A GAME THEORETIC APPROACH TO ECONOMIC POLICY ANALYSIS

Peter Norman

A DISSERTATION

in

Economics

Presented to the Faculties of the University of Pennsylvania in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy

1997

Andrew Postlewaite, Supervisor of Dissertation

George Mailath, Graduate Chair

UMI Number: 9800907

Acknowledgments

I am very grateful for the opportunity to write my thesis at University of Pennsylvania. Most of all I want to thank the three persons who have been supporting me constantly since I started the work with my dissertation: I stand in great debt to my Dissertation supervisor Andrew Postlewaite for insightful comments, excellent guidance, interesting discussions on my work as well as on more general issues in economics and his incredible patience. I am equally indebted to Stephen Morris who provided excellent general as well as many detailed comments on my work, made me avoid several dead ends and taught me a lot of game theory. Stephen Coate has been of enormous help in formulating interesting questions for research and has also provided many very useful comments and pointed out interesting interpretations of my results. All three also provided necessary moral support and encouragement in darker times and have always been willing to sacrifice their time: I could not have had a better team of advisors.

Many other faculty-members, visitors and fellow graduate students have been very helpful during my time here at Penn. In particular I want to thank Illtae Ahn, Han-Ming Fan, Marcos Lisboa, George Mailath, Andrea Moro, Motty Perry, Anna Rubinshick, Ken Wolphin and Alvaro Sandroni for comments and discussions. I also want to thank Regina Forlano and Thea Turnage for good guidance on how to find my way through the administrative labyrinths of the university.

I also want to thank many people at the Institute of International Economic Studies in Stockholm for continuous support, encouragement and hospitality during the summer months when Philadelphia is not the most livable place on earth. In particular I want to thank Henrik Horn for good advise, comments, discussions and moral support, Torsten Persson for comments and discussions and Nils Gottfries for talking me into crossing the Atlantic. I have also learned a lot in discussions with many others, in particular David Domeij, Harry Flam, Lars Persson, Björn Segendorff, Jakob Svenson and Johan Stennek. I also want to express my deep gratitude to Jörgen W. Weibull at Stockholm School of Economics who have been very helpful and supportive, both before and after I was admitted at University of Pennsylvania.

Finally I want to thank Anna, Samuel and Vendela for making a career path as a research economist at all possible.

ABSTRACT

A GAME THEORETIC APPROACH TO ECONOMIC POLICY ANALYSIS

Peter Norman

Andrew Postlewaite

In the first chapter, *Legislative Bargaining and Coalition Formation,* I analyze a widely used model of endogenous policy determination. The main question is if the stationary equilibrium, a popular equilibrium selection in applied work, can be rationalized as approximating the unique backwards induction equilibrium with a long finite horizon. The answer is negative: I show that there is a continuum subgame perfect equilibrium outcomes in the finite horizon model. With sufficiently patient players a "folk theorem" applies. In contrast, I obtain generic uniqueness of equilibria in a generalized model where differences in time preferences are allowed. However, a unique equilibrium is always non-stationary and even if differences in discount rates are arbitrarily small, the non-stationarity is non-negligible. In the second chapter, *Affirmative Action in a Competitive Economy* (with Andrea Moro), the objective is to investigate how affirmative action affects incentives for human capital acquisition in a model of statistical discrimination with endogenous human capital. Affirmative action may "fail" in the sense that there may still be discrimination in equilibrium. However, the incentives to invest for the disadvantaged group are better in *any* equilibrium under affirmative action than in the most discriminatory equilibrium without the policy. Welfare effects are ambiguous: the policy may even hurt the intended beneficiaries, also when the initial equilibrium is the most discriminatory equilibrium. The third chapter, *Statistical Discrimination and Efficiency,* investigates if there is an efficiency rationale for policies aimed at statistical discrimination. In the same framework as the second chapter I show that there is always an efficiency rationale for intervention. However, the inefficiencies arise because of a "free-riding" problem, which occurs independently of whether there is discrimination or not. A planner may want to discriminate between two identical groups of workers. The reason is that discrimination makes it possible for certain workers to specialize as qualified workers, which reduces the problem to match workers with jobs efficiently. This positive effect must be weighted against inefficiencies in investment behavior created by discrimination, but examples can be constructed where the surplus maximizing allocation involves discrimination.

CONTENTS

1 Legislative Bargaining and Coalition Formation 1

1.1 Introduction 2

1.2 The Model 5

1.3 Equilibria 7

1.3.1 The Role of Sequential Voting 15

1.4 Heterogeneous Time Preferences 16

1.5 Concluding Remarks 21

1.6 Appendix 22

1.6.1 Proof of Proposition 2 22

1.6.2 Proof of Lemma 1 34

1.6.3 Proof of Lemma 3 36

2 Affirmative Action in a Competitive Economy 39

2.1 Introduction 40

2.2 The Model 44

2.2.1 The Game 45

2.3 Characterization of Equilibria 48

2.3.1 Why Use a Strategic Model? 53

2.4 The Model with Two Identifiable Groups of Workers 54

2.4.1 The Extended Model 55

2.4.2 Equilibrium in the Extended Model 55

2.5 Affirmative Action 58

2.5.1 The Model with Affirmative Action 59

2.5.2 Welfare Effects of Affirmative Action 66

2.6 Discussion 68

2.7 Appendix 69

2.7.1 Proof of Proposition 4 69

2.7.2 Proof of Proposition 6 75

2.7.3 Proof of Proposition 7 78

2.7.4 Proof of Proposition 8 79

3 Statistical Discrimination and Efficiency **88**

3.1 Introduction 89

3.2 The Model 93

3.2.1 The production technology 93

3.2.2 Human capital investments and the screening technology 93

3.3 Optimal Policies in a Command Economy 94

3.3.1 The planning problem 95

3.3.2 Characterization of the surplus maximizing plan 96

3.3.3 Example: Discrimination By the Planner 101

3.4 The Decentralized Model 106

3.4.1 The Model 106

3.4.2 Equilibrium 107

3.4.3 Underinvestment in equilibrium 108

3.5 Discussion 110

3.6 Appendix 111

3.6.1 Proof of Lemma 22 111

3.6.2 Proof of Lemma 23 113

3.6.3 Proof of Proposition 10 113

List of Figures

2.1 The task-assignment problem 50

2.2 An example with a unique interior equilibrium 53

2.3 Not an equilibrium 61

2.4 Equilibrium wage schedules under affirmative action 62

2.5 A profitable deviation 74

1. LEGISLATIVE BARGAINING AND COALITION FORMATION

Peter Norman[1]

University of Pennsylvania

November 21 1996

Abstract

The finite horizon version of a much studied legislative bargaining model due to Baron and Ferejohn is investigated. It is shown that if there are three or more rounds of bargaining, then a continuum of distributions are supportable as subgame perfect equilibria. In fact, the result holds true even if one restricts attention to equilibria in Markov strategies. If the players are sufficiently patient a folk theorem applies: any distribution of benefits such that all players receive strictly positive shares can be supported as a subgame perfect equilibrium. In contrast to this result we obtain a generic uniqueness result when allowing for differences in the players time preferences. However, whenever there is a unique equilibrium it is also highly non-stationary and the non-stationarity does not disappear in the limit as the discount factors converge towards a common discount factor. Hence, the unique backwards induction equilibrium will not converge to the stationary equilibrium in the original game when we consider a sequence of games with payoff functions converging towards the payoffs of the original game.

[1] I thank Stephen Coate, George Mailath, Andrea Moro, Stephen Morris, Motty Perry and Andrew Postlewaite for comments and interesting discussions. Remaining errors are mine. Correspondence to *pnorman@econ.sas.upenn.edu*.

1.1. Introduction

In an influential paper, Baron and Ferejohn [6] set up a tractable model of legislative voting with endogenous agenda setting. The model has later been generalized and applied to a wide variety of questions in economics and political science. For example, Baron and Ferejohn [7] use the model to analyze the role of committees in the legislative process, Baron [5] studies how legislative equilibria depend on characteristics on the goods provided, Chari et al [11] use the model in their analysis of split ticket voting and in McKelvey and Riezman [20], Baron and Ferejohn's model is an important building block in a political economy model where a seniority system is derived as an equilibrium outcome.

The general idea behind the approach is that real world legislators must obey certain rules on how and when to make proposals and vote. Thus it seems natural to proceed as in the literature on bargaining and capture these constraints within some dynamic game and use non-cooperative solution concepts to obtain predictions. Since the rules of the game creates "frictions" in the bargaining process by restricting each agents' possibilities of making proposals one may now obtain testable implications even in situations where a non-institutional approach provides weak or no predictions at all. In particular, there may now be equilibria when there for each proposal exists a counterproposal preferred by a majority, so there is hope that situations where legislators have profoundly conflicting preferences can be handled by this type of models.

This view is advanced by Baron and Ferejohn [6], who consider a stylized model of legislative bargaining where the task of the legislature is to divide a given budget among its members. One may interpret the model as one where each legislator represents a particular district and where the problem for the legislators is how to allocate spending over the districts. The formal game is essentially an $n-$player version of Rubinstein's [26] bargaining model, with the important difference that a proposal is accepted whenever a majority of the players vote in favor of the proposal, while unanimous agreement is typically required in sequential bargaining models.

Baron and Ferejohn [6] show that if players are sufficiently patient, then any division can be supported as a subgame perfect equilibrium in the infinite horizon version of their model. However, they argue that equilibria based on infinitely nested punishments may be difficult to enforce and restrict attention to *stationary* (Markov perfect) equilibria in the infinite game, i.e. equilibria where all players behave in the same way in all structurally equivalent subgames. They show that there is a unique stationary equilibrium and in their paper, as well as in most applications of their model, the focus is exclusively on the stationary equilibrium.

There are several attractive features of the stationary equilibrium. Obviously, stationary equilibria makes the analysis more tractable and in the context of Baron and Ferejohn's model it can be made precise that the stationary equilibrium is the simplest equilibrium in a well defined sense (see Baron and Kalai [8]). An alternative argument in favor of this equilibrium selection that has also been proposed: there is an "essentially" unique subgame perfect equilibrium in the finite horizon version of the model and the stationary equilibrium can be viewed as the limit of the backwards induction solution as the number of bargaining rounds is taken to infinity. Note that this claim corresponds well to the usual intuition about equilibria in strategies that are only a function of payoff relevant variables: when agents are not allowed to condition play on payoff irrelevant parts of the history, then "bygones are bygones" and the analysis is as if one solved for the backwards induction equilibrium of the corresponding game with a finite horizon.

In this paper we study the finite horizon version of Baron and Ferejohns model. We show that, for this particular game, the intuition that the stationary equilibrium corresponds to (the unique) backwards induction solution of the corresponding finite horizon model is wrong: there will simply not be a unique backwards induction equilibrium. In fact, with three or more rounds of bargaining a continuum of divisions are supportable as subgame perfect equilibrium outcomes. It is also shown that if there are sufficiently many rounds of bargaining and the players are sufficiently patient, then any distribution of benefits such that all players receive a strictly positive share is supportable as a subgame perfect equilibrium.

Since we study a game of perfect information, the non-uniqueness may be surprising at a first glance. As is well known, there is a unique backwards induction solution for any game of perfect information with generic payoffs, so there must be something non-generic with the payoffs of the game. Interestingly enough, what creates the non-genericity is in a sense what makes the model interesting: the proposer in the penultimate session must "bribe" sufficiently many players to get a majority for her proposal. However, by symmetry of the game, the value of the game in the beginning of the last period is the same for all players, so all players have the same acceptance rules in the penultimate period. This means that the proposer in the penultimate session can choose what players to "bribe" in any way he wants. Hence it is costless to punish and reward players by making the identities of the players to be included in the winning majority dependent on the history of play. Of course, non-uniqueness of equilibria follows directly from the observation that the identities of the players selected by the proposer is indeterminate. However, if there are three or more rounds of bargaining this in turn implies that we obtain a more interesting multiplicity: if players are selected to be included in the "winning majority" with different probabilities, then the value of the game in the beginning of the penultimate period is different for different players. Thus, the acceptance rules will be different in the preceding session, so some players will be more expensive to bribe than other players. As we will see, this makes it possible to support a continuum of divisions as equilibrium outcomes *even in strategies that are independent of payoff irrelevant history.*

Since the argument above depends crucially on the symmetry of the game one may guess that the intuition that the stationary equilibrium may be obtained as a limit of a sequence of backwards induction equilibria may be restored by breaking the symmetry of the model in such a way so that the indifferences for the proposer in the penultimate period disappears. To see if this intuition is correct we allow the discount factors to vary between individuals. In this case we show that for generic choices of discount factors there is in fact a unique subgame perfect equilibrium of the model. Here, generic means that the set of (vectors of) discount factors such that there is not a unique equilibrium has Lebesgue measure zero, so we may interpret the result as saying that if the n players discount factors are independent draws from a uniform distribution over the unit interval, then we would get discount factors generating a unique equilibrium with probability one.

However, whenever we obtain a unique equilibrium this turns out to be highly non-stationary. Moreover, the non-stationarity does not disappear when the discount factors converge to a common discount factor for all players. Hence, while we typically get uniqueness of the equilibria when we recognize the possibility that the players may have different time-preferences this does not really help us to select equilibria in the infinite game. The same thing holds true for other natural asymmetries. For example, if there are spillover effects between "nearby districts", the same logic goes through. Also, if the payoff functions are as in the original model, but the probabilities of getting the opportunity of making an offer are different for different players, the proof of the generic uniqueness result goes through with some minor modifications and the qualitative properties of the equilibrium is as with differences in time preferences.

While one may interpret these results as mainly negative they do actually provide some insights as to how sequential bargaining with majority rule (or any other rule different from unanimity rule) differs from the standard setup where unanimous agreement is required: if no player has veto power it is possible that it is undesirable to have a strong position in later stages of the game. The reason is that being strong in the future makes the player more expensive to bribe in earlier stages and therefore more likely to be excluded from the "winning coalition". Because *all* players must get at least their respective value of waiting to be willing to accept a proposal in standard sequential bargaining models, these type of considerations do not emerge in models where everybody must agree in order for a proposal to be implemented.

The remainder of this chapter is organized as follows. In the next section the basic model is set up. In Section 1.3 some simple examples are analyzed in detail and the finite horizon folk theorem is stated. In Section 1.4 the model is extended to allow for heterogenous time preferences. The discussion in Section 1.5 concludes the chapter. Most proofs are relegated to the appendix in Section 1.6.

