gupta, 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 property is stolen during ethnic unrest, is the perpetrator viewed as opportunistic toward everyone or as someone who can still be trusted in his own community? If a woman is unfairly denied a promotion, do her male colleagues expect a similar fate 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 policyholders, do others 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 weigh the immediate gains from cheating against 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, might people from another group just ignore it and proceed with business as usual? If so, then the incentive to cheat a person can depend on that person's group identity, so that it might be profitable to cheat members of one group but not another.

Despite the central role that opportunism plays in many areas of economics, the idea that discrimination can involve some people being "cheated" is surprisingly absent from economic theories of discrimination, including models of 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), search costs (Black, 1995; Mailath *et al.*, 2000), identity (Akerlof and Kranton, 2000), coordination (Eeckhout, 2006), association (Peski and Szentes, 2013), and implicit bias (Greenwald *et al.*, 1998; Bertrand *et al.*, 2006). These models do not capture the idea that people from some groups are more likely

to be taken advantage of than others. Such discrimination might be due to differential ability to enforce contractual remedies against opportunism because of unequal access to the legal system (Douglass, 1879), but equal access alone does not ensure fairness due to the inherent difficulty of enforcing contracts in many environments (Williamson, 1985). We examine how opportunism can be discriminatory in an environment where the primary constraint on opportunism is reputational rather than contractual or legal.

Most discrimination models predict that there is no effect of group size on the susceptibility to discrimination, or that discrimination is actually worse for larger groups.[1] In particular, preference-based models of discrimination find that the effects of bias become worse as the group facing bias becomes larger and its members find it increasingly difficult to find positions among unbiased firms. In contrast, the opportunistic discrimination model we analyze predicts 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 U.S. 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).

To understand opportunistic discrimination against a minority group, we model a simple 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 a non-contractible investment or other resource commitment, and the firm then either cheats the individual by taking all the gains of the investment, or lets the individual benefit as well. 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,

[1] Search-based models predict that either the minority or majority can be discriminated against, with the exception of Lundberg and Startz (2007) in which there is less return to learning about the ability of small groups, and of Black (1995) in which biased firms survive at a lower rate as the minority population increases. In a repeated prisoner's dilemma where cooperation is encouraged by lack of attractive outside options, Eeckhout (2006) finds that there can exist segregation equilibria where the minority is worse off.

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 grim trigger strategy where everyone initially trusts the firm but if the firm cheats anyone then no one trusts the firm again.

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 grim trigger strategy, which we refer to as the "solidarity trigger strategy". Given that individuals follow 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 its reputation among one group 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.

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 or other biases, we find that a discrimination equilibrium with discrimination against the minority is supported by a wider range of discount factors 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

[2]Smith (1766) argues that trust increases with the frequency of commercial exchange and that opportunism is therefore most problematic in undeveloped regions like his native Scotland. His emphasis on the frequency of interactions as determining the possibility for trust also appears in his claim that opportunism is more likely in political and diplomatic activities where transactions are less frequent than they are in commerce.

from standing with the minority, so the minority is vulnerable to opportunism.[3]

A problem with repeated trust games is that it is often not credible to punish cheating since the players have an incentive to forgive the cheater.[4] To address this renegotiation problem we assume that with some probability in any period the firm becomes inept and cheats the individual because it cannot generate sufficient surplus to reward him. Because of this small chance, individuals believe a firm that has cheated is no longer capable of being trustworthy, so individuals optimally respond to cheating by refusing to deal with the firm. Even if all or some individuals act cooperatively, they have no incentive to forgive a cheating firm and start trusting again so the equilibria we analyze are coalition-proof equilibria that are immune to renegotiation. Hence by following the "separating" approach to reputation in which a firm maintains its reputation by behaving differently than a bad type of firm (Mailath and Samuelson, 2006),[5] we also resolve the renegotiation problem.

To examine the interaction of reputation effects with implicit bias (Greenwald *et al.*, 1998; Bertrand *et al.*, 2006) and preference-based discrimination (Becker, 1957), we allow for a small possibility that the firm is a biased firm that always cheats one group.[6] From an implicit bias perspective, the firm might not appreciate the capabilities of members of that group, e.g., be unable to perceive that a member of one group deserves a promotion or deserves financing for a project. From a preference perspective, the firm might literally prefer to cheat members of one group. We are interested in the case where bias is unlikely so its direct effects are limited and we allow for bias against either the majority or the minority. There need not be any actual bias — the players just have to believe it is a

[3]Note that in our model there is no competition for resources between the majority and minority as in models of ethnic conflict (Esteban and Ray, 2008), so the majority does not gain directly from firm discrimination against the minority.

[4]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.

[5]The alternative "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.

[6]The patterns of employment discrimination against African-Americans are consistent with firm bias (Charles and Guryan, 2008).

possibility.

We find that if the potential bias is against the majority there is no interaction with reputation. But if the potential bias is against the minority then such bias interacts with reputation to make discrimination against the minority the unique type-stationary coalition-proof equilibrium when a firm is 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 with the firm and continue business as usual. A non-biased firm of intermediate patience then has an incentive to pool with biased firms and thereby reap both the short-term benefits from cheating the minority and the long-term benefits of a good reputation with the majority. For instance, an employer might literally add insult to injury after cheating an employee in order to suggest to other employees that his opportunism is limited to a particular group.[7]

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).[8] 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, and the Annen (2003) model of inclusive networks in which communication becomes harder as networks expand. The tradeoff between network size and communication is formalized in the Dixit (2003) model of trade networks in which distant trade is both more valuable and harder to monitor. We differ from the social capital literature in considering how inefficient discrimination equilibria can arise even with public information about who cheats. Since individuals are aware of opportunism against members of other groups, they must decide when it is in their interest to punish opportunism against some people but not others.