1.2. The Model

The set up is as in Baron and Ferejohn[6]. Following their interpretation of the model we may think of a body of legislators, each representing the voters of their district who are to make a decision about how to distribute some benefits. Specifically we let:

the set of players (legislators) be given by $I = \{1, .., n\}$,

the set of potential "outcomes" be given by $X = \{0\} \cup \left\{ x \in R^n_+ \mid \sum_{i=1}^{n} x^i \leq w \right\}$, where $w > 0$[2],

each players' instantaneous utility function be given by $u_i : X \to R$ defined by $u_i(x) = x^i$. Payoffs are discounted by the common discount factor $\delta \in (0, 1]$

Hence, the "bargaining problem" is a n player cake-splitting problem. In Baron and Ferejohn [6] one of the main objectives is to study how the equilibria of the game depends on the agenda setting rules, so several extensive forms are considered. In this paper we will restrict attention to the simplest case [3]. The "institutional" assumptions are as follows;

Proposals. If the legislature has not agreed on any proposal in periods $1, ..., t-1$ and if $t \leq T$, one randomly chosen player is *recognized* to make a proposal in the beginning of period t. The probability distribution over who is selected at time t is assumed to be a time and history independent distribution $\mu \in \Delta^n$. As in Baron and Ferejohn we take $\mu_i = \frac{1}{n}$.

[2] I.e. we assume "free disposal". We could at the cost of some additional complexity of the proofs take $X = \{x \in R^n_+ \mid \sum x^i = w\}$. However, since it is feasible that no agent receives nothing it is not clear why one would want to restrict agents to propose allocations on the boundary.

[3] This specification is referred to as "the closed rule" in Baron and Ferejohn [6] and Baron and Ferejohn [7].

Voting. When a player j has made a proposal $x \in X$, the legislature votes sequentially[4] and according to a predetermined order between accepting and rejecting the proposal. If a majority accepts x, the proposal is implemented and the game ends with payoffs $\delta^t x^i$ for all players $i \in I$. If a majority decides to reject and $t = T$ the game ends with the zero payoffs being realized for all players. If a majority decides to reject and $t < T$ the game goes on to the next stage.

A history at the beginning of time t is given by a list $h_t = (a_1, a_2, ..., a_{t-1})$ where $a_\tau = (i_\tau, x_\tau, v_\tau^1, ..., v_\tau^n)$ is the "action profile" for periods $\tau = 1, 2, ...t-1$, $i_\tau \in I$ is the identity of the proposer in session τ, $x_\tau \in X$ is the proposal made at time τ and $v_\tau^j \in \{yes, no\}$ is the vote cast by player j in session τ. We let H_t denote the set of all possible (beginning of) time t histories. However, voting is performed sequentially, so each player knows not only h_t, but also how all players that are voting before her in the current session have voted. The order of voting in every session is predetermined and known by all players. To translate this into formalism we assume that for each t there is an invertible function $k_t : I \to I$ that generates a voting ordering in session t[5]. We let h_t^k denote a generic history when $k-1$ votes have been cast in session t and denote H_t^k the set of all such histories. Note that we can write a history when the kth player is just about to vote as $h_t^k = \left(h_t, j_t, x_t, v_t^{k_t^{-1}(1)}, ..., v_t^{k_t^{-1}(k-1)}\right)$. For example if k_t is the identity function so that player 1 votes first, player 2 second and so on, then $h_t^k = \left(h_t, j, v_t^1, ..., v_t^{k-1}\right)$. A (pure) strategy s^i for player $i \in I$ is given by $s^i = \{p_t^i, v_t^i\}_{t=1}^T$ where $p_t^i : H_t \to X$ lists a proposal for each possible history at time t and $v_t^i : H_t^{k_t(i)} \to \{yes, no\}$ describes the voting behavior of player i for every possible history. A typical mixed (behavioral) strategy for i will be denoted $\sigma_i = \{\widetilde{\sigma}_t^i, \widehat{\sigma}_t^{ik}\}_{t=1}^T$, where $\widetilde{\sigma}_t^i : H_t \to \Delta(X)$ and $\widehat{\sigma}_t^{ik} : H_t^k \to [0,1]$.

[4]Whether voting is performed sequentially or simultaneously does not make much of a difference in the finite horizon model, given that we in the case of simultaneous voting eliminate conditionally weakly dominated strategies.

[5]Hence, $k_t(i)$ is the position in the voting order for player i in session t and $k_t^{-1}(i)$ gives the identity of the person in the i^{th} place in the voting order at time t. For the particular equilibria we construct in this paper the voting order does not matter, indeed we could even make the voting order to depend on the history h_t. However, there may be other equilibria where the order actually matters.

1.3. Equilibria

To gain some intuition and to make the mechanics of the model clear, we will begin this section by considering two examples. In the first example, it is shown that with three rounds of bargaining there is multiplicity in the set of divisions that are supportable as equilibrium outcomes even with history independent strategies. The result that there is a continuum of distributions supportable as (Markov perfect) equilibrium outcomes follows as a slight generalization of the example. In the second example, it is shown that there are subgame perfect equilibria that are at odds with Riker's *minimal winning coalition* theorem in the sense that more than $(n+1)/2$ players may get a positive share of the surplus in equilibrium. After these examples the general result is stated and discussed.

Example 1.

We will consider the simplest example possible, with $T = 3$, $n = 3$ and $\delta = 1$. The size of the pie is without loss of generality normalized to unity and we proceed by solving the game backwards.

The Last Session

Suppose that player j is recognized in the last session and proposes x_3 with $x_3^j = 1$ and $x^i = 0$ for $i \neq j$. Since all players get 0 if the proposal is rejected the voting rule "accept any proposal after any history" implies Nash equilibrium play in every subgame starting after the last sessions proposal. Given these voting rules there is no profitable deviation for the proposer Thus, there are subgame perfect equilibria where all player proposes to take everything if called to make a proposal in the final period and where these proposals are accepted for sure.

In fact, this is the unique proposal in the last session consistent with subgame perfection. To see this, suppose that $x_3^i > 0$ for at least two players. We first argue that there are no voting rules consistent with subgame perfection such that the proposal is rejected. The argument illustrates how sequential voting makes it unnecessary to eliminate weakly dominated strategies. Consider the choice of the last of the two players who receives strictly positive payoffs under x^i. If she is pivotal, she will always vote in favor of the proposal. Hence the first of the two players with strictly positive payoffs knows that if she accepts, then the proposal will be implemented. The first player in the voting order that receive strictly positive payoffs must therefore vote to accept in any subgame perfect equilibrium, *unless the other players follow such voting rules so that the proposal will be implemented no matter how she votes.* Thus the proposal must always be accepted in any subgame perfect equilibrium. Now, if $x_3^i > 0$ for some $i \neq j$ there exists ε small enough so that $x_3^i - \varepsilon > 0$. This proposal will be accepted for sure, but so will the alternative proposal y with $y_j = x_3^j + \varepsilon$, $y_i = x_3^i - \varepsilon$ (and the third players share unchanged). Since the proposer is better off, x_3 cannot be proposed in the last session in any subgame perfect equilibrium. Thus we conclude that the only proposals in the last period consistent with subgame perfection is for the last proposer to take the whole pie for herself. I.e. player 1 proposes $(1, 0, 0)$, player 2 $(0, 1, 0)$ and player 3 $(0, 0, 1)$ and these proposals will be accepted for sure.

The Second Session

Since each player gets 1 with probability $\frac{1}{3}$ and 0 with probability $\frac{2}{3}$ if the game continues to the final stage, the continuation value for player any player i in the beginning of the last session is $\frac{1}{3}$. Thus if a player is pivotal in the voting stage of the second session, she must accept proposals with $x_i^j(2) \geq \frac{1}{3}$ and reject if $x_i^j(2) < \frac{1}{3}$.[6] Consequently, if two players receive $x_i^j(2) \geq \frac{1}{3}$ the proposal will be accepted, while any other proposal must be rejected in any subgame perfect equilibrium. Hence, anyone selected in the second period will offer $\frac{1}{3}$ to one of the other players and keep $\frac{2}{3}$ for herself[7].

A Symmetric Equilibrium

Consider the following proposal rule in the second session: regardless of history, each player selects one player by randomizing with equal probabilities over the remaining players. She then proposes to give this randomly select player $\frac{1}{3}$ and keep $\frac{2}{3}$ for herself. By the analysis above these proposals will always be accepted in any subgame perfect equilibrium. Since proposals that gives the proposer a larger share must be rejected in any equilibrium, the proposer has no incentive

[6]As in most other bargaining models, ties in the voting stage must be resolved in favor of accepting the proposal with probability one in any subgame perfect equilibrium.

[7]Since the proposer can randomize arbitrarily between the remaining players when selecting who is to receive part of the cake this means that there is a continuum of equilibria in any subgame starting in session 2. The rest of the example illustrates that there is also multiplicity in terms of distribution of benefits that goes beyond the set of possible permutations of $\left(\frac{2}{3}, \frac{1}{3}, 0\right)$.

to deviate. Hence the proposed strategy profile constitutes a subgame perfect equilibrium in the game with two rounds of bargaining. We can compute the value of the game in the beginning of the second period (before identity of the proposer is known) as

$$\Pr\left[i \text{ recognized}\right]\frac{2}{3}+\left(1-\Pr\left[i \text{ recognized}\right]\right)\Pr\left[j \text{ gives } \frac{1}{3} \text{ to } i\right]\frac{1}{3}=\frac{1}{3}\frac{2}{3}+\frac{2}{3}\frac{1}{2}\frac{1}{3}=\frac{1}{3}$$

for $i = 1, 2, 3$. Hence, *given that all players follow these proposal rules in the second session*, the voting rules in the first session must be the same as in the second session and following the analysis of the second session we conclude that any player selected in the first period to make a proposal must keep $\frac{2}{3}$ for herself and give $\frac{1}{3}$ to one of the other players[8].

An Asymmetric Equilibrium

In the second period, let the players follow the following proposal rules after any history of play:

If player 1 is recognized, she proposes $x^1(2) = \left(\frac{2}{3}, \frac{1}{3}, 0\right)$.

If player 2 is recognized, she proposes $x_2^2(2) = \left(\frac{1}{3}, \frac{2}{3}, 0\right)$.

If player 3 is recognized, she proposes $x_2^3(2) = \left(\frac{1}{3}, 0, \frac{2}{3}\right)$.

All proposals according to these rules must be accepted in any subgame perfect equilibrium. Given these proposal rules, the continuation values of the game in the beginning of the second session, which we denote by $V^i(2)$ for players $i = 1, 2, 3$, are given by:

$V^1(2) = \mu_1\frac{2}{3} + \sum_{i=2,3}\mu_i \Pr\left[i \text{ gives } \frac{1}{3} \text{ to } 1\right]\frac{1}{3} = \frac{1}{3}\frac{2}{3} + \frac{1}{3}\frac{1}{3} + \frac{1}{3}\frac{1}{3} = \frac{4}{9}$

$V^2(2) = \mu_2\frac{2}{3} + \sum_{i=1,3}\mu_i \Pr\left[i \text{ gives } \frac{1}{3} \text{ to } 2\right]\frac{1}{3} = \frac{1}{3}\frac{2}{3} + \frac{1}{3}\frac{1}{3} = \frac{3}{9}$

$V^3(2) = \mu_3\frac{2}{3} + \sum_{i=1,2}\mu_i \Pr\left[i \text{ gives } \frac{1}{3} \text{ to } 3\right]\frac{1}{3} = \frac{1}{3}\frac{2}{3} = \frac{2}{9}$

Consider the following voting rules for the first session:

Player 1 accepts the proposal $x^j(1)$ if and only if $x_1^j(1) \geq \frac{4}{9}$

[8]One possibility is for all players to randomize with equal probabilities in the first period as well which implies symmetric ex ante payoffs. As is easily seen we can proceed this way with arbitrary many rounds of bargaining.

Player 2 accepts the proposal $x^j(1)$ if and only if $x_1^j(1) \geq \frac{3}{9}$

Player 3 accepts the proposal $x^j(1)$ if and only if $x_1^j(1) \geq \frac{2}{9}$

Since the continuation strategies are history independent one easily verifies that these voting rules are sequentially rational given the continuation play[9]. Given these voting rules it is clear that it is much cheaper to select player 3 in a winning majority than to select player 1. It is readily verified that taking these continuation strategies into account the optimal way to make proposals for the respective players in the first period is:

If player 1 is recognized in period one she proposes $x^1(1) = \left(\frac{7}{9}, 0, \frac{2}{9}\right)$

If player 2 is recognized in period one she proposes $x^2(1) = \left(0, \frac{7}{9}, \frac{2}{9}\right)$

If player 3 is recognized in period one she proposes $x^3(1) = \left(0, \frac{3}{9}, \frac{6}{9}\right)$

The proposed strategies constitutes a subgame perfect equilibrium of the game by construction. We see that two of these three distributions on the outcome path do not correspond to the symmetric equilibrium. Hence we have shown that there is multiplicity of equilibria that goes beyond the identities of the players who are selected for the winning majority. To see that there is indeed a continuum of divisions supportable as subgame perfect equilibrium outcomes, just let π be the probability that players behave as in the (particular) asymmetric equilibrium in the second session and $(1-\pi)$ the probability that the players behave as in the symmetric equilibrium. With this specification the ordering of the players continuation payoffs remains the same for any $\pi \in (0,1]$ and it is easy to see that by choosing π appropriately we can support any $x^1 = \left(s^1, 0, 1-s^1\right)$, where $s^1 \in \left[\frac{2}{3}, \frac{7}{9}\right]$, as an equilibrium outcome (realized if player one proposes in first session).

If we increase the number of sessions, any distribution supportable as a subgame perfect equilibrium in the 3-period model is supportable in the T-period model. In other words, the set of equilibrium outcomes can only get larger as the number of periods increases. To see this, note that if, after any history h_t, the continuation values at time t are given by $V^1(t) = V^2(t) = V^3(t) = \frac{1}{3}$, then a proposal will be accepted if and only if at least two players get at least $\frac{1}{3}$. Thus, the player recognized to be the proposer at time $t-1$ will propose to keep $\frac{2}{3}$ and give $\frac{1}{3}$ to one of the other players. If the indifference is resolved by letting $\Pr\left[i \text{ gives } \frac{1}{3} \text{ to } j \text{ at time } t-1 \mid i \text{ recognized}\right] = \frac{1}{2}$ for all i,j the continuation value before the random draw at time $t-1$ is again given by $V^i(t-1) = \frac{1}{3}$

[9]There are also more complicated voting rules consistent with subgame perfection, but all we are doing here is to construct a particular equilibrium.

for $i = 1, 2, 3$[10]. By induction it follows that we can construct a symmetric equilibrium for any $T-$period game with continuation payoffs given by $\left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right)$ in each period. Now consider the class of equilibria of the T-period game where all players are playing in accordance to the symmetric equilibrium strategies whenever they are called to play in the third session or later. The (history independent) continuation payoffs in the beginning of session 3 (before the identity of the proposer is known) in this class of equilibria equals the (history independent) continuation payoff in the beginning of the last session in our 3-period example. Hence, anything that can be supported as an equilibrium with three rounds of bargaining can also be supported with T rounds. We can then conclude that there is a continuum of divisions supportable as subgame perfect equilibria if there are more than three rounds of bargaining.

While we only considered the case with three players and no discounting, the logic had nothing to do with these assumption. Arguing as in the example above one can show:

Proposition 1. *Suppose that $\delta > 0$, $T \geq 3$ and $n \geq 3$. Then there is a continuum of divisions supportable as subgame perfect equilibria of the model.*

The proof is a straightforward extension of the example above and is omitted. Arguing as in the example one shows that in the second session from the end, any player who is selected to make a proposal will propose to distribute δ/n to $(n-1)/2$ of the other players and keep $1 - \delta(n-1)/2n$ for herself. We can then construct a symmetric and some asymmetric equilibrium as in the example. The last step is then to show that, no matter which player i who is recognized in the first session, one can always support any division in between the two particular equilibrium divisions.

One interesting feature of the asymmetric equilibrium in the example is that the expected payoff is for player one is lower than for the other two players. At the same time, player one is the player who, conditionally on reaching it, has the strongest position in the second round. While this helps player one to take a somewhat larger share *if making a proposal in the initial session*, it is a disadvantage if any of the other players is making a proposal: the other players have less to gain by turning down proposals, so player one will simply be excluded from the winning coalition in the first session.

[10]Note that symmetric continuation values can also be achieved by pure strategies. For example if player 1 selects player 2, player 2 selects player 3 and player 3 selects player 1 we get the same continuation valuations as from randomizing with equal probabilities.

Example 2 (history dependent strategies).

The purpose of this example is to show that the set of equilibrium outcomes is enlarged even further if history dependent strategies are allowed. In particular, there are now equilibria that are qualitatively different from the set of equilibria in Markov strategies: there are now equilibria where all players receive a positive share of the benefits, that is, equilibria at odds with Riker's[25] *minimal winning coalition theorem.*

To minimize the algebraic complexity we again consider the case with $\delta = 1, n = 3$ and $T = 3$. As in the previous example these restrictions are purely for expositional convenience. For $i = 1, 2, 3$ we let $x^i \in X$ be some outcome (that we want to support as an equilibrium outcome when i is making a proposal in the initial session) and suppose the players follow the following strategies if called to make a proposal in the second session.

If player $i \in \{1, 2, 3\}$ proposed x^i in session 1 and is recognized again in session 2 she keeps $\frac{2}{3}$ for herself and randomizes with equal probabilities over which of the other players should be offered $\frac{1}{3}$.

If player $i \in \{1, 2, 3\}$ proposed $x' \neq x^i$ in session 1 and is recognized again in session 2 she keeps $\frac{2}{3}$ and give $\frac{1}{3}$ to the player who was offered the most in period 1 (if $x^{ij} = x^{ik}$ the proposer flips a fair coin)

If player $j \neq i$ proposed $x' = x^j$ in session 1 player i proposes to take $\frac{2}{3}$ for herself and give $\frac{1}{3}$ to player j if recognized to make a proposal in session 2.

If player $j \neq i$ proposed $x' \neq x^j$ in session 1 player i proposes to take $\frac{2}{3}$ for herself and give $\frac{1}{3}$ to player who did not propose in the first session.

Since some permutation of $(\frac{2}{3}, \frac{1}{3}, 0)$ with the proposer getting $\frac{2}{3}$ is proposed after any history of play we know by the analysis in Example 1 that these proposal rules together with the uniquely determined backwards induction continuation strategies implies Nash equilibrium play in any subgame from the beginning of the second session and on.

Given the above proposal rules the continuation values of the game in the beginning of the second session will depend on the history of play, i.e. whether the "target outcome" x^i was proposed or not.

Suppose that player i was recognized in session 1, proposed a division $x' \neq x^i$ such that $x'^j \neq x'^k$ and that this proposal was rejected by a majority. After any such history the continuation values in the beginning of the second session are given by;

V_2^i $\left(i \text{ defected from } x^i\right) = \frac{1}{3}\frac{2}{3} = \frac{2}{9}$ for the player who proposed in session one

V_2^j $\left(i \text{ defected from } x^i\right) = \frac{1}{3}\frac{2}{3} + \frac{2}{3}\frac{1}{3} = \frac{4}{9}$ for the player who was offered the most in period 1

V_2^k $\left(i \text{ defected from } x^i\right) = \frac{1}{3}\frac{2}{3} + \frac{1}{3}\frac{1}{3} = \frac{3}{9}$ for the player who was offered the least in period 1

If the deviation is such that $x'^j = x'^k$, then the continuation value for the player who proposed is unchanged, but is $\frac{7}{18}$ for the other two players. Combining these two cases we see that a first period deviation from x^i will be accepted if and only if either[11],

the proposer offered at least $\frac{4}{9}$ to one of the other players or,

the proposer offered at least $\frac{3}{9}$ to one of the other players and strictly more than $\frac{3}{9}$ to the other or,

the proposer offered at least $\frac{7}{18}$ to both the other players.

Inspecting the alternatives above we see that best way for the proposer to deviate from x^i is to take $\frac{5}{9}$ and give $\frac{4}{9}$ to any of the other players.

Next we consider the value of the game in the second session after a history where the rejected proposal was in accordance to the proposed equilibrium strategies. Then:

V_2^i $\left(i,\ x^i\right) = \frac{1}{3}\frac{2}{3} + \frac{2}{3}\frac{1}{3} = \frac{4}{9}$ for the first period proposer

V_2^j $\left(i, x^i\right) = \frac{1}{3}\frac{2}{3} + \frac{1}{3}\frac{1}{2}\frac{1}{3} = \frac{5}{18}$ for the other two players

Since a proposal is accepted if and only if at least two players get at least their continuation value of the game we conclude that if $x^{ii} \geq \frac{4}{9}$ and $x^{ij} \geq \frac{5}{18}$ for some $j \neq i$, then x^i will be accepted if proposed in first session given the continuation strategies above. The best deviation for the proposer gives a payoff of $\frac{5}{9}$, so if $x^{ii} \geq \frac{5}{9}$ and if the proposal is accepted in equilibrium, then there is no profitable deviation from x^i for the proposer in the first period. Hence we can for example construct an equilibrium where the target outcome is $x^1 = \left(\frac{5}{9}, \frac{5}{18}, \frac{3}{18}\right)$if player 1 is recognized,

[11]To be complete we should also include the provision that the proposer offered at least $\frac{2}{9}$ to herself, since otherwise she would have a strict incentive to reject her own proposal. The conditions below assumes that the proposer will vote to accept.

$x^2 = \left(\frac{3}{18}, \frac{5}{9}, \frac{5}{18}\right)$ if player 2 is recognized and $x^3 = \left(\frac{5}{18}, \frac{5}{18}, \frac{5}{9}\right)$ if player 3 is recognized in session 1[12].

The model as specified in this paper assumes that the utility functions are linear in the share of benefits. In this case an equilibrium is efficient if and only if the players agree in the first period (this is of course only relevant with $\delta < 1$) and no resources are thrown away. However, in many applications it is natural to assume that there is some curvature in the utility functions, reflecting attitudes towards risk or simply that the marginal benefits of additional spending in a district may be decreasing. In this case the type of strategies described in Example 2 actually serves a purpose: it will now be possible to increase the ex ante expected utility for all players by threatening to punish proposers for making "too unfair" proposals in the initial session.

It is not surprising that the set of divisions supportable as equilibrium outcomes increases as the number of sessions increases. However, if the players are not patient enough or if there are too few (i.e. 3) players, then it is not possible to support an arbitrary distribution as a subgame perfect equilibrium. The general result is:

Proposition 2. *Suppose there are an odd number of players $n \geq 3$. Then, for any distribution of benefits x such that $x_i > 0$ for all $i \in I$ there is a $T \in N$ and some $\delta^* < 1$ such that x is supportable as a subgame perfect Nash equilibrium in the model if $\delta \geq \delta^*$ and if there are at least T rounds of bargaining.*

The proof is relegated to the appendix in Section 1.6.2. The idea is to construct strategies such that, [1] in the voting stage, voters are punished if an equilibrium proposal is rejected, [2] in the voting stage, the "right group" of voters is rewarded if a deviant proposal is rejected, [3] the proposer is punished if he proposes something else than the equilibrium proposal. The main complication in the argument is that two different constructions have to be used. The equilibrium strategies constructed are similar to the strategies in the second example towards the end of the game. In sessions when there are not so many rounds left to go the strategies works roughly the following way: A proposer that deviates from her "target proposal" will be excluded from the winning majority in the next session. However, if a proposal is accepted the game ends, so discipline must be induced for the voters as well. To do this (the first $(n+1)/2$) voters that reject deviant proposals are rewarded by inclusion in the winning majority in the next session. It is

[12] If we allow the players to throw away resources we can also support inefficient equilibria. For example it is possible to support $x^1 = \left(\frac{5}{9}, \frac{5}{18}, 0\right), x^2 = \left(0, \frac{5}{9}, \frac{5}{18}\right), x^3 = \left(0, \frac{5}{18}, \frac{5}{9}\right)$.

shown that these strategies makes it possible to induce the proposer to propose less for herself as the number of sessions remaining is increased. Also, individual voters can be induced to accept proposals that give them a smaller share, since they will otherwise be punished by exclusion from the winning majority.

While the set of distributions supportable by these strategies is increasing in the number of sessions remaining, they cannot support the whole simplex. But, if there are sufficiently many periods continuation strategies of the type discussed above can be used to support equilibria with sufficient flexibility in continuation valuations so that it is possible to "match" deviant proposals in the beginning of the game. This matching works roughly as follows: for any deviant proposal, let the equilibrium proposal for all players in the next period be such that the $(n+1)/2$ players that get the least under the deviation get something that in present value terms is at least as great as under the deviation (and strictly more if recognized to propose). A player who deviates when proposing is always excluded, unless recognized again in the next session. Working backward we show that as the number of bargaining sessions go to infinity, the share the proposer gives to herself goes to zero in this equilibrium. Using the specific equilibrium we construct, we can show that irrespective how we pick an initial proposal (in the interior of the simplex), any player selected to make a proposal in the first period will follow the recommendation and the proposal will be accepted by a majority.

Note that for large legislatures, the condition that $\delta > \overline{\delta}$ is not a restriction since $\overline{\delta} \rightarrow 0$ as $n \rightarrow \infty$. The reason for the requirement that $x_i > 0$ for all players is that for any finite horizon there is a probability for the same player to be recognized in every session. Hence the proposer must get something strictly possible in order not to have incentives to deviate.

1.3.1. The Role of Sequential Voting

As is well known, simultaneous move voting games typically yield multiple equilibria. For example, if 3 players all prefer alternative A to alternative B it is still a Nash equilibrium for everyone to vote for B since no player can affect the outcome by a unilateral deviation. To avoid these unnatural equilibria it is standard to restrict attention to weakly undominated Nash equilibria in voting models.

These considerations are irrelevant when voting is performed sequentially. Consider the example above, but suppose that player 1 votes first. Then, after observing how player 1 voted, player 2 votes and finally player 3 votes after observing the actions of the two other players. Now, if player

3 is pivotal he will vote for A, his most preferred alternative. Foreseeing the optimal response by player 3, player 2 will vote for A if player 1 voted for B and is indifferent if player one voted for A. Finally player 1 can also vote for anything since no matter what he does there will always be a majority for A, his most preferred outcome.

The logic of this example generalizes so that the set of subgame perfect equilibrium outcomes[13] of the game cannot be further refined by iterative elimination of conditionally weakly dominated strategies. In fact, if conditionally weakly dominated voting strategies are eliminated in the model with simultaneous voting, then the set of equilibrium outcomes is the same as for the model with sequential voting.

1.4. Heterogeneous Time Preferences

In the examples and in the proof of Proposition 2 we exploited the fact that the set of optimal proposals in the second period from the end is to give $\frac{\delta}{n}$ to any group of $\frac{n-1}{2}$ other players and keep the rest. Since the proposer is indifferent between a proposal where a particular set of players are included in the set of "winners" and a strategy where say player i is excluded and player j is included the proposer can punish and reward players without costs

Assume instead that the proposer in the penultimate period is not indifferent over whom to select to receive positive benefits. This could be due to differences in the probabilities of recognition, differences in the discount factors or payoff externalities between nearby districts. Then there is a unique equilibrium proposal in the penultimate stage and weighting by the respective probabilities of recognition we can compute the unique equilibrium continuation values in the beginning of the second period from the end. Now if all these are distinct there is a unique optimal proposal in the next period (give the discounted continuation payoff to the $(n-1)/2$ players with the lowest continuation payoff). Proceeding backwards, as long as the continuation values of the game are distinct in each stage backwards induction produces a unique solution.

Consider the case with heterogenous time preferences and, say, three players with $\delta_1 > \delta_2 > \delta_3$. Then if players 1 or 2 is to propose in the second stage from the end they will propose to give $\frac{\delta_3}{3}$ to player 3 since player 3 will be the cheapest player to bribe. For the same reason player 3 would select player 2. Now the unique equilibrium continuation values can be computed as $V_1(2) = \frac{1}{3}\left(1 - \frac{\delta_3}{3}\right)$, $V_2(2) = \frac{1}{3}\left(1 - \frac{\delta_3}{3}\right) + \frac{1}{3}\frac{\delta_2}{3}$ and $V_3(2) = \frac{1}{3}\left(1 - \frac{\delta_2}{3}\right) + \frac{2}{3}\frac{\delta_3}{3}$. So the question

[13] Through the paper the term "outcome" refers to a physical allocation.

whether the proposals at the third stage from the end are unique or not is a question whether $\delta_1 V_1(2) \neq \delta_2 V_2(2) \neq \delta_3 V_3(2)$ or not. Intuitively, if we would select $(\delta_1, \delta_2, \delta_3)$ randomly we would be quite surprised if we found that we had equality. Actually it is not hard to see that for almost all choices of $(\delta_1, \delta_2, \delta_3)$ we have that $\delta_1 V_1(2) \neq \delta_2 V_2(2) \neq \delta_3 V_3(2)$.

It is natural to conjecture that this logic can be extended beyond the three person case and the three period case. This conjecture turns out to be right and the purpose of this section is to show this. The main result is that, for any choice of discount factors for the players from a subset of $[0,1]^n$ with full Lebesgue measure, there is a unique subgame perfect equilibrium. Hence the results for the basic model of Section 1.2 are only interesting if we have any reasons to believe that the probabilities of recognition and discount factors are exactly equal[14]. If we believe that there are slight differences, the generic uniqueness result implies that the symmetric model is not a good approximation of the "right" model.

Before proceeding to the formal analysis, note that it does not matter whether the asymmetry is in time preferences or in probabilities of recognition. Actually, if the probabilities of recognition are allowed to be different across periods, then a simpler argument than the argument below can be made to establish generic uniqueness. However, it is easier to interpret the case with differences in discount factors or time invariant differences in the probability of recognition, so we stick to this case.