[7] Reference to ethnic or gender stereotypes and use of epithets can make the "otherness" of the victim clear, but communication can also be more subtle. After refusing to pay a large promised reward to Serbian inventor Nikola Tesla for work on generators, Thomas Edison famously remarked, "You don't understand our American sense of humor."

[8] As analyzed in the social capital literature (Loury, 1977; Bowles *et al.*, 2010; Munshi, 2011), 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.

While we follow Smith (1766) in emphasizing the frequency of interactions in determining the incentives for opportunism, the information issues emphasized in the social capital literature are clearly important. When we relax the complete information assumption to allow one group to be better informed about the firm's history of opportunism against anyone, we find that a sufficiently less informed group can be more susceptible to discrimination even if it is the majority. If we assume that individuals are well informed about opportunism against their own group, but ignorant of any opportunism against the other group, then we find that the smaller group is always more susceptible to opportunism and there is no possibility of attaining the solidarity equilibrium.

In many countries differential access to the legal system allows some groups but not others to gain contractual protection against opportunism. To capture this possibility we extend the model to allow for both contractual and reputational constraints on opportunism. We find that increased contractual protection for the majority reduces the majority's reliance on reputational sanctions, thereby weakening reputational constraints on opportunism against either group and leaving the minority worse off than if both groups were forced to rely on reputation alone. Generally, 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 the firm's behavior toward the other group. These results that enforcement against opportunism can either aggravate or ameliorate opportunistic discrimination depending on the nature of enforcement adds to the large literature which examines whether legal and contractual constraints are complements to or substitutes for reputational constraints (e.g., Ostrom, 2000; Poppo and Zenger, 2002; Lazzarini *et al.*, 2004).

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, we consider

reputations for trustworthiness *toward* different groups.[9] To emphasize this focus of the paper, the firm in our model is not required to have a group identity.[10]

In the following section we provide our basic model of a trust game and of the interaction between reputation and preference-based discrimination. In Section 3 we extend the model to include the interaction between reputation and explicit enforcement against opportunism, and to allow for imperfect knowledge of opportunism. Section 4 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.[11] 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 made by an employee, a loan provided by a creditor, a transaction-specific investment made by a supplier, etc. If the individual trusts the firm (trades with it), then either the firm is fair and the individual receives a net payoff of $\alpha - c > 0$ or the firm cheats and the individual receives $-c$. The total gross value of the trade is normalized to 1 so the firm receives $1 - \alpha$

[9] 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.

[10] 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).

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

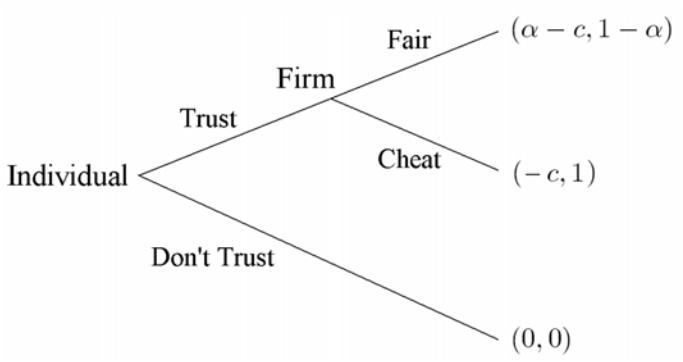

Figure 1: Trust game in each period.

from being fair and 1 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 $n > 2$ 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 $\gamma_p = n_p/n$.[12] The firm discounts the period between transactions by a common knowledge factor $\delta \in (0, 1)$. If the firm has a higher discount factor we say it 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 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 by providing a payoff of $-c$ to the individual and $0 < \beta < c$ to itself.[13] For instance, the

[12] 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. Note that the model is equivalent to each individual matching randomly with the firm according to an independent Poisson distribution.

[13] To keep the probability of ineptness from accumulating in periods with no trade, we assume that a transition can only happen after a trade. One could also allow for a small chance in each period that the firm transitions back to being normal. Such transitions back and forth between types are used in the literature on impermanent reputations (e.g., Ekmekci *et al.*, 2012).

firm cannot meet its obligations due to financing problems, changes in ownership, or the loss of key employees. Individuals only observe the history of payoffs to individuals so do not know whether observed cheating is by a normal firm or by an inept firm.[14] We examine the limiting case where ε goes to zero. As discussed in the introduction, this small chance that a firm might be inept allows us to follow the "separating" approach to reputation and also solves the credibility problem in repeated trust games that players have an incentive to renegotiate away from strategies that punish opportunism. Since individuals are cheated along the equilibrium path by inept firms, the simple response of no longer dealing with firms who cheat is an equilibrium strategy even when we allow for renegotiation by the players.[15] As discussed later, to capture such renegotiation we will consider a coalition-proofness refinement of perfect Bayesian equilibria.