Let $\delta = (\delta_1, ..., \delta_n) \in [0,1]^n$ be the vector of discount factors for the players. The *δ-perturbation* of the symmetric game G^T will be denoted $G^T(\delta)$ and is identical to the basic model described in Section 1. 2 with the exception that the discount factors need not be identical. The set of subgame perfect equilibria of $G^T(\delta)$ is denoted $SPE(T, \delta)$. The (ex ante) continuation value of the game for player $i \in I$ given strategy profile σ and history $h_t \in H_t$ will be denoted $V_i(h_t \mid \sigma_{|t})$, where $\sigma_{|t}$ denotes the truncation of σ starting at time t. Note here that in Section 1. 2 we reserved the notation H_t for the set of possible histories at the beginning of session t[15]and that $V_i(h_t \mid \sigma_{|t})$ is evaluated before it is known who is to make a proposal in the session. If for all $\sigma, \widetilde{\sigma} \in SPE(T, \delta)$ and all $h_t, \widetilde{h}_t \in H_t$ it is the case that $V_i(h_t \mid \sigma_{|t}) = V_i(\widetilde{h}_t \mid \widetilde{\sigma}_{|t})$ we will denote the unique equilibrium continuation value for i at time t as $V_i(t)$.

Lemma 1. *Suppose that* $V_i(t) = V_i(h_t \mid \sigma_{|t}) = V_i(\widetilde{h}_t \mid \widetilde{\sigma}_{|t})$ *for all* $\sigma, \widetilde{\sigma} \in SPE(T, \delta)$ *and all*

[14]Actually, as long as the product of the two is equal we get a continuum of equilibria.

[15]The set of possible histories at nodes where a proposal is currently voted on are denoted H_t^j, where j indexes how many players in the predetermined voting order have already cast their votes.

$h_t, \widetilde{h}_t \in H_t$ *and that there is a player* $m \in I$ *such that the sets* $A = \{i \in I \mid \delta_i V_i(t) > \delta_m V_m(t)\}$ *and* $B = \{i \in I \mid \delta_i V_i(t) < \delta_m V_m(t)\}$ *both have cardinality* $\frac{n-1}{2}$. *Then there is a unique equilibrium proposal at time* $t+1$ *for any player* $i \in I$, x_{t+1}^{*i}.

The proof is in the appendix in Section 1.6.2, but the idea is simple. All that the Lemma says is that, under the hypothesis of the Lemma, any player must select $(n-1)/2$ players with the highest indices to receive their discounted continuation valuations. Clearly, any other proposal would be suboptimal, since then the proposer would have to give away a larger share to other players.

Lemma 2. *Suppose the assumptions made in Lemma 1 holds. Then for each* $i \in I$ *there exists a unique value* $V_i(t+1)$ *such that* $V_i(t+1) = V_i(h_{t+1} \mid \sigma_{|t+1}) = V_i\left(\widetilde{h}_{t+1} \mid \widetilde{\sigma}_{|t+1}\right)$ *for all* $\sigma, \widetilde{\sigma} \in SPE(T, \delta)$ *and all* $h_{t+1}, \widetilde{h}_{t+1} \in H_t$.

Proof. By Lemma 1 there is a unique equilibrium proposal at time $t+1$ for any $i \in I$. This proposal is accepted for sure and the continuation value at $t+1$ for a particular player is simply the sum over all players of the share the player gets if j is recognized multiplied by the probability that j is recognized. Since the proposals are unique the numbers computed in this way are obviously unique for each player. ■

Note that these continuation values can be written as

$$V_i(t+1) = \begin{cases} \frac{1}{n}\left(1 - \sum\limits_{j=\frac{n+3}{2}}^{n} \delta_j V_j(t)\right) & \text{for } i < \frac{n+1}{2} \\ \frac{1}{n}\left(1 - \sum\limits_{j=\frac{n+3}{2}}^{n} \delta_j V_j(t)\right) + \frac{(n-1)\delta_i V_i(t)}{2n} & \text{for } i = \frac{n+1}{2} \\ \frac{1}{n}\left(1 + \delta_i V_i(t) - \sum\limits_{j=\frac{n+1}{2}}^{n} \delta_j V_j(t)\right) + \frac{(n-1)\delta_i V_i(t)}{n} & \text{for } i > \frac{n+1}{2} \end{cases} \tag{1.1}$$

Using Lemma 1 we now proceed to prove that a generic perturbation of the original game $G^T(\delta)$ has a unique equilibrium outcome. To be precise, any two equilibria differ only with respect to the voting strategies given a particular proposal at a particular time and these differences do not affect the outcome of the voting over this particular proposal.

Proposition 3. *For any finite* T *there exists a set* $D^T \subset [0,1]^n$ *with full Lebesgue measure such that if* $\delta \in D^T$ *and if* $\sigma, \widetilde{\sigma} \in SPE(T, \delta)$ *the following holds true*

1. $V_i(t) = V_i(h_t \mid \sigma_{|t}) = V_i(\tilde{h}_t \mid \tilde{\sigma}_{|t})$ *for all* $\sigma, \tilde{\sigma} \in SPE(T,\delta)$ *all* $h_t, \tilde{h}_t \in H_t$ *all* $i \in I$ *and all* $t \leq T$*. Furthermore* $V_i(t) > 0$ *for all* $i \in I$ *and all* t[16].

2. *For any* $h_t, \tilde{h}_t \in H_t$ *, any* $i \in I$ *and all* $t \leq T$ *:*

   a *The local strategy at a node where* i *makes a proposal is pure.*

   b *The proposal made at time* t *are history-independent . I.e. (abusing notation) for any* $\sigma \in SPE(T,\delta)$ $\sigma_t^i : H_t \to \Delta^n$ *satisfies* $\sigma_t^i(h_t) = \sigma_t^i(\tilde{h})$ *for all* $h_t, \tilde{h}_t \in H_t$*.*

   c *All equilibrium strategies and all possible histories induces the same proposal if any particular player* i *is recognized at time* t*. I.e.* $\sigma_t^i(h_t) = \tilde{\sigma}_t^i(\tilde{h})$ *for all* $h_t, \tilde{h}_t \in H_t$*.*

   d *All proposals on the equilibrium path are accepted for sure.*

Hence the equilibrium outcome[17] *is the same for all* $\sigma, \tilde{\sigma} \in SPE(T,\delta)$

Proof. The proof is by induction. Before getting into the details of the proof some words about the logic might be helpful. We assume that when there are t periods left to go there is a unique equilibrium proposal for any player i for all $\delta \in [0,1]^n \setminus X_t$ where $\mu(X_t) = 0$ and μ denotes the Lebesgue measure defined over $[0,1]^n$. We show that the set $N_{t+1} \subset [0,1]^n$ such that there is not a unique equilibrium proposal at time $t+1$ satisfies $\mu(N_{t+1}) = 0$. Since the union of two sets of measure zero has measure zero it follows that there is a unique equilibrium proposal for all $\delta \in [0,1]^n \setminus X_{t+1}$ where $X_{t+1} = X_t \cup N_{t+1} \Rightarrow \mu(X_{t+1}) = 0$.

The Last stage

In the last stage we argue as in the case with a common discount factor to show that, for any $\delta \in [0,1]^n$, the unique equilibrium outcome is for the proposer to suggest to take everything and this being accepted. Hence the continuation values in the beginning of the last session are uniquely determined, $V_i(1) = V_i(h_1 \mid \sigma_{|1}) = V_i(\tilde{h}_1 \mid \tilde{\sigma}_{|1}) = \frac{1}{n}$ for all $\sigma, \tilde{\sigma} \in SPE(T,\delta)$ all $h_1, \tilde{h}_1 \in H_1$, so the statement of Proposition 3 holds for $t = 1$.

The Induction Step

[16]This is actually true in any subgame perfect equilibrium of the finite game. The intuitive reason is that for all players there is a probability that the player is recognized in every session.

[17]Remember that we refer distributions of benefits as outcomes.

Suppose that the statement holds for the model with t rounds of bargaining. By Lemma 1 there will then be a unique equilibrium proposal at $t+1$ if

$$\delta_{\frac{n-1}{2}} V_{\frac{n-1}{2}}(t) > \delta_{\frac{n+1}{2}} V_{\frac{n+1}{2}}(t) > \delta_{\frac{n+3}{2}} V_{\frac{n+3}{2}}(t),$$

where we have relabeled the players so that $\delta_1 V_1(t) \geq \delta_2 V_2(t) \geq ... \geq \delta_n V_n(t)$. Our task is to show that for almost all $\delta \in [0,1]^n$ the two strict inequalities above is satisfied. Clearly, a sufficient condition for the above inequality to hold is if $\delta_i V_i(t) \neq \delta_j V_j(t)$ for *all* i, j with $i \neq j$ (For $t = 1$ we have $V_i(t) = V_j(t) = \frac{1}{n}$. Since $\{\delta \in [0,1]^n \mid \delta_i = \delta_j \text{ for some pair } (i,j)\}$ has measure zero, this shows that the proposition holds for the case with two rounds of bargaining). For an arbitrary $t > 2$ we need to introduce some notation. It is convenient to first transform (1.1) to matrix notation. We can express (1.1) as

$$x(t) = A_t x(t-1) \tag{1.2}$$

where $x(t)' = (\delta_1 V_1(t), ..., \delta_n V_n(t))$ and A_t is given by

$$\underset{n\times n}{A_t} = \begin{pmatrix} \Lambda_t & B_t \\ \mathbf{0} & C_t \end{pmatrix} \tag{1.3}$$

where Λ_t is of order$\left(\frac{n-1}{2} \times \frac{n-1}{2}\right)$, B_t is of order $\left(\frac{n-1}{2}, \frac{n+1}{2}\right)$, $\mathbf{0}$ is of order $\left(\frac{n+1}{2}, \frac{n-1}{2}\right)$ and C_t is of order $\left(\frac{n+1}{2} \times \frac{n+1}{2}\right)$. The explicit formulas for the sub-matrices are given by

$$\Lambda_t = \begin{pmatrix} \frac{1}{nV_1(t-1)} & 0 & \cdots & 0 \\ 0 & \frac{1}{nV_2(t-1)} & 0 & \cdots \\ \cdots & 0 & \cdots & 0 \\ 0 & \cdots & 0 & \frac{1}{nV_{\frac{n-1}{2}}(t-1)} \end{pmatrix}, \quad B_t = \begin{pmatrix} 0 & -\frac{\delta_1}{n} & \cdots & -\frac{\delta_1}{n} \\ 0 & -\frac{\delta_2}{n} & & -\frac{\delta_2}{n} \\ \cdots & \cdots & & \cdots \\ 0 & -\frac{\delta_{\frac{n-1}{2}}}{n} & \cdots & -\frac{\delta_{\frac{n-1}{2}}}{n} \end{pmatrix} \tag{1.4}$$

$$C_t = \begin{pmatrix} \frac{1}{nV_{\frac{n+1}{2}}(t-1)} + \frac{\delta_{\frac{n+1}{2}}(n-1)}{2n} & -\frac{\delta_{\frac{n+1}{2}}}{n} & \cdots & \cdots & -\frac{\delta_{\frac{n+1}{2}}}{n} \\ -\frac{\delta_{\frac{n+3}{2}}}{n} & \frac{1}{nV_{\frac{n+3}{2}}(t-1)} + \delta_{\frac{n+3}{2}} & -\frac{\delta_{\frac{n+3}{2}}}{n} & \cdots & -\frac{\delta_{\frac{n+3}{2}}}{n} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ & & \cdots & \cdots & -\frac{\delta_{n-1}}{n} \\ -\frac{\delta_n}{n} & -\frac{\delta_n}{n} & \cdots & -\frac{\delta_n}{n} & \frac{1}{nV_n(t-1)} + \delta_{\frac{n+3}{2}}\delta_n \end{pmatrix} \tag{1.5}$$

To get these expressions we multiply both sides of the expression for $V_i(t)$ by δ_i and to rearrange

Lemma 3. *A_t has full rank*

A direct proof of Lemma 3 is found in the appendix in Section 1.6.3 . It is also possible to show that A_t is a dominant diagonal matrix, which implies that it must have full rank. Since the date t is arbitrary, the Lemma holds for all dates $t-1, ..., 3$ (by pre-multiplication by appropriately chosen (full rank) permutation matrix in each step we can in principle keep track of the agents). Hence, $x(t) = Q_t \frac{\delta}{n}$, where Q_t is some permutation of $A_t A_{t-1} ... A_3$. Since $A_3, ... A_t$ all have full rank it follows that Q_t has full rank.

Let q_i be row i in Q_t and suppose that δ^* is critical for period $t+1$. By Lemma 1 it follows that $\delta_i^* V_i(t) = \delta_j^* V_j(t)$ for some i, j with $i \neq j$. Thus:

$$\delta_i^* V_i(t) = \delta_j^* V_j(t) \Leftrightarrow x_i(t) = x_j(t) \Leftrightarrow q_i \delta^* = q_j \delta^* \Leftrightarrow (q_i - q_j)\delta^* = 0$$

Therefore the set of critical values at time $t+1$, is given by

$$N_{t+1} = \{\delta \in [0,1]^n \mid (q_i - q_j)\delta^* = 0 \text{ for some pair } q_i, q_j \text{ i.e. some pair of rows of } Q_t\}$$

But Q_t has full rank, so each pair q_i, q_j are linearly independent, so $(q_i - q_j)$has some strictly positive element. Therefore N_{t+1} is the intersection between $[0,1]^n$ and a finite set of hyperplanes in R^n. It follows that $\mu(N_{t+1}) = 0$. Since $X_t = \{\delta \in [0,1]^n \mid \delta \text{ critical for some } \tau \leq t\}$ has measure zero by the induction hypothesis it follows that

$$X_{t+1} = \{\delta \in [0,1]^n \mid \delta \text{ critical for some } \tau \leq t+1\} = \cup_{t'=1}^{t} N_{t+1}$$

has measure zero. By induction, X_T has measure zero for any T, which proves the result. ■

1.5. Concluding Remarks

We have considered the finite horizon version of a legislative bargaining model due to Baron and Ferejohn[6]. We showed that there is a continuum of divisions supportable as an equilibrium outcome for any (common) discount factor as long as there are three or more rounds of bargaining. Given that the players are sufficiently patient we showed that for any division of benefits such that all players receive a strictly positive share, there is a T large enough, so that if there are at least T rounds of bargaining, the proposed division can be supported as an equilibrium outcome.

For the specification of the model considered in this paper, there are no particular reasons why we would expect the players to coordinate on the rather complicated equilibrium strategies used in the examples and in the proof of the "limit folk theorem". However, in most interesting applications of the model there is some curvature of the players' utility function. While I don't have a proof that covers this more general case, it seems that the logic of the arguments have nothing to do with the linear payoffs. Thus, I conjecture that the indeterminacy results, if anything, only would be made stronger by assuming concave payoff functions. Moreover, in this case there would be an efficiency rationale for coordination on an equilibrium where all players receive positive shares with probability 1.

In Section 1. 4 the model is slightly generalized to allow for differences in the discount rate across players. One can think of this as reduced form for different probabilities of reelection or effects of term limits etc. It should however be noted that the

model with asymmetric payoff functions considered in Section 1. 4 we showed that, for generic choices of discount factors, there is a unique subgame perfect equilibrium outcome. In this generically unique equilibrium the continuation strategies induces a unique distribution over the physical outcomes after any history of play. These distributions will typically be varying over time, but not history dependent.

The equilibrium outcomes of the non-symmetric model will have one important qualitative property in common with the stationary equilibrium of the infinite model: only a majority of the barest possible size will receive strictly positive benefits. Thus, Rikers' [25] minimal winning coalition theorem holds generically.

If we take limit as the number of sessions goes to infinity and consider the sequence of equilibrium proposals, then we note that these sequences can not converge unless the discounted value of the game is the same for all players when there is some finite number of sessions left. But, if the discounted value of the game is the same for all players, then we don't have uniqueness. Hence, when there is a unique equilibrium it can not approach the stationary equilibrium when the discount factors converges towards a common discount factor and the number of sessions goes to infinity.

1.6. Appendix

1.6.1. Proof of Proposition 2

Proof. Time is counted backwards and we assume that n is odd. We let $\{m_t\}_{t=1}^{T}$ be a sequence of real numbers (to be specified later) with $m_t \in [0,1]$ for each t. A strategy profile will be constructed where each player $i \in I$ follows symmetric proposal strategies $p^i : H^t \to X$ when called to make a proposal. Suppose that $i' \in I$ was called to make a proposal at time $t+1$ (i.e. the round before round t), proposed $p^{i'}(h_{t+1})$ and also voted in favor of this proposal, but that this proposal was rejected by a majority. Furthermore, let the proposer at t be indexed by i. After all histories h_t with these properties we let $p^i(h_t) = p^{i1}(h_t), \ldots, p^{in}(h_t)$[18] be given by

$$p^{ij}(h_t) = \begin{cases} m_t & \text{for } j = i \\ 0 & \text{for } \frac{n-1}{2} \text{ randomly select agents who voted to reject } p(h_{t+1}) \\ \frac{2(1-m_t)}{n-1} & \text{otherwise} \end{cases} \quad (1.6)$$

If on the other hand the proposer himself voted to reject the proposal and a different agent is selected to make a proposal at time t, then

$$p^{ij}(h_t) = \begin{cases} m_t & \text{for } j = i \\ 0 & \text{for the proposer at } t+1 \text{ and } \frac{n-3}{2} \text{ randomly select rejectors} \\ \frac{2(1-m_t)}{n-1} & \text{otherwise} \end{cases} \quad (1.7)$$

Note in particular that (regardless of whether the proposer at time $t+1$ voted to reject or not) all agents who voted in favor of $p^{i'}(h_{t+1})$ get a positive share with probability one. Also, since $\frac{n-1}{2}$ players receive $\frac{2(1-m_t)}{n-1}$ we have that $\sum_j p^j(h_t) = m_t + \frac{n-1}{2}\frac{2(1-m_t)}{n-1} = 1$, so the proposal is always feasible for any $m_t \in [0,1]$. For histories where in the last period agent i' was selected to propose and proposed $x_{t+1} \neq p^{i'}(h_{t+1})$ and the current proposer i was among the first $\frac{n+1}{2}$ players who voted to reject[19], then

$$p^{ij}(h_t) = \begin{cases} m_t & \text{for } j = i \\ \frac{2(1-m_t)}{n-1} & \text{for } j \neq i' \text{ among first } \frac{n+1}{2} \text{ players who voted to reject } x_{t+1} \\ 0 & \text{otherwise} \end{cases} \quad (1.8)$$

If the current proposer was not among the $\frac{n+1}{2}$ first agents who voted to reject one of the agents in this group is drawn at random to receive nothing, i.e.

$$p^{ij}(h_t) = \begin{cases} m_t & \text{for } j = i \\ \frac{2(1-m_t)}{n-1} & \frac{n-1}{2} \text{ randomly select players from } \frac{n+1}{2} \text{ first rejectors} \\ 0 & \end{cases} \quad (1.9)$$

Observe that p^i specifies what proposal to make after an arbitrary history no matter how the sequence m_t is specified. Our goal is to select the sequence m_t so that p^i is part of a subgame perfect equilibrium, that is, it is always sequentially rational for the current proposer to make proposals in accordance to p^i. Hence, we must see under what conditions deviant proposals are rejected and under what conditions $p^i(h_t)$ is accepted.

[18]The first superscript refers to the identity of the proposer and the second to the receiver.

[19]Note that it is not assumed that $i' \neq i$. In fact, with equal probabilities of recognition, the probability that $i' = i$ is $1/n$.

Assume that the strategy profile s is such that at time t each player $i \in I$ follows proposal rule $p^i(h_t)$ and that this proposal is accepted. Let j be an agent who was among first $\frac{n+1}{2}$ players (other than the proposer) who voted to reject a deviant proposal at time $t+1$. Now we can compute agent j^s ex ante continuation payoff at time t (before identity of proposer is known) as

$$V^j\left(h_t, s_{|h_t}\right) = \frac{1}{n}m_t + \frac{n-1}{n}\left(\frac{1}{2}+\frac{1}{2}\left(1-\frac{2}{n+1}\right)\right)\frac{2(1-m_t)}{n-1} \equiv V_t(x,0) \quad (1.10)$$

The interpretation is as follows: with probability $\frac{1}{n}$ the agent is called to make a proposal which gives him m_t. with probability $\frac{n-1}{n}$ the agent is not called to make a proposal. In this case he will get $\frac{2(1-m_t)}{n-1}$ for sure if someone in the group of $\frac{n+1}{2}$ rejectors is proposing (which happens with probability $\frac{1}{2}$ conditional on that someone else is proposing), while if the proposer is not from this group (conditional probability $\frac{1}{2}$) he may be the unlucky agent who does not get anything (with conditional probability $\frac{2}{n+1}$). Simplifying we get that the unconditional probability that the agent will receive $\frac{2(1-m_t)}{n-1}$ is $\frac{n-1}{n+1}$, so

$$V_t(x,0) = \frac{1}{n}m_t + \frac{n-1}{n+1}\frac{2(1-m_t)}{n-1} \quad (1.11)$$

Still assuming all agents follow p if called to propose and that any proposal in accordance to p is accepted at time t we can compute the ex ante continuation payoff for any agent j who was not among the first $\frac{n+1}{2}$ to reject. Since these agents will receive zero benefits unless called to propose we have that $V^j\left(h_t, s_{|h_t}\right) = \frac{1}{n}m_t = V_t(x,1)$. The continuation payoff for the proposer of the previous period is $V^i\left(h_t, s_{|h_t}\right) = \frac{1}{n}m_t$, irrespective of whether the proposer voted to accept or reject his own proposal. Now consider the case where the proposal rejected in the last period was in accordance to p. Then the ex ante continuation payoff for an agent j who voted to accept the proposal is given by

$$V^j\left(h_t, s_{|h_t}\right) = \frac{1}{n}m_t + \frac{n-1}{n}\frac{2(1-m_t)}{n-1} \equiv V_t(p,1), \quad (1.12)$$

while if j is in the set of $r \geq \frac{n+1}{2}$ agents who voted to reject $p(h_{t+1})$ (and if the proposer is not among the agents who voted to reject) we have that

$$\begin{aligned} V^j\left(h_t, s_{|h_t}\right) &= \frac{m_t}{n} + \left(\frac{r-1}{n}\left(\frac{2r-n-1}{2(r-1)}\right) + \frac{n-r}{n}\left(\frac{2r-n+1}{2r}\right)\right)\frac{2(1-m_t)}{n-1} \\ &= V_t(p,r,0). \end{aligned} \quad (1.13)$$

A similar formula can be derived in the case when the proposer is among the agents who voted to reject, but, as we will see, the proposer will always be better off by deviating from the proposal rule than from following the proposal rule and then vote to reject.

Note that if $p(h_{t+1})$ was rejected by a minimal winning coalition, i.e. $r = (n+1)/2$, then $(2r-n-1)/2(r-1) = 0$ and $(2r-n+1)/2r = 2/(n+1)$: in this case agent j will only receive something with positive probability when someone outside the rejecting majority from last period is called to propose (and only one agent is needed from this group to get a majority together with the group of acceptors from last period so the probability of getting a positive share conditional on that someone outside the rejecting majority is recognized is $2/(n+1)$).

A voting rule is a mapping $d_t^i : H_t^{k_t(i)} \to \{1,0\}$ and we could in principle work out exactly what these must be if in the next period any player i selected to make a proposal at $t-1$ will do this in accordance to p^i and if any such proposal is accepted. However, given the particular proposal rules we consider, the incentives for the proposer is only affected by whether proposal is accepted or not and the proposers own voting behavior. For this reason it will be sufficient to know whether the outcome of the voting stage is to accept or not, which simplifies the analysis a great deal.

Lemma 4. *Let $i' \in I$ denote the proposer at time t and let $V_{t-1}^{i'}(x,0) = \frac{m_{t-1}}{n}$ and $V_{t-1}^{i}(x,0) = V_{t-1}(x,0)$ for all $i \neq i'$. Suppose that all agents $i \in I$ follow the proposal rule p^i at time $t-1$ and that the voting rules at time $t-1$ are such that $p^i(h_{t-1})$ is always accepted for any h_{t-1} and any $i \in I$. Then if the current proposer i' proposes some allocation $x_t \neq p^{i'}(h_t)$ we have that:*

i If $x_t^j > \delta V_{t-1}^j(x,0)$ for at least $\frac{n+1}{2}$ agents $j \in I$ there exists no sequentially rational voting rules such that x_t is rejected.

ii If $x_t^j < \delta V_{t-1}^j(x,0)$ for at least $\frac{n+1}{2}$ agents $j \in I$ there exists no sequentially rational voting rules such that x_t is accepted.

iii If $x_t^j \geq \delta V_{t-1}^j(x,0)$ for at least $\frac{n+1}{2}$ agents $j \in I$ there exists some sequentially rational voting rules such that x_t is accepted.

Proof. i) Suppose $x_t^j > V_{t-1}^j(x,0)$ for all $j \in J$, where $|J| = \frac{n+1}{2}$. Without loss of generality relabel the players so that $j=1$ is the first agent in J who votes at round t, 2 is the second agent in J to vote,....,and $j = (n+1)/2$ is the last agent in J to vote. First consider the decision of agent $(n+1)/2$ after an arbitrary sequence of votes. Since agent $(n+1)/2$ is not necessarily the last agent to vote he must in general take the continuation strategies of the other players into consideration But if $(n-1)/2$ votes have already been cast in favor of the proposal and if the continuation strategies are such that there is a positive probability that the proposal is rejected if $(n+1)/2$ votes to reject there is a strict incentive for the agent to vote in favor of proposal. Hence the proposal must be accepted with probability 1 if when the last agent in J is called to vote, $(n-1)/2$ votes have already been cast in favor of the proposal. Now consider $j < (n+1)/2$. Suppose that the continuation strategies are such that:

1. x_t accepted for sure if j votes in favor of proposal.

2. x_t rejected with positive probability if j votes to reject.

For the same reason as above j has a strict incentive to vote in favor of x_t. Now suppose that x_t would be rejected. Then there must be some agent $j \in J$ who votes to reject[20], since otherwise there is a majority for x_t. Let j' be the first agent who votes to reject. Since j' strictly prefers to accept the proposal there must be some $j'' \in J$ with $j'' > j'$ (that is, voting after j') such that j'' would reject even if j' voted to accept. But j'' also strictly prefers the proposal so there must be $j''' \in J$ with $j''' > j''$ such that j''' would reject even if both j' and j'' voted to accept and continuing by induction we find that $j = (n+1)/2$ must reject even if all other $j \in J$ has voted to accept. But since agent $(n+1)/2$ strictly prefers x_t this is a contradiction.

ii) Let J be a set of $(n+1)/2$ agents such that the discounted value of being among the first $(n+1)/2$ rejectors is strictly higher than the payoff from the current proposal and argue as in i).

iii) Proceed by backwards induction. If at least $(n-1)/2$ votes have already been cast in favor of the proposal it is a best response for $j = (n+1)/2$ to vote in favor of proposal (in fact, when exactly $(n-1)/2$ votes have been cast in favor of x_t it is a conditionally weakly dominated action to accept). If $j' = (n-1)/2$ assumes that the last player in j will always accept after any sequence of votes where at least $(n-1)/2$ votes are in favor, it is a best response for j' to vote in favor of x_t after any history where at least $(n-3)/2$ votes are already in favor of x_t. The result follows by (backwards) induction.. ■

Lemma 5. *Suppose that all agents $i \in I$ follow the proposal rule p^i at time $t-1$ and that the voting rules at time $t-1$ are such that $p^i(h_{t-1})$ is always accepted for any h_{t-1} and any $i \in I$. Furthermore, suppose that $m_{t-1} < 1$. Then there exists no sequentially rational voting rules and no history h_t such that $p^i(h_t)$ is rejected by more that $(n+1)/2$ players.*

[20]In the rest of this argument, "reject" means reject with positive probability.

Proof. Suppose $(n+1)/2$ players have already voted to reject $p^i(h_t)$ when it is agent j^s turn to vote. If j votes in favor of proposal, her discounted expected payoff is

$$\delta V_{t-1}(p,1) = \delta\left(\frac{1}{n}m_{t-1} + \frac{n-1}{n}\frac{2(1-m_{t-1})}{n-1}\right), \tag{1.14}$$

while if j votes to reject her discounted expected payoff is

$$\delta V_{t-1}(p,0,r) = \delta\left(\frac{m_{t-1}}{n} + \left(\frac{r-1}{n}\left(\frac{2r-n-1}{2(r-1)}\right) + \frac{n-r}{n}\left(\frac{2r-n+1}{2r}\right)\right)\frac{2(1-m_{t-1})}{n-1}\right) \tag{1.15}$$

where $r \geq (n+1)/2$. To get the result we need to verify that $V_{t-1}(p,1) > V_{t-1}(p,0,r)$. Assume that this it not the case and that $m_{t-1} < 1$. Then

$$n-1 \leq (r-1)\left(\frac{2r-n-1}{2(r-1)}\right) + (n-r)\left(\frac{2r-n+1}{2r}\right).$$

It is easily verified that $\frac{2r-n-1}{2(r-1)} < 1$ and $\frac{2r-n+1}{2r} < 1$, which implies that the right hand side of the expression is strictly less than $n-1$, which is a contradiction. Thus $V_{t-1}(p,1) > V_{t-1}(p,0,r)$ and the result follows. ■

Lemma 6. *Suppose that all agents $i \in I$ follow the proposal rule p^i at time $t-1$ and that the voting rules at time $t-1$ are such that $p^i(h_{t-1})$ is always accepted for any h_{t-1} and any $i \in I$ (where it is assumed that $m_{t-1} < 1$). Then:*

i If $p^{ij}(h_t) > \delta V_{t-1}\left(p,0,\frac{n+1}{2}\right)$ for at least $\frac{n+1}{2}$ agents $j \in I$ there exists no sequentially rational voting rules such that x_t is rejected.

ii If $p^{ij}(h_t) < \delta V_{t-1}\left(p,0,\frac{n+1}{2}\right)$ for at least $\frac{n+1}{2}$ agents $j \in I$ there exists no sequentially rational voting rules such that x_t is accepted.

iii If $p^{ij}(h_t) \geq \delta V_{t-1}\left(p,0,\frac{n+1}{2}\right)$ for at least $\frac{n+1}{2}$ agents $j \in I$ there exists some sequentially rational voting rules such that x_t is accepted.