To incorporate the possibility that the firm has an implicit bias or a preference-based bias, we allow for a small independent probability $\phi_p \geq 0$ that the firm always cheats a member of group p. The firm may be biased against either group or even both groups. We look at the case where the probability of bias is low enough that, absent any history of opportunism against a group, members of each group will still trust the firm if firm types who are neither inept nor biased against them do not cheat, i.e., $(1-\varepsilon)(1-\phi_p)(\alpha-c) \geq \left[1-(1-\varepsilon)(1-\phi_p)\right]c$, or in the limit as ε goes to zero, $\phi_p \leq (\alpha-c)/\alpha$ for $p = x, y$.[16] Note that a biased but otherwise normal firm always behaves rationally with respect to individuals against whom they have no bias. In particular, if type p individuals trust and type q individuals never trust then observationally a firm that is biased against q but not against p behaves as if it were a normal and unbiased firm.

In the trust literature it is typical to concentrate on the no-trust strategy in which

[14] In some settings normal firms might be able to partially or fully distinguish themselves from inept firms by financial statements or other means. If normal firms are not expected to cheat in equilibrium, then individuals should still interpret unexpected cheating as by inept firms unless such information is fully distinguishing.

[15] Similar to our interpretation, Mailath and Samuelson (2001) model "inept" types who cannot meet their obligations, while Ghosh and Ray (1996) model "myopic" types who always choose not to meet their obligations. We differ from these approaches in the assumption that the firm can go bad during the course of the game.

[16] Since each of the finite number of individuals interacts with the firm again in the future there is also some information value to testing whether the firm is biased, so the stated sufficient condition is only tight as n becomes large.

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 grim 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.*

Our main equilibrium concept is a *pure strategy perfect Bayesian equilibrium* (or just "equilibrium"), 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 focus on equilibria that are *type-stationary* in that equilibrium strategies for normal firms, 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. Expecting such cheating, individuals never trust the firm. If an individual deviates and trusts the firm, the firm's stage payoff from cheating is higher than from not cheating, and the continuation payoff remains at zero, so the firm's best response is to cheat. Therefore the individual's best response is to not trust, and the no-trust equilibrium always exists. Note that this same logic that an individual deviation to trusting does not benefit the individual applies whenever the firm's strategy is to cheat the individual on or off the equilibrium path.

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 otherwise (off the equilibrium path) to cheat every individual. Since

the firm maintains its reputation if it is fair to every individual and also does not become inept, the value V_s of a reputation for being fair when individuals follow the solidarity trigger strategy is $V_s = 1-\alpha+\delta\left[(1-\varepsilon)V_s+\varepsilon\beta\right]$, or $V_s = \left(1-\alpha+\delta\beta\varepsilon\right)/\left[1-\delta\left(1-\varepsilon\right)\right]$. If the firm cheats the individual then the firm is expected to continue cheating, so by the above argument no individual will deviate to trusting the firm in the future and the firm earns a payoff of 1. Therefore the discount factor δ_s such that a (normal) firm is just indifferent between being fair to and cheating an individual is given by $1-\alpha+\delta\left[\left(1-\varepsilon\right)V_s+\varepsilon\beta\right]=1$ or, substituting,

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

Since the payoff from being fair is increasing in δ, the solidarity-trust equilibrium exists if and only if $\delta \geq \delta_s$.

Now suppose type p individuals follow the discriminatory trigger strategy and type q individuals, expecting to be cheated, follow the no-trust strategy. For the corresponding *q-discrimination equilibrium* (which we will refer to as the minority-discrimination equilibrium for $q = y$ and the majority-discrimination equilibrium for $q = x$) 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. If the firm faces a q individual then, as in the no-trust equilibrium, there is no incentive to deviate from cheating the individual so the individual will not trust. For members of group p, let V_p be the value of a reputation for treating them 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, $V_p = \gamma_p\left(1-\alpha\right)+\delta\left[(1-\gamma_p\varepsilon)V_p+\gamma_p\varepsilon\beta\right]$ or, substituting, $V_p = \gamma_p\left[(1-\alpha)+\delta\varepsilon\beta\right]/\left(1-\delta+\delta\varepsilon\gamma_p\right)$. Given this reputation value, the discount factor δ_p such that the firm is indifferent between between being fair to a p individual and earning a payoff of 1 from cheating, is given by $1-\alpha+\delta\left[(1-\gamma_p\varepsilon)V_p+\gamma_p\varepsilon\beta\right]=1$ or, substituting,

$$\delta_p = \frac{\alpha}{\alpha+\gamma_p\left[\left(1-\alpha\right)\left(1-\gamma_p\varepsilon\right)+\varepsilon\left(\beta-\alpha\right)\right]}. \tag{2}$$

Since the payoff from being fair is increasing in δ, this discrimination equilibrium exists if and only if $\delta \geq \delta_p$.