The proof is using essentially the same arguments as in the proof of Lemma 4 and is omitted.

Lemma 7. *Suppose that all agents $i \in I$ follow the proposal rule p^i at time $t-1$ and that the voting rules at time $t-1$ are such that $p^i(h_{t-1})$ is always accepted for any h_{t-1} and any $i \in I$. Furthermore, suppose that all agents' voting rules in round t are sequentially rational given the continuation strategies and that $p^i(h_t)$ is accepted with probability one by these sequentially rational voting rules. Then, if $m_t \geq \max\left\{1 - \frac{(n-1)}{2}\delta V_{t-1}(x,0), \frac{\delta m_{t-1}}{n}\right\}$ there is no profitable deviation from $p^i(h_t)$ for agent i when called to make a proposal at time t after history h_t.*

Proof. By Part iii of Lemma 4, a necessary condition for a proposal $x_t \neq p(h_t)$ to be accepted is that $x_t^j \geq \delta V_{t-1}^j(x,0)$ for at least $(n+1)/2$ agents $j \in I$. Assuming that the proposer votes in favor of his own proposal (and that everyone else accepts when indifferent) it follows that the best deviation for the proposer is to give $\delta V_{t-1}(x,0)$ to $(n-1)/2$ other players and keep $1-\frac{n-1}{2}\delta V_{t-1}(x,0)$ for himself. On the other hand, if agent i^s proposal is rejected by a majority the discounted expected value of the proposers payoffs is $\delta m_{t-1}/n$. Since m_t is what the proposer get if he sticks to the proposed strategy p^i if $p^i(h_t)$ is accepted ,by assumption, the voting rules are such that $p^i(h_t)$ is accepted when proposed there is consequently no profitable deviation if $m_t \geq \max\left\{1-\frac{(n-1)}{2}\delta V_{t-1}(x,0), \frac{\delta m_{t-1}}{n}\right\}$. ■

Lemma 8. *Let $m_1 = 1$ and $m_t = 1-\frac{\delta(n-1)}{2}\left(\frac{m_{t-1}}{n}+\frac{2(1-m_{t-1})}{n+1}\right)$ for $t = 2,3,...T$. Then there exists voting rules such that these voting rules together with the proposal rules $(p^1,...,p^n)$ constitutes a subgame perfect equilibrium, where after any history of play h_t any proposal in accordance to p^i is always accepted by a majority.*

Proof. The unique subgame perfect equilibrium outcome for the one-period game has the proposer suggesting that he takes everything and this proposal is accepted. With $m_1 = 1$ this is also what the proposal rule p^i specifies each player i do propose after any history h_1. Now consider some period $t > 1$ and suppose that the proposal rules $p^1,...,p^n$ (with m_t recursively defined above) together with some voting rules which are such that all proposals in accordance to p^i are accepted are consistent with Nash equilibrium play in every subgame from period $t-1$ on. We want to show that under these hypothesizes, there exists sequentially rational voting rules such that: i) any proposal in accordance to p^i is accepted by a majority and ii) that there is no deviant proposal that is accepted under these voting rules that makes the proposer strictly better off.

i) By Lemma 6, if $p^{ij}(h_t) \geq \delta V_{t-1}\left(p,0,\frac{n+1}{2}\right)$ for at least $\frac{n+1}{2}$ agents $j \in I$ there exists some sequentially rational voting rules such that $p^i(h_t)$ is accepted (and if the equality is strict for $\frac{n+1}{2}$ players the proposal is accepted under all sequentially rational voting rules). By applying the definition of p^i we see that if

$$\begin{aligned} m_t &\geq \delta V_t\left(p,0,\frac{n+1}{2}\right) \text{ and} \\ \frac{2(1-m_t)}{n-1} &\geq \delta V_t\left(p,0,\frac{n+1}{2}\right) \end{aligned} \tag{1.16}$$

then the value of the proposal is higher than the discounted "rejection" continuation value for the proposer and all agents who receive a share $\frac{2(1-m_t)}{n-1}$, so by Lemma 6 it is consistent with subgame perfection that $p^i(h_t)$ is accepted if we can verify that the inequalities above holds. Note that

$$\begin{aligned} m_t &= 1-\frac{\delta(n-1)}{2}\left(\frac{m_{t-1}}{n}+\frac{2(1-m_{t-1})}{n+1}\right)=1-\delta\frac{(n-1)}{2}V_{t-1}(x,0) \quad (1.17)\\ &\Updownarrow \\ \frac{2(1-m_t)}{n-1} &= \delta V_{t-1}(x,0) \end{aligned}$$

and that if $m_{t-1}<1$ then

$$\begin{aligned} V_{t-1}(x,0) &= \frac{1}{n}m_{t-1}+\frac{(n-1)2(1-m_{t-1})}{(n+1)(n-1)} > \quad (1.18)\\ &> \frac{1}{n}m_{t-1}+\frac{(n-1)2(1-m_{t-1})}{n(n+1)(n-1)}=V_{t-1}\left(p,0,\frac{n+1}{2}\right) \end{aligned}$$

Hence, if $m_{t-1}<1$ and $\delta>0$ it follows that $p^{ij}(h_t)>\delta V_{t-1}\left(p,0,\frac{n+1}{2}\right)$ for the $\frac{n-1}{2}$ agents who receive $\frac{2(1-m_t)}{n-1}$ under the proposal $p^i(h_t)$. Left to verify is that the current proposer has no incentives to reject the proposal. But given that the proposer would vote to reject her own proposal $p^i(h_t)$ she is always excluded from the winning majority in the next period unless called to propose again. The discounted continuation value for the proposer if she rejects is thus $\frac{\delta m_{t-1}}{n}$. Hence, if

$$\frac{\delta m_{t-1}}{n}\le 1-\frac{\delta(n-1)}{2}\left(\frac{m_{t-1}}{n}+\frac{2(1-m_{t-1})}{n+1}\right), \quad (1.19)$$

then the proposer has no incentives to reject. Suppose for contradiction that this inequality is not met. After some rearranging we then have

$$\begin{aligned} \frac{\delta m_{t-1}}{n}\frac{(n+1)}{2} &> 1-\frac{\delta(1-m_{t-1})(n-1)}{n+1} \quad (1.20)\\ &\Updownarrow \\ m_{t-1}+\frac{(1-m_{t-1})n(n-1)}{(n+1)^2} &> \frac{2n}{\delta(n+1)} \end{aligned}$$

Intuitively we see that the left hand side goes to one as $n\to\infty$, while the right hand side is always strictly larger than one, so we see directly that if the legislature is large enough, then the proposer has a strict incentive to vote to accept the proposal. However, as we will show next, there is no incentive to deviate even if the legislature is small. The right hand side of (1.20) is decreasing in δ. Thus, if the inequality above is satisfied for some $\delta\in(0,1)$, it is also satisfied for $\delta=1$. Hence, the proposer has an incentive to reject proposal only if

$$\begin{aligned} m_{t-1}+\frac{(1-m_{t-1})n(n-1)}{(n+1)^2} &> \frac{2n}{(n+1)}\Leftrightarrow \quad (1.21)\\ m_{t-1}\left((n+1)^2-n(n-1)\right)+n(n-1) &= m_{t-1}(3n+1)+n(n-1) > \\ &> 2n(n+1) \end{aligned}$$

Since $m_t < 1$ for all t we must then have that

$$3n + 1 + n(n-1) > 2n(n+1) \Leftrightarrow 1 > n^2 \tag{1.22}$$

which is a contradiction. We conclude that the proposer and $\frac{n-1}{2}$ other players always have an incentive to accept the proposal. It follows that if all agents follow the specified proposal rules after any history, then there exists sequentially rational voting rules such that any proposal in accordance to these rules will be accepted.

Left to verify is that the proposer has no incentive to deviate when making a proposal. But, by construction, m_t equals the payoff the proposer would get under the best possible deviation that would be accepted by a majority. Hence, we only need to check that the proposer does not want to propose something that would be rejected. But such proposal would give a discounted expected payoff of $\frac{\delta m_{t-1}}{n}$, so the calculations above shows that this is not a profitable deviation. ■

While the proposal rules described together with appropriate voting rules together with some "initial proposal" satisfying certain conditions always constitutes a subgame perfect equilibrium, there are some limits as to what these strategies can achieve. Let m^* be defined as

$$m^* = \frac{1 - \frac{\delta(n-1)}{(n+1)}}{1 - \frac{\delta(n-1)^2}{2n(n+1)}}. \tag{1.23}$$

Note that $m^* \to 0$ as $\delta \to 1$ and $n \to \infty$. This has the implication that if n is large and δ is close to one, then almost every allocation can be supported as a subgame perfect equilibrium already with the proposal rules considered so far (however, as we will see later, we can do better than this).

Lemma 9. *There exists some $T < \infty$ such that if there are at least T rounds of bargaining, then any initial proposals $p(h_T) = (p^1(h_T), ..., p^n(h_T))$ where for each player $i \in I$, $p^{ii}(h_T) > m^*$ and $p^{ij}(h_T) > \delta\left[\frac{1}{n}m^* + \frac{2}{n(n+1)}(1-m^*)\right]$ for $\frac{n-1}{2}$ players $j \neq i$ can be supported as a subgame perfect equilibrium.*

Proof. We first show that $\{m_t\}$ is a monotonically decreasing sequence with $m_t \in [0,1]$ for all t and $m_t \to m^*$. To see that $m_t \in [0,1]$ for all t we note first that $m_1 = 1 \in [0,1]$. If $m_{t-1} \in [0,1]$ we have that

$$\begin{aligned} m_t &= 1 - \frac{\delta(n-1)}{2}\left(\frac{m_{t-1}}{n} + \frac{2(1-m_{t-1})}{n+1}\right) = \\ &= 1 + \frac{m_{t-1}\delta(n-1)^2}{2n(n+1)} - \frac{\delta(n-1)}{(n+1)} < 1 + \frac{\delta(n-1)^2}{2n(n+1)} - \frac{\delta(n-1)}{(n+1)} < 1 \end{aligned} \tag{1.24}$$

and

$$m_t = 1 + \frac{m_{t-1}\delta(n-1)^2}{2n(n+1)} - \frac{\delta(n-1)}{(n+1)} > 1 - \frac{\delta(n-1)}{(n+1)} > 0. \tag{1.25}$$

By induction it follows that $m_t \in [0,1]$ for all t. To show that the sequence is monotonically decreasing we note that $m_2 < m_1 = 1$ and that

$$m_t - m_{t-1} = \frac{\delta(n-1)^2}{2n(n+1)}(m_{t-1} - m_{t-2}). \tag{1.26}$$

It follows by induction that m_t is monotonically decreasing. Since $\{m_t\}$ is a monotonic sequence in a compact set it must have a limit in $[0,1]$ and it follows by straightforward algebra that this limit is given by m^*. The result then follows since if the inequalities hold in the limit, then there exists some T large enough such that $p^{ii}(h_T) > m_T$ and $p^{ij}(h_T) > \delta\left[\frac{1}{n}m_T + \frac{2}{n(n+1)}(1-m_T)\right]$ for $\frac{n-1}{2}$ players $j \neq i$. By Lemma 8, the proposer is (weakly) better off by proposing and accepting m_T than any other deviation. Hence the proposer has a strict incentive to propose and accept $p^i(h_T)$. By choosing T large enough, $m_T - m_{T-1}$ can be made arbitrarily small, so $\frac{1}{n}m_T + \frac{2}{n(n+1)}(1-m_T) \approx \frac{1}{n}m_{T-1} + \frac{2}{n(n+1)}(1-m_{T-1})$, which is the value of the game in the beginning of time $T-1$ for an agent (different from the proposer) who voted to reject. Thus, if all inequalities holds, then at least $\frac{n+1}{2}$ players who have a strict incentive to accept the initial proposal. The result then follows since there by Lemma 8 is some subgame perfect equilibrium where the proposals are according to p^i for all players $i \in I$ and all histories $h_t \in H_t$ and where these proposals are accepted for sure. ■

To complete the proof we must augment the proposal rule used after histories where the proposal in the last period was not according to the candidate equilibrium strategies. As above we will specify the rule in terms of an unknown sequences of numbers $\{y_t\}_{t=T'}^{T}$, $\{w_t\}_{t=T'}^{T}$ where

$$\begin{aligned} y_T &= m_T \\ w_T &= \delta\left[\frac{1}{n}m_{T-1} + \frac{2}{n(n+1)}(1-m_{T-1})\right], \end{aligned} \tag{1.27}$$

and the particular choices of y_t, m_t for $t > T$ will be determined below. The reader may however want to keep in mind that the idea is to construct these as decreasing sequences approaching zero.

Let $t \geq T'$ (to be specified below) and suppose that i' proposed $x_{t+1} \neq p^{i'}(h_{t+1})$ in round $t+1$. Then, for all histories h_t following h_{t+1} such that x_{t+1} is rejected let $L \subset I$ be an arbitrary (but commonly known before the voting) subset consisting of a group of $\frac{n+1}{2}$ players (other than the proposer i') who receives the least under x_{t+1}[21]. If

$$y_t + \sum_{j \in L} \max\left\{\frac{x_{t+1}^j}{\delta}, w_t\right\} < 1 \tag{1.28}$$

and if the proposer at time t ,$i \in L$, then let $p^i(h_t)$ be given by

$$p^{ij}(h_t) = \begin{cases} y_t + \max\left\{\frac{x_{t+1}^j}{\delta}, w_t\right\} & \text{for } i = j \\ \max\left\{\frac{x_{t+1}^j}{\delta}, w_t\right\} & \text{for all } j \in L \mid \{i\} \\ 0 & \text{otherwise} \end{cases} \tag{1.29}$$

while if $i \notin L$ then

$$p^{ij}(h_t) = \begin{cases} y_t & \text{for } i = j \\ \max\left\{\frac{x_{t+1}^j}{\delta}, w_t\right\} & \text{for all } j \in L \\ 0 & \text{otherwise} \end{cases} \tag{1.30}$$

For all x_{t+1} such that

$$y_t + \sum_{j \in L} \max\left\{\frac{x_{t+1}^j}{\delta}, w_t\right\} \geq 1 \tag{1.31}$$

we let the proposal rules be as above, with m_t replaced by y_t.

Lemma 10. *Let $i' \in I$ denote the proposer at time t and suppose that all agents $i \in I$ follow the proposal rule p^i at time $t-1 \geq T'$ and that the voting rules at time $t-1$ are such that $p^i(h_{t-1})$ is always accepted for any h_{t-1} and any $i \in I$. Then, if the current proposer i' proposes some allocation $x_t \neq p^{i'}(h_t)$ such that $y_t + \sum_{j \in L} \max\left\{\frac{x_{t+1}^j}{\delta}, w_t\right\} < 1$ the proposal must be rejected under any sequentially rational voting rules.*

Proof. It is easily seen that the discounted value of the continuation payoffs for all $j \in L$ are strictly higher than payoffs if x_t is accepted, thus $\frac{n+1}{2}$ agents have a strict incentive to reject the proposal. Applying the backwards induction argument in Part i of Lemma 4 the result follows. ■

Lemma 11. *Let $t > T+1$ and suppose that $y_t \geq k y_{t-1}$ and suppose there exists some subgame perfect equilibrium where all agents are proposing in accordance to p^i at time $t-1$ and that any such proposal is accepted. Then if $y_{t-1} + \frac{n-1}{2} w_{t-1} < 1$ there exists some $k \in (0,1)$ and some $\delta^* < 1$ such that if $\delta \geq \delta^*$, then the proposer has no incentive to deviate from the specified proposal rule.*

[21] I.e. without loss of generality, let $i' = n$ be the proposer and assume that $x_1 \leq x_2 \leq .... \leq x_{\frac{n+1}{2}} \leq x_{\frac{n+3}{2}} ... \leq x_{n-1}$. Then L is determined by payoffs only if $x_{\frac{n+1}{2}} < x_{\frac{n+3}{2}}$, while if $x_{\frac{n+1}{2}} = x_{\frac{n+3}{2}}$ there are choices how to specify the set. No matter what is used as a "tie-breaker" all agents agree.

Proof. Without loss of generality, index the agents so that n is the proposer and that $x^1 \leq ... \leq x^{n-1}$ for the other players. For ease of notation we let $m = (n-1)/2$. By Lemma 10 a necessary condition for a deviant proposal to be accepted at time t is that

$$y_{t-1} + \sum_{i=1}^{m} \max\left\{\frac{x^i}{\delta}, w_{t-1}\right\} \geq 1 \tag{1.32}$$

Furthermore, a deviant proposal is profitable only if $x^n > y_t \geq k y_{t-1}$. By 1.32, if $y_{t-1} + \frac{n-1}{2} w_{t-1} < 1$, then a necessary condition for x to be a profitable deviation is that

$$\frac{1}{\delta}\sum_{i=1}^{m} x_i \geq 1 - y_{t-1} - \frac{n-1}{2} w_{t-1} \tag{1.33}$$

Thus, if δ and k are sufficiently close to unity, a necessary condition for x to be a profitable deviation is that

$$\begin{aligned} 1 &\geq \sum_{i=1}^{n} x_i = \sum_{i=1}^{m} x_i + \sum_{i=m+1}^{n-1} x_i + x_n > \\ &> 1 - y_{t-1} - \frac{n-1}{2} w_{t-1} + \sum_{i=m+1}^{n-1} x_i + y_{t-1} - \epsilon = \\ &= 1 - \frac{n-1}{2} w_{t-1} + \sum_{i=m+1}^{n-1} x_i - \epsilon \end{aligned} \tag{1.34}$$

where by choosing δ, k arbitrarily close to one, ϵ can be made arbitrarily small. But this means that there is at least one agent $i \in \{m+1, ..., n-1\}$ for whom $x_i \leq w_{t-1} + \frac{2\epsilon}{n-1}$. Hence $x^i \leq w_{t-1} + \frac{2\epsilon}{n-1}$ for $i = 1, ..., m$. But now, if $y_{t-1} + \frac{n-1}{2} w_{t-1} < 1$, ϵ is small enough and δ is sufficiently close to unity, then $y_{t-1} + \sum_{i=1}^{m} \max\left\{\frac{x^i}{\delta}, w_{t-1}\right\} < 1$. Hence there can be no profitable deviation. ■

Proof. (Proposition 2) To finish the argument we now only have to construct appropriate sequences $\{y_t\}_{t=T}^{T'}, \{w_t\}_{t=T}^{T'}$ such that $y_t, w_t \to 0$ and such that there always is a majority that has an interest to accept the proposal after any history of play. To do this let

$$\begin{aligned} y_t &= \min\left(k y_{t-1}, 1 - \frac{\delta(n-1)}{2}\left(\frac{m_{t-1}}{n} + \frac{2(1-m_{t-1})}{n+1}\right)\right) \\ w_t &= \delta\left(\frac{y_{t-1}}{n} + \frac{2(1-y_{t-1})}{n+1}\right) \end{aligned} \tag{1.35}$$

Given that the proposal in the next period will be according to the rule specified, no agent receiving w_t has an incentive to reject proposal (same argument as in Lemma 8). Since the proposer get a continuation payoff of $\frac{\delta y_{t-1}}{n}$ if rejecting his own proposal, proposer has a strict incentive to accept the proposal. Thus, there are $\frac{n+1}{2}$ players who has an incentive to accept $p^i(h_t)$ and it follows that there exists sequentially rational voting rules so that the proposal is accepted with probability one if proposed. By Lemma 8, if $y_{t-1} + \frac{n-1}{2} w_{t-1} < 1$, then there is no profitable deviation for the proposer, so left to verify is that this condition holds for every $t \geq T$. We can find T such that the inequality holds for $t = T$ if

$$\begin{aligned} m^* + \frac{n-1}{2}\left[\frac{1}{n}m^* + \frac{2}{n(n+1)}(1-m^*)\right] &< 1 \\ &\Updownarrow \\ m^* &< \frac{2(n^2+1)}{2n(n+1)+(n-3)(n+1)} \end{aligned} \tag{1.36}$$

which, since $m^* < 1$, obviously holds as long as $n \geq 3$. Since y_t is strictly decreasing for $t > T$ and since consequently also w_t is strictly decreasing the inequality holds for all $t > T$. Thus, at each time t, the proposer will stick to the candidate equilibrium proposal rule and any such proposal will be accepted. Since $y_t, w_t \to 0$ as $T' \to \infty$, we can thus support an arbitrary initial proposal as an equilibrium, given that δ is sufficiently close to unity. ■

1.6.2. Proof of Lemma 1

Proof. Without loss of generality we rename the players so that

$$\delta_1 V_1(t) \geq ... \geq \delta_{\frac{n-1}{2}} V_{\frac{n-1}{2}}(t) > \delta_{\frac{n+1}{2}} V_{\frac{n+1}{2}}(t) > \delta_{\frac{n+3}{2}} V_{\frac{n+3}{2}}(t) \geq ... > \delta_n V_n(t). \tag{1.37}$$

Thus, $A = \left\{1, .., \frac{n-1}{2}\right\}$, $B = \left\{\frac{n+3}{2}, ..., n\right\}$ and $m = \frac{n+1}{2}$. We want to show that the all subgame perfect equilibria has the property that regardless of what player is recognized to make a proposal at time $t+1$, this player will propose to give $\delta_i V_i(t)$ to the (uniquely determined) set of $(n-1)/2$ players different from the proposer such that the discounted continuation payoff is lower than for all other players. I.e. the unique subgame perfect proposal rules at time $t+1$ are given by $p^j : H_{t+1} \to X$, where if the proposer $j \notin A$, then

$$p_{t+1}^{ji}(h_{t+1}) = \begin{matrix} 1 - \sum_{i \in A} \delta_i V_i(t) & \text{for } i = j \\ \delta_i V_i(t) & \text{for } i \in A \\ 0 & \text{otherwise} \end{matrix}, \tag{1.38}$$

while if $j \in A$, then

$$p_{t+1}^{ji}(h_{t+1}) = \begin{matrix} 1 - \sum_{i \in A \cup \left\{\frac{n+1}{2}\right\}} \delta_i V_i(t) & \text{for } i = j \\ \delta_i V_i(t) & \text{for } i \in A \cup \left\{\frac{n+1}{3}\right\} \\ 0 & \text{otherwise} \end{matrix}, \tag{1.39}$$

Clearly, everybody following the voting rule "accept an offer x at time $t+1$ if and only if $x^i \geq \delta_i V_i(t)$ is sequentially rational give continuation player according to any subgame perfect equilibrium (since by assumption the continuation valuation at time t are the same in all subgame perfect equilibria). Given these continuation strategies the best proposal that is accepted are given by p^j and since

$$\delta_j V_j(t) \leq \delta_j \left(1 - \sum_{i \neq j} V_j(t)\right) \leq \delta_j p_{t+1}^{jj}(h_{t+1}) \tag{1.40}$$

the proposer has no incentive to propose something that would be rejected. Hence the proposed strategies constitutes a subgame perfect equilibrium of the game with $t+1$ rounds of bargaining. Next we want to show that there are no other proposals that are consistent with subgame perfection. For contradiction, suppose that there is a subgame perfect equilibrium proposal that has at least $(n+1)/2$ players to receive shares $x^i < \delta_i V_i(t)$. Now, if the last player among these in the voting order is pivotal, then she has a strict incentive to reject. Hence if $(n-1)/2$ players have already voted in favor of rejection when the $(n+1)/2$th player is to vote, the proposal must be rejected for sure. Foreseeing this, the $(n-1)/2$th player in the voting order (among those with strict incentive to reject) knows that if at least $(n-3)/2$ players have already voted to reject and she votes to reject, then the proposal will be rejected for sure. Thus, the proposal must be rejected for sure in any subgame perfect equilibrium, given that $(n-3)/2$ votes have been cast in favor of rejection when the $(n-1)/2$th voter with a strict incentive to reject is to vote. Proceeding by (backwards) induction it follows that the proposal must be rejected for sure in any subgame perfect equilibrium. Hence there is no alternative subgame perfect equilibrium that gives the proposer a higher payoff.

A symmetric argument shows that if the proposer and $(n-1)/2$ other players are strictly better off when the proposal is implemented, then the proposal must be accepted for sure in any subgame perfect equilibrium. So suppose that there is some subgame perfect equilibrium where the proposer offers x at time $t+1$ and is worse off than in the equilibrium with proposal rule p_{t+1}^j. Then we can construct an alternative proposal y such that the proposer is better off and $(n-1)/2$ other players get a strictly higher share than $\delta_i V_i(t)$. But, such proposal must be accepted for sure in any subgame perfect equilibrium and leaves the proposer strictly better off. Thus the original proposal x could not be a part of any subgame perfect equilibrium. Hence, there is no alternative subgame perfect equilibrium that gives the proposer a strictly lower payoff. Combining the fact that the payoff of the proposer must be given by $p_{t+1}^{jj}(h_{t+1})$ in all subgame perfect equilibria with the condition that unless $(n-1)/2$ players are getting a share of at least $\delta_i V_i(t)$, it follows that the equilibrium proposal rules are uniquely determined by p_{t+1}^j. ■

1.6.3. Proof of Lemma 3

Proof. Suppose A_t has not full rank. Then at least one row in A_t can be expressed as a linear combination of the others. Since the first $\frac{n-1}{2}$ rows have a strictly positive entry in a column where all other rows have zeros it follows that any row from $\begin{pmatrix} \Lambda_t & B_t \end{pmatrix}$ is linearly independent of all other rows in A_t. Thus, A_t has full rank if and only if C_t has full rank. From basic linear algebra we know that if C_t and D are conformable matrices with full rank, then DC_t has full rank. This is useful since if D is chosen wisely we can simplify the problem. For easy of notation we relabel player $(n+1)/2$ as player1, player $(n+3)/2$ as player 2,...., player n as player $(n+1)/2 = m$. Let D be given by

$$D = \begin{pmatrix} \frac{1}{\delta_1} & 0 & \dots & & 0 \\ 0 & \frac{1}{\delta_2} & 0 & \dots & \dots \\ \dots & & \dots. & & \\ & & & & \dots \\ \dots & & & & 0 \\ 0 & \dots & & \dots\ 0 & \frac{1}{\delta_m} \end{pmatrix} \tag{1.41}$$

Then $C_t^* = DC_t$ is given by

$$\begin{pmatrix} \frac{1}{n\delta_1 V_1(t-1)} + \frac{n-1}{2n} & -\frac{1}{n} & -\frac{1}{n} & \dots & \dots & -\frac{1}{n} \\ -\frac{1}{n} & \frac{1}{n\delta_2 V_2(t-1)} + 1 & -\frac{1}{n} & \dots & & \dots \\ \dots & -\frac{1}{n} & \dots & & & \\ & \dots & & & & \\ & & & & & \\ \dots & & & & \dots & -\frac{1}{n} \\ -\frac{1}{n} & \dots & & & -\frac{1}{n} & \frac{1}{n\delta_m V_m(t-1)} + 1 \end{pmatrix} \tag{1.42}$$

If this matrix has full rank, then C_t as well as A_t has full rank, which is what we want to show. Although this matrix looks simple, the result is by no means obvious. The intuitive reason for why C_t^* has full rank is that the diagonal elements are pretty large relative to the off diagonal elements, so the matrix is "sufficiently close to the identity matrix". In fact, it can be shown that C_t^* is a "dominant diagonal matrix" and since any such matrix has full rank, C_t^* has full rank. However, we will not follow this route, but prove the full rank of C_t^* by a direct calculation.