Finally consider the case where both types of individuals follow the discriminatory trigger strategy. For the corresponding *independent-trust equilibrium* we define a normal

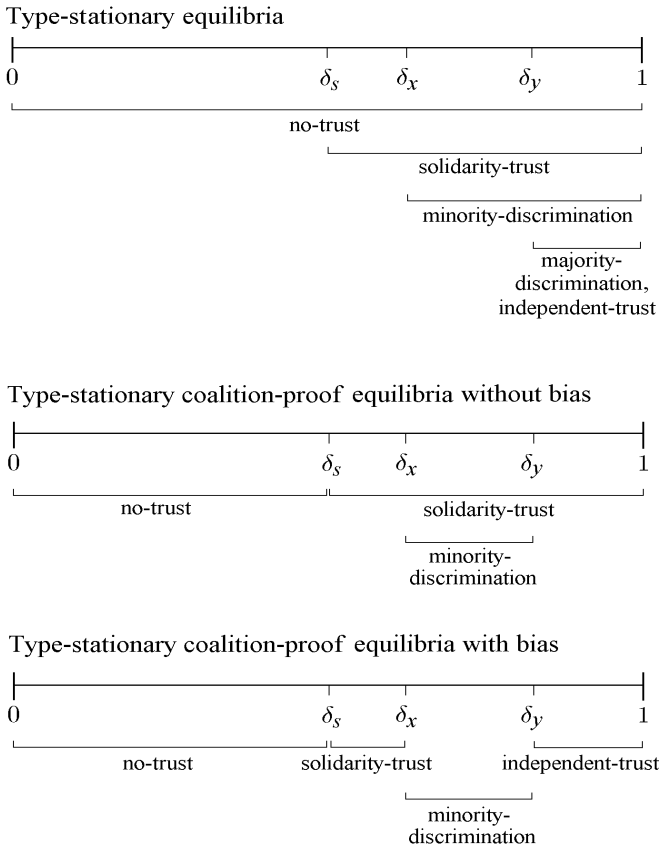

Figure 2: Equilibrium ranges for $c = 1/3$, $\alpha = 1/2$, and $\gamma_x = 3/4$.

firm's strategy as to be fair to every individual if no individual has ever been cheated and otherwise (off the equilibrium path) to cheat every individual. Following the above logic, if $\delta \geq \delta_p$ for $p = x, y$ then each group is sufficiently large to deter opportunism even though it receives no help from the other group. Note that for a normal and unbiased firm, the independent-trust equilibrium has the same equilibrium outcomes as the solidarity-trust equilibrium.

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.[17] Hence, based on the above analysis, we have the following result. Note that it

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

holds for all $0 \leq \phi_x, \phi_y < (\alpha - c)/\alpha$ so it does not depend on the existence of biased firms.

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

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 Figure 2 that discrimination against the minority is supported by a wider range of discount factors must hold.

Corollary 1 *The range of discount factors supporting a minority-discrimination equilibrium is larger than the range supporting a majority-discrimination equilibrium.*

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 minority-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

model there is no gain to limiting the punishment period.

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 equilibria that are often argued to be unreasonable (e.g., Bernheim *et al.*, 1987; Farrell and Maskin, 1989; Milgrom and Roberts, 1996). For instance, for $\delta \geq \delta_s$ the solidarity-trust 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 argued that the players should be able to talk their way to it. Similarly, for $\delta \geq \delta_y$, 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, these equilibria are similarly unreasonable.

Such arguments do not eliminate all inefficient equilibria. Consider the minority-discrimination equilibrium in the range $\delta \in [\delta_x, \delta_y)$. The solidarity-trust equilibrium also exists in this range and both the minority and the firm are strictly better off in it. However, the difference from the previous cases is that the minority and the firm alone cannot induce a change to this better equilibrium, but are dependent on the majority changing its strategy to the solidarity strategy. Since the majority is already doing as well as they can in the current equilibrium, it is unclear why they would switch.[18] 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.

The 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 (Bernheim *et al.*, 1987; Farrell

[18] If there is any chance that the firm is biased against the minority, they strictly lose from the lost trade opportunity. They also strictly lose from a solidarity strategy 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. We do not model these latter two possibilities.

and Maskin, 1989; Milgrom and Roberts, 1996). To capture the effect of joint deviations in the simplest way, we say a perfect Bayesian 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 that are consistent along the equilibrium path. Note that any deviation must be to an equilibrium, but this equilibrium need not itself be coalition proof.[19]

As we show in the following proposition, coalition-proofness limits the multiplicity of (perfect Bayesian) equilibria in Proposition 1 in accordance with the above discussion. Note that coalition-proofness allows a coalition to be formed at any period, so it incorporates the possibility of renegotiating a planned punishment strategy following unexpected opportunism by the firm.[20] 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.

The following result is for the case where the firm is definitely not biased, and the subsequent result is for the case where there is some small chance the firm is biased.

Proposition 2 *For the case of no firm bias, $\phi_x, \phi_y = 0$, 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 minority-discrimination or solidarity-trust equilibrium if $\delta \in [\delta_x, \delta_y)$; and iv) the solidarity-trust equilibrium if $\delta \in [\delta_y, 1)$.*

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 adopt a punishment strategy that stops the firm from cheating anyone, but any one group alone cannot stop such opportunism. For the higher range $\delta \in [\delta_y, 1)$ the firm is sufficiently patient that if either

[19] In a normal form game, 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.

[20] The coalition-proofness literature generalizes the two-player renegotiation models of Bernheim and Ray (1989) and Farrell and Maskin (1989).

group switches from the no-trust strategy to the discriminatory trigger strategy then 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 coalition-proof, and it must be directed against the minority.

Proposition 2 shows the difficulty of getting the majority to punish opportunism against the minority when renegotiation is possible. This problem is exacerbated if the firm is with some small probability biased. 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. Such renegotiation 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 give the firm another chance. Under our assumption that ε goes to zero, this is always true for fixed $\phi_x > 0$.[21] 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 in this range. Therefore, even though we allow for bias against either group, the interaction effect of bias with coalition-proofness always works against the minority.

Proposition 3 *For the case of potential firm bias, $\phi_x, \phi_y > 0$, 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 minority-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 the small possibility of bias can have a large impact in choosing between multiple equilibria. If the majority starts with a solidarity strategy

[21] More generally, since $\phi_p/(\phi_p + \varepsilon)$ is the probability that the firm was biased against p, and there is still a chance that the firm is biased against q or has become inept in the current period, the majority will still trust the firm if $\left[\phi_p/(\phi_p + \varepsilon)\right](1-\varepsilon)(1-\phi_q)(\alpha - c) > \left[1 - (\phi_p/(\phi_p + \varepsilon)\right](1-\varepsilon)(1-\phi_q)c$. If, contrary to our assumption, the biases also go to zero then this condition depends on the rate that they go to zero relative to ε. If, for instance, $\phi_p = \phi_q = r\varepsilon$ for some constant $r > 0$, then in the limit as ε goes to zero the condition is satisfied for $r > c/(\alpha - c)$.

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.[22] This is shown in the bottom section of Figure 2 for the case of $\phi_p = \phi_q = 1/100$.

Based on the results of Propositions 2 and 3 we have the following corollary.

Corollary 2 *A coalition-proof minority-discrimination equilibrium always exists for some range of discount factors and is sometimes the unique type-stationary equilibrium, but a coalition-proof majority-discrimination equilibrium never exists for any discount factor.*

We have restricted attention to type-stationary equilibria, i.e., to equilibria that are either non-discriminatory or are discriminatory based only on group identity. If we assume that individuals are only distinguishable by their group identity, so that firm strategies can depend on the sequence of play and group identity but not individual identity, then it is straightforward to show that any equilibrium with strategies based on the sequence of play can be improved on so it is not coalition-proof. Hence in this case the results of Propositions 2 and 3 apply to all coalition-proof equilibria.

3 Extensions

The simplicity of the above model makes it easily extendable in a number of directions. In the following extensions we assume that firms are definitely not biased, $\phi_x, \phi_y = 0$, so as to focus on the new issues raised by the extensions.[23]

3.1 Contractual and legal constraints on opportunism

To see how direct penalties against opportunism can interact with reputation effects, we consider three enforcement strategies against opportunism. First, the government can

[22] 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.

[23] Other important extensions are beyond the scope of this paper. Notably, we do not consider the self-selection of individuals into firms as in Becker (1957), nor the effects of competition on reputation as in Hörner (2002) and Bar-Isaac (2005).

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. 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 for intermediate discount factors 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 focus on ranges where for at least one group $\pi_p < \alpha$ so the penalty does not simply eliminate all opportunism against everyone but rather allows for some interaction between the penalty and reputation effects. With sufficiently high penalties against everyone all opportunism can in theory be prevented, but in practice such strong enforcement is unlikely to be possible as Williamson (1985) emphasizes.

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-trust equilibrium is δ_s^π such that $1-\alpha+\delta_s^\pi\left((1-\varepsilon)V_s+\varepsilon\beta\right)=1-\min\{\pi_x,\pi_y\}$. Substituting and taking the limit as ε goes to zero,[24]

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

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 $1-\alpha+\delta\left((1-\gamma_p\varepsilon)V_p+\gamma_p\varepsilon\beta\right)=1-\pi_p$, or, substituting and taking the limit as ε goes to zero,

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

We can now use these cutoffs just as in the previous analysis. In particular, the ranges for coalition-proof equilibria are the same as given in Proposition 2, except with these penalty-adjusted cutoffs.

[24] For notational convenience we define the δ cutoffs at the limit of $\varepsilon = 0$, so the cutoffs are now the lower limit for each region but not in the region.

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-trust equilibrium region $(\delta_s^\pi, \delta_x^\pi]$ shrinks and the minority-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 $\pi_x^* = \alpha\gamma_y$, 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 further increases in π_x make the majority better off and the minority is not hurt further.

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 minority-discrimination equilibrium if $\delta \in (\delta_y^\pi, \delta_y]$ and it cannot cause a switch into the minority-discrimination region since it does not affect δ_x^π. However, it too can be counterproductive. Letting ε go to zero, if π_y is higher than

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

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-trust equilibrium would only survive in the region $\delta \in (\delta_s, \delta_x]$, this is a net loss for $\delta \in (\max\{\delta_s, \delta_y^\pi\}, \delta_x]$.

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-trust equilibrium survives. And if $\delta \in (\delta_y^\pi, \delta_y]$ then it induces a switch from the minority-discrimination region to a region where only the solidarity-trust (or independent-trust) 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-trust equilibrium survives to the region where the minority-discrimination equilibrium survives. Without any enforcement the majority would have to follow the

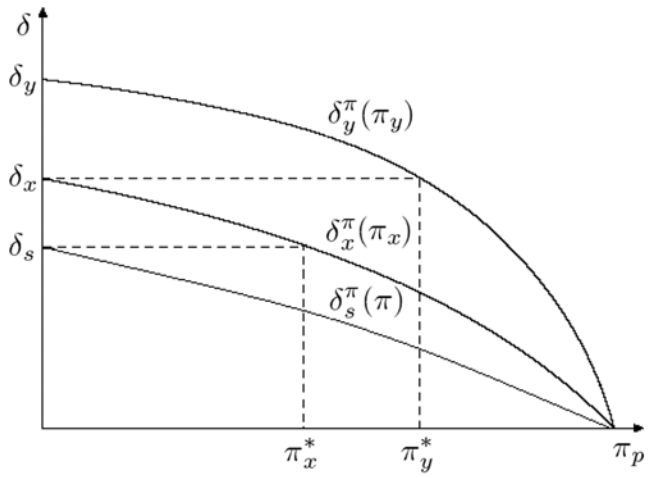

Figure 3: Impact of enforcement strategies on equilibrium cuttoffs.

solidarity strategy to avoid the no-trust equilibrium, but the combination of reputation and explicit sanctions makes it possible for the majority 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,[25] except that reverse discrimination cannot result even if the penalty is higher than π_y^* because discrimination against the majority will also be penalized. Therefore the original minority-discrimination region disappears, a majority-discrimination region is not created, and only the solidarity-trust and independent-trust equilibria survive for $\delta \geq \delta_s$. Comparing these different enforcement strategies, we have the following result.

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

[25] 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.

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 sometimes make reputation less effective and increase overall opportunism. 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.

3.2 Differential knowledge of firm history

The model makes the simplifying assumption that the firm's history is common knowledge so any opportunism is known by everyone. In practice, knowledge of opportunism is likely to be imperfect so that opportunism does not necessarily lead to a loss of business with everyone. Moreover, groups often differ in their ability to learn about acts of opportunism. For instance, tourists are more vulnerable to opportunism in part because they are less likely to share information than local customers. We want to check the robustness of the results to the more realistic assumption of imperfect knowledge by everyone and also to investigate how differential knowledge affects discrimination.

To do this we now allow for the possibility that an individual is not aware of a firm's past opportunism. We allow individuals from the two groups to differ in the likeliness that they are aware, and also allow this difference to depend on which group was cheated. In particular we assume that if the firm has cheated a member of group q in the past then with probability θ_p^q a member of group p is aware that the firm has cheated a member of group q and with probability $1 - \theta_p^q$ she is ignorant.[26] Because of the complexity of forming coalitions between informed and uninformed members, we will focus only on perfect Bayesian equilibria and therefore for simplicity assume $\varepsilon = 0$.

Any chance that opportunism will not be observed makes sustaining a punishment strategy more difficult, so the discount factor cutoffs for the different types of equilibria

[26] That private information is not problematic in our context follows from the simplicity of one-sided incentive problems (Kandori, 1992). Note that this assumption implies that ignorant individuals forget their own history with the firm. Allowing them to remember their own history but not know the history of other individuals does not change the results qualitatively as long as there are multiple members of each group, and the quantitative results converge to the present case as the number of individuals becomes large.

will rise. To see the impact on discrimination equilibria, note that the discounted value of a reputation for fairness to p individuals when q individuals do not trust is still V_p, and that a firm which decides instead to cheat a p should do so immediately and then continue to cheat ignorant p individuals. Therefore the cutoff discount factor δ_p^h such that the firm is indifferent between cheating the first p and not is $1-\alpha+\delta_p^h V_p = 1+(\delta_p^h/(1-\delta_p^h))\gamma_p(1-\theta_p^p)$ or

$$\delta_p^h = \frac{\alpha}{(1-\gamma_p(1-\theta_p^p))\alpha + \gamma_p\theta_p^p(1-\alpha)}. \tag{6}$$

As expected, if $\theta_p^p < 1$ then $\delta_p^h > \delta_p$. The difference $\delta_y^h - \delta_x^h$ is proportional to

$$(\gamma_x\theta_x^x - \gamma_y\theta_y^y) - (\gamma_x - \gamma_y)\alpha. \tag{7}$$

If $\theta_y^y = \theta_x^x = \theta$ so that knowledge of firm history is imperfect but each group is equally (un)informed, then $\delta_y^h - \delta_x^h$ is positive, implying that the result from Corollary 1 of discrimination against the minority still holds. And if $\theta_y^y \neq \theta_x^x$ then the result still holds if we replace the condition that group x is larger than group y with the condition that (7) is positive, which holds for $\gamma_x\theta_x^x - \gamma_y\theta_y^y$ sufficiently large. The following proposition therefore follows from application of these new cutoffs δ_p^h to Proposition 1.

Proposition 5 *The range of δ supporting a minority-discrimination equilibrium is larger than the range supporting a majority-discrimination equilibrium iff $\gamma_y\theta_y^y$ is sufficiently smaller than $\gamma_x\theta_x^x$.*

This confirms that the basic insight about the quality of communication from the social capital and related literatures extends to this model and implies an intuitive adjustment to the results. Finally, note that one interpretation of the social capital approach is that individuals in a group or network are aware of opportunism if and only if it is directed against a member of their group, $\theta_p^p = 1$ and $\theta_p^q = 0$. Such an assumption turns the game into two separate games with the different groups, and implies that fairness toward group p is possible if and only if $\delta \geq \delta_p$. Since cheating the minority is most tempting a solidarity-equivalent equilibrium exists if and only if $\delta \geq \delta_y$. A better outcome is possible for $\delta < \delta_y$ only if the majority has some information about cheating against the minority and is willing to use it to punish the firm.

4 Conclusion

Most economic theories of discrimination assume contracts are sufficiently complete to preclude opportunism, so such models do not allow for the simple idea that discrimination involves some people but not others being "cheated". To examine the potential for such opportunistic discrimination, we use a standard repeated trust game that is widely applied in many areas of economics to capture the effects of imperfect contracting. We find that the minority is more susceptible to opportunism than the majority even without discriminatory preferences or differences in individual attributes. The vulnerability of the minority to opportunism 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.

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 the possibility of implicit bias or preference-based bias might make particular divisions focal. We find the stronger result that such biases can interact with reputational concerns to make discrimination the unique type-stationary coalition-proof equilibrium. Hence the existence of bias against a minority, even if not widespread, may help explain why some divisions have a large effect on trust and others do not.

Appendix

Proof of Proposition 2: Proposition 1 gives the set of type-stationary PBE. We now show which of these equilibria are also coalition proof when $\phi_x = \phi_y = 0$.

First consider the no-trust equilibrium. For $\delta < \delta_s$, no-trust is unique so there is no scope for renegotiation. For $\delta \geq \delta_s$, all agents get strictly positive expected payoffs from a deviation by the grand coalition to solidarity-trust. Since solidarity-trust is an equilibrium in this range, the deviation is credible and therefore no-trust is not coalition proof.

Now consider the solidarity-trust equilibrium which exists for $\delta \geq \delta_s$. Since $1-\alpha+\delta[(1-\varepsilon)V_s+\varepsilon\beta]$ is the maximum discounted expected payoff, the only issue is whether or not the punishment subgame where a normal firm cheats is coalition-proof. If the firm has cheated in the past, all individuals believe that the firm has become inept and therefore will not trust the firm so there is no scope for renegotiation and the solidarity-trust equilibrium is coalition proof.

Now consider the minority-discrimination (y-discrimination) equilibrium which exists for $\delta \geq \delta_x$. First, just as in the solidarity-trust equilibrium, cheating an x will lead individuals to infer the firm has become inept so type x individuals will never trust again and thus there is no scope for renegotiation out of the punishment subgame. Now consider deviations that involve trust between a normal firm and 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 so consider a deviation by just the firm and y's. The deviation that is best for the firm is where the y's play a discriminatory-trust or solidarity-trust strategy. In either case the firm's payoff from cheating a y is $1+\delta[(1-\gamma_x\varepsilon)V_x+\gamma_x\varepsilon\beta]$ and from not cheating is $1-\alpha+\delta[(1-\varepsilon)V_s+\varepsilon\beta]$. But if x's stick with the discrimination strategy, we know that in the limit as $\varepsilon \to 0$, if $\delta < \delta_y$ then $1 > 1-\alpha+\delta V_y$ which implies that $1+\delta V_x > 1-\alpha+\delta V_s$ so that the firm always benefits from cheating a y individual in any subgame. Therefore for $\delta \in [\delta_x, \delta_y)$ there is no deviation to a PBE that improves on the discrimination equilibrium for those players who deviate, so the equilibrium is coalition proof. For $\delta \geq \delta_y$ such a deviation by the y's and a normal and unbiased firm to the independent-trust strategy is an equilibrium, so in this range the minority-discrimination equilibrium is not coalition proof.

Now consider the majority-discrimination (x-discrimination) equilibrium which exists for $\delta \geq \delta_y$. In this case there is an improving joint deviation by the x's and a normal and unbiased firm to the independent-trust equilibrium where the x's also adopt the discriminatory trigger strategy. Hence the majority-discrimination equilibrium is not coalition proof.

Finally, consider the independent-trust equilibrium which exists for $\delta \geq \delta_y$. For $\phi_x, \phi_y = 0$ it is equivalent to the solidarity-trust equilibrium so again the only question is whether there is scope for renegotiation and there is not by the same arguments. ■

Proof of Proposition 3: Recall that under the restrictions on ϕ_x and ϕ_y, individuals of type p will initially trust a firm if a normal firm's strategy is not to cheat, so the question is how the presence of biased firms affects renegotiation to better equilibria and renegotiation away from punishment strategies.

First consider the no-trust equilibrium. For $\delta < \delta_s$, no-trust is unique so there is no scope for renegotiation. For $\delta \geq \delta_s$, a joint deviation by the grand coalition to the solidarity-trust equilibrium is improving so no-trust is not coalition proof.

Now consider the solidarity-trust equilibrium. For $\delta \in [\delta_s, \delta_x)$ no individual can gain by changing to another equilibrium and the firm cannot gain by deviating, so the question is whether or not the punishment subgame is credible. Suppose that a normal and unbiased firm cheats the first type p individual it meets. By the assumption on the magnitude of ϕ_p relative to ε, all individuals conclude that the firm is most likely biased against members of population p, so q individuals have an incentive to form a coalition with the firm and break out of the grim punishment equilibrium. But p's no longer trust the firm and since for $\delta < \delta_q$, $1 - \alpha + \delta[(1 - \gamma_q\varepsilon)V_q + \gamma_q\varepsilon\beta] < 1$, the firm will not treat q's fairly. Thus it is impossible to jointly deviate out of the no-trust punishment equilibrium and therefore the solidarity-trust equilibrium is coalition-proof. For $\delta \geq \delta_x$, if a normal and unbiased firm cheats the first y it encounters then individuals conclude that the firm is most likely biased against members of population y and therefore x individuals have an incentive to form a coalition with the firm and continue trading. Since $\delta \geq \delta_x$, even if the y's never trust the firm again, encounters with x's are sufficiently frequent and the firm is sufficiently patient that the firm will treat x's fairly so the punishment is not coalition-proof (i.e., $\alpha - c > 0$, $V_x > 0$ and $1 - \alpha + \delta[(1 - \gamma_x\varepsilon)V_x + \gamma_x\varepsilon\beta] \geq 1$). Anticipating this, a firm that is neither inept nor biased against y's gets a payoff of $1 + \delta[(1 - \gamma_x\varepsilon)V_x + \gamma_x\varepsilon\beta]$ by cheating the first y it encounters. The firm will cheat the y whenever this is greater than $1 - \alpha + \delta[(1 - \varepsilon)V_s + \varepsilon\beta]$. In the limit as $\varepsilon \to 0$, the former is greater than the latter whenever $1 > 1 - \alpha + \delta(V_s - V_x) = 1 - \alpha + \delta V_y$ which holds whenever $\delta < \delta_y$ (equation (2)). Since all firms will therefore pool with firms that are biased against y's, the solidarity-trust equilibrium is not coalition proof.

Now consider the minority-discrimination equilibrium. For $\delta \in [\delta_x, \delta_y)$, in the event of a joint deviation to solidarity, the firm's payoff to treating a y fairly is $1-\alpha+\delta[(1-\varepsilon)V_s+\varepsilon\beta]$ whereas the payoff from cheating the y is $1 + \delta[(1 - \gamma_x\varepsilon)V_x + \gamma_x\varepsilon\beta]$. But as we've just

shown, in the limit as $\varepsilon \to 0$, the former is strictly less than the latter whenever $\delta < \delta_y$. Therefore there is no improving joint deviation to a PBE, and the minority-discrimination equilibrium is coalition-proof. For $\delta \geq \delta_y$, there is an improving joint deviation by the x's and a normal and unbiased firm to the independent-trust equilibrium where the x's also adopt the discriminatory trigger strategy, so for this range the minority-discrimination equilibrium is not coalition proof.

Now consider the majority-discrimination equilibrium. For $\delta \geq \delta_y$ there is an improving joint deviation by the y's and a normal and unbiased firm to the independent-trust equilibrium, so majority-discrimination is not coalition proof.

Finally, consider the independent-trust equilibrium which exists for $\delta \geq \delta_y$. Since no individual can gain by changing to another equilibrium and the firm cannot gain by deviating, the question is whether the punishment strategies for x and y types are credible. Suppose that a normal and unbiased firm cheats the first p it encounters. All individuals conclude that the firm is either biased against members of population p or that the firm has become inept. In either case non-trust is a dominant strategy for members of population p, so the punishment is credible. ■

References

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

Annen, K. (2003), "Social Capital, Inclusive Networks, and Economic Performance," *Journal of Economic Behavior and Organization*, 50:449–463.

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

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

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

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

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

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

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

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

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

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

Black, D. A. (1995), "Discrimination in an Equilibrium Search Model," *Journal of Labor Economics*, 13:309–334.

Bowles, S., G. C. Lowry and R. Sethi (2010), "Group Inequality," working paper.

Charles, K. K. and J. Guryan (2008), "Prejudice and Wages: An Empirical Assessment of Becker's The Economics of Discrimination," *Journal of Political Economy*, 116:773–809.

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

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

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

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

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

Douglass, F. (1879), "Negro Exodus from the Gulf States," speech before American Social Science Association in Saratoga, New York.

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

Eeckhout, J. (2006), "Minorities and Endogenous Segregation," *Review of Economic Studies*, 33:31–53.

Ekmekci, M., O. Gossner and A. Wilson (2012), "Impermanent Types and Permanent Reputations," *Journal of Economic Theory*, 32:162–178.

Esteban, J. and D. Ray (2008), "On the Salience of Ethnic Conflict," *American Economic Review*, 98(5):2185–2202.

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

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

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

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

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

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

Ghosh, P. and D. Ray (1996), "Cooperation in Community Interaction Without Information Flows," *Review of Economic Studies*, 63:491–519.

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

Greenwald, A. G., D. E. McGhee and J. L. K. Schwartz (1998), "Measuring Individual Differences in Implicit Cognition: The Implicit Association Test," *Journal of Personality and Social Psychology*, 74:1464–1480.

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

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

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

Kandori, M. (1992), "Social Norms and Community Enforcement," *Review of Economic Studies*, 59:63–80.

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

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

Kreps, D. M., P. Milgrom, J. Roberts and R. Wilson (1982), "Rational Cooperation in the Finitely Repeated Prisoners' Dilemma," *Journal of Economic Theory*, 27(2):245–252.

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

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

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

— (2007), "Information and Racial Exclusion," *Journal of Population Economics*, 20:621–642.

Mailath, G. J. and L. Samuelson (2001), "Who Wants a Good Reputation?" *Review of Economic Studies*, 68:415–441.

— (2006), *Repeated Games and Reputations: Long-Run Relationships*, Oxford University Press.

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

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

Munshi, K. (2011), "Strength in Numbers: Networks as a Solution to Occupational Traps," *Review of Economic Studies*, 40:1069–1101.

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

Peski, M. and B. Szentes (2013), "Spontaneous Discrimination," *American Economic Review*, 103:2412–36.

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

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

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

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

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

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

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