Suppose first that the first row, c_1, is linearly dependent of the other rows. Then there exists scalars $k_2, ..., k_m$ such that $c_1 = \sum_{i=2}^{m} k_i c_i$. Exploiting the structure of the matrix C_t^* :

$$\begin{cases} \frac{1}{\delta_1 V_1(t-1)} + \frac{(n-1)}{2} = -\sum_{i=2}^{m} k_i & \text{(1 equation)} \\ -1 = -\sum_{i=2}^{m} k_i + k_j \left(n+1+\frac{1}{\delta_j V_j(t-1)}\right) & \text{for } j = 2, ..., \frac{n+1}{2} \\ & (\frac{n-1}{2} \text{ equations}) \end{cases} \tag{1.43}$$

Combining these equations we get

$$0 = \left(\frac{1}{\delta_1 V_1 (t-1)} + \frac{(n+1)}{2}\right) + k_j \left(n+1+\frac{1}{\delta_j V_j (t-1)}\right) \tag{1.44}$$

so a necessary condition for linear dependence is that

$$k_j = -\frac{\left[\frac{1}{\delta_1 V_1(t-1)} + \frac{(n+1)}{2}\right]}{\left[n+1+\frac{1}{\delta_j V_j(t-1)}\right]} \text{ for } j = 2, ..., \frac{n+1}{2} \tag{1.45}$$

Now summing over j we get

$$\sum_{j=2}^{m} k_j = -\left[\frac{1}{\delta_1 V_1 (t-1)} + \frac{(n+1)}{2}\right] \sum_{j=2}^{m} \left[\frac{1}{n+1+\frac{1}{\delta_j V_j(t-1)}}\right]$$

Hence if all $\frac{n+1}{2}$ conditions for c_1 to be linearly dependent on $c_2, ...c_m$ are to be satisfied

$$\frac{\frac{1}{\delta_1 V_1(t-1)} + \frac{(n-1)}{2}}{\frac{1}{\delta_1 V_1(t-1)} + \frac{(n+1)}{2}} = \sum_{j=2}^{m} \left[\frac{1}{n+1+\frac{1}{\delta_j V_j(t-1)}}\right] \tag{1.46}$$

We can show that

$$\frac{\frac{1}{\delta_1 V_1(t-1)} + \frac{(n-1)}{2}}{\frac{1}{\delta_1 V_1(t-1)} + \frac{(n+1)}{2}} > \frac{n-1}{n+1} \tag{1.47}$$

And that

$$\begin{aligned} \delta_j V_j (t-1) &< \frac{2}{n+1} \text{ for all } j \Leftrightarrow \frac{1}{\delta_j V_j (t-1)} > \frac{n+1}{2} \Leftrightarrow \\ &\Leftrightarrow \frac{1}{n+1+\frac{1}{\delta_j V_j(t-1)}} < \frac{1}{n+1+\frac{n+1}{2}} = \frac{2}{3(n+1)} \Longrightarrow \\ &\Longrightarrow \sum_{j=2}^{m} \left[\frac{1}{n+1+\frac{1}{\delta_j V_j(t-1)}}\right] < \frac{(n-1)}{3(n+1)} < \frac{n-1}{n+1} \end{aligned} \tag{1.48}$$

Hence

$$\frac{n-1}{n+1} < \frac{\frac{1}{\delta_1 V_1(t-1)} + \frac{(n-1)}{2}}{\frac{1}{\delta_1 V_1(t-1)} + \frac{(n+1)}{2}} = \sum_{j=2}^{m} \left[\frac{1}{n+1+\frac{1}{\delta_j V_j(t-1)}}\right] < \frac{(n-1)}{3(n+1)} < \frac{n-1}{n+1} \tag{1.49}$$

which is a contradiction. For any other row, say the k^{th} we can proceed in the same way. If r_k is a linear combination of the other rows there are scalars $k_1, ..., k_{k-1}, k_{k+1}, ..., k_m$ such that

$$\begin{cases} -1 = k_1 \left(\frac{1}{\delta_1 V_1(t-1)} + \frac{(n+1)}{2}\right) - \sum_{i\neq k} k_i & \text{(1 equation)} \\ \frac{1}{\delta_k V_k(t-1)} + n = -\sum_{i\neq k} k_i & \text{(1 equation)} \\ -1 = -\sum_{i\neq k} k_i + k_j \left(n+1+\frac{1}{\delta_j V_j(t-1)}\right) & \text{for } j \neq 1, k \ \left(\frac{n-3}{2} \text{ equations}\right) \end{cases} \tag{1.50}$$

So we assume that $\frac{1}{\delta_k V_k(t-1)} + n = -\sum_{i\neq k} k_i$ and solve for k_i to get a set of necessary conditions for row k to be a linear combination of the other rows,

$$\begin{cases} k_1 = -\frac{\left[n+1+\frac{1}{n\delta_k V_k(t-1)}\right]}{\left(\frac{(n+1)}{2}+\frac{1}{\delta_1 V_1(t-1)}\right)} \\ k_j = -\frac{\left[n+1+\frac{1}{n\delta_k V_k(t-1)}\right]}{\left(n+1+\frac{1}{\delta_j V_j(t-1)}\right)} & \text{for } j \neq 1, k \end{cases} \tag{1.51}$$

summing

$$\sum_{j\neq k} k_i = -\left[n+1+\frac{1}{n\delta_k V_k(t-1)}\right]\left[\frac{1}{\left(\frac{n+1}{2}+\frac{1}{\delta_1 V_1(t-1)}\right)} + \sum_{j\neq 1,k}\frac{1}{\left(n+1+\frac{1}{\delta_j V_j(t-1)}\right)}\right] \tag{1.52}$$

Hence in order for the assumption $n + \frac{1}{\delta_k V_k(t-1)} = -\sum_{i\neq k} k_i$ to be fulfilled it must for instance be that

$$\frac{\left[n+\frac{1}{\delta_k V_k(t-1)}\right]}{\left[n+1+\frac{1}{n\delta_k V_k(t-1)}\right]} = \left[\frac{1}{\left(\frac{n+1}{2}+\frac{1}{\delta_1 V_1(t-1)}\right)} + \sum_{j\neq 1,k}\frac{1}{\left(n+1+\frac{1}{\delta_j V_j(t-1)}\right)}\right] \tag{1.53}$$

This is however impossible. To see this we note that

$\frac{1}{\left(n+1+\frac{1}{\delta_j V_j(t-1)}\right)} < \frac{1}{n+1} \Rightarrow \sum_{j\neq 1,k}\frac{1}{\left(n+1+\frac{1}{\delta_j V_j(t-1)}\right)} < \frac{n-3}{2(n+1)}$ and $\frac{1}{\left(\frac{n+1}{2}+\frac{1}{\delta_1 V_1(t-1)}\right)} < \frac{2}{n+1}$. Combining these facts we find that

$$\begin{aligned} LHS &= \left[\frac{1}{\left(\frac{n+1}{2}+\frac{1}{\delta_1 V_1(t-1)}\right)} + \sum_{j\neq 1,k}\frac{1}{\left(n+1+\frac{1}{\delta_j V_j(t-1)}\right)}\right] < \\ &< \frac{2}{n+1} + \frac{n-3}{2(n+1)} = \frac{1}{2} \end{aligned} \tag{1.54}$$

Also we have that

$$RHS = \frac{\left[n+\frac{1}{\delta_k V_k(t-1)}\right]}{\left[n+1+\frac{1}{n\delta_k V_k(t-1)}\right]} > \frac{n}{n+1} \tag{1.55}$$

Combining these results

$$\frac{n}{n+1} < RHS = LHS < \frac{1}{2} \Rightarrow n < 1 \tag{1.56}$$

which is a contradiction. ■

2. AFFIRMATIVE ACTION IN A COMPETITIVE ECONOMY

Andrea Moro*
University of Pennsylvania
and
Peter Norman*
University of Pennsylvania

October 29, 1996

Abstract

This paper analyzes statistical discrimination in a model with endogenous human capital formation and a frictionless labor market. It is shown that in the presence of two distinguishable but *ex ante* identical groups of workers discrimination is sustainable as an equilibrium outcome. This is true irrespective of whether there are multiple equilibria when the groups have no distinguishable characteristics. When an affirmative action policy consisting of an employment quota is introduced in the model it is shown that affirmative action can "fail" in the sense that there may still be equilibria where the groups are treated differently. However, the incentives to invest for agents in the disadvantaged group are better in any equilibrium under affirmative action than in the most discriminatory equilibrium without the policy. Thus, the lower bound on the fraction of agents from the disadvantaged group who invest in their human capital is raised by the policy. The welfare effects are ambiguous. It is demonstrated that the policy may increase the incentives

[0]We are grateful to Stephen Coate, Marcos Lisboa, George Mailath, Stephen Morris, Andrew Postlewaite and seminar participants at University of Pennsylvania and Institute of International Economic Studies, Stockholm for many helpful comments and interesting discussions.

to invest and reduce the expected payoffs for all agents in the target group simultaneously. Indeed, the policy may hurt the intended beneficiaries even when the initial equilibrium is the worst equilibrium for the targeted group.

2.1. Introduction

Since its introduction in the sixties, affirmative action has been and remains one of the most controversial policies to combat discrimination in the labor market. An economist has little to say about issues on fairness and constitutionality, which are extensively discussed in the popular and political debate. However, there are important aspects of affirmative action that can be analyzed using economic theory and relatively little has been done.

In particular, the popular debate often focuses on the effects on incentives of the intended beneficiaries. On the one hand side, opponents of affirmative action often argue that affirmative action makes it easier for unqualified members of the target groups to obtain relatively well paid jobs. This, it is argued, reduces the incentives to invest in their skills for members of the target groups, which means that the real problem, namely that skills are unevenly distributed across groups, is only aggravated by affirmative action. On the other side, proponents of affirmative action argue that minorities and in some cases women are at least partially excluded from the more attractive parts of the labor market and that they for this reason simply do not have the same incentives to make human capital investments. Affirmative action policies with numerical goals for hirings of candidates from the discriminated groups helps overcome the situation by forcing employers to hire people from the disadvantaged groups and therefore create incentives for members of these groups to invest in their personal skills.

The purpose of this paper is to analyze what effects affirmative action policies may have on the incentives to invest, in particular for workers from the groups the policy is intended to help. Furthermore, since opponents often claim that the policy only helps already well situated members of the minority groups, we are also interested in identifying winners and losers of affirmative action. However, while our framework in principle allows us to do this, our understanding of the welfare effects of affirmative action is still very incomplete.

In order to study the effects of affirmative action and other anti discriminatory policies we need a model with discrimination as a possible equilibrium outcome. Here there are two main strands in the literature. One approach, pioneered by Becker [9], explains discrimination from preferences.

In this class of models employers prefer to hire candidates from the same group, workers prefer to work with coworkers from the same group or consumers are unwilling to buy products produced by firms' employing workers from other groups.

The main alternative to these taste based models is a statistical theory of discrimination. building on work by Arrow [4] and Phelps [24]. Here the main idea is that when worker skills are imperfectly observable discrimination may occur although firms maximize profits and workers have no preferences about their coworkers' group identity: race, sex, religion etc... may serve as a proxy for productivity if the distributions are different across groups. When each worker can affect their own productivity by human capital investments discrimination may occur in equilibrium even if the groups are identical in terms of "intrinsic abilities" or costs of investment in human capital. In this paper we consider the effects of affirmative action within a model of statistical discrimination.

While there is a large theoretical literature on discrimination in general and discrimination on the labor market in particular, surprisingly little attention has been paid to policy analysis. Notable exceptions are Lundberg and Startz [19], Lundberg [18] and Coate and Loury [12][1]. In Lundberg and Startz [19] it is shown that an equal opportunity policy prohibiting the firms from making wages dependent upon group identity may be an efficiency enhancing policy in a model with statistical discrimination. In Lundberg [18], it is noted that enforcement of this type of policy may be very difficult since there will be incentives for firms to evade the policy by using other variables as proxies for group identity. The main concern of the paper is to find regulatory policies that implement the equal opportunity laws under different informational assumptions.

The paper most closely related to our work is Coate and Loury [12] where the effects of employment quotas are studied in a setup where discrimination is in job assignments rather than in wages. In their model, output can be produced using two different technologies and workers face a costly human capital investment, which if undertaken makes them productive in the more advanced technology. The sole decision made by employers' is how to assign a number of randomly drawn workers in jobs using either of the two technologies based on an imperfect signal of each workers' productivity in the more advanced job. Whenever there are multiple equilibria in the model there will be equilibria where groups are treated differently.

It is shown that there are circumstances under which all equilibria with the affirmative action policy are such that investment behavior is the same in both groups. However, under equally

[1] All these papers are focusing on statistical discrimination. Welch [30] and Kahn [17] studies employment quotas in models where discrimination is taste based.

plausible circumstances there are still equilibria where groups behave differently and the employers (rationally) perceive members of one of the groups to be less capable. Indeed, it is shown that group disparity of investment behavior may actually increase as a result of affirmative action.

The intuition for this possible failure of affirmative action is simple. Consider a situation where the fraction of investors is lower in group a than in group b and make the thought experiment that these fractions remains the same even after the introduction of affirmative action. In order to comply with the policy this means that employers must employ agents from group a in the more advanced job who are (rationally) perceived to have a lower probability of being productive than all agents from group b . Hence both agents who have invested and agents who have not invested are more likely to be employed in the skilled job and whether this improves the incentives to invest for agents in group a or not depends on particularities of the probability distributions of the noisy signal.

While the logic may sound compelling the analysis in Coate and Loury [12] raises some questions. Wages as well as the distribution of workers available for any firm are fixed exogenously in their model. In a world where firms are competing with each other to attract workers these assumptions do not make much sense. Rather one would think that equilibrium wages would depend on investment behavior of the workers and policy parameters, which means that the change in the incentives to invest would also depend on how the policy affects wages. In particular, since the expected marginal productivity is increasing in the signal for agents in the complex technology one would think that wages would also be increasing in the signal. But then the expected wage conditional on the agent being employed in the more advanced technology will be higher for agents who undertake the investment and it seems that if firms were forced to employ more workers from the disadvantaged group in the advanced job this would indeed create better incentives to invest.

In our paper, human capital accumulation as well as the information technology is modeled as in Coate and Loury [12]. Individual workers have to decide whether to undertake a costly investment in human capital or not. This choice is unobservable to the firms but there is a publicly observable test available that contains information about the likelihood that a particular worker has undertaken the investment.

Instead of randomly assigning workers between firms we assume that the labor market works without frictions. Firms compete in a Bertrand fashion by offering wage schedules, where the wage is a function of the noisy signal. Apart from the fact that wages are endogenized our model departs from that of Coate and Loury in that the production technology exhibits complementaries between tasks. To be specific we assume that production requires input of labor in two tasks, a *complex* task and a *simple* task. It is assumed that only workers who have undertaken the investment are productive in the complex task, whereas all workers can perform the simple task effectively. Output is generated from the two types of labor input according to a standard neoclassical production function.

When we introduce two groups of workers which only differ by some payoff irrelevant but observable characteristic we show that discrimination is possible due to self confirming expectations about differences in behavior between the groups.

The complementarity in the production technology has several interesting consequences. Even if there is a unique equilibrium in the model where there are no observable payoff irrelevant characteristics there will, under mild conditions, be equilibria with discrimination. The intuition is that groups can specialize as high quality and low quality workers respectively. While this hurts the group that specializes as low quality workers and also creates inefficiencies in investment behavior it does reduce the informational problem for the firms[2]. It should be noted in this context that in models where discrimination is explained as different groups coordinating on different equilibria in some "base model", as for example in Spence [29], Akerlof [2] and Coate and Loury [12], there are no conflicts of interests between groups. The discriminated group is discriminated simply because of coordination on a worse equilibrium than the other group and if this coordination failure could be resolved the other group need not be affected at all. In our model on the other hand the group with the higher fraction of investors unambiguously gains from discrimination since the supply of qualified workers is more scarce than otherwise.

The complementarity in production also has the consequence that group size matters in the determination of equilibria with discrimination. We find that the larger the group is, the more stringent are the conditions that must be satisfied in order to support a (particular type of) discrimination and the smaller is the differences in average earnings between groups (given that discrimination is still sustainable). In a loose sense, we interpret this to mean that in our model discrimination of a smaller group is more likely than discrimination of a larger group. To us this seems to conform with the stylized facts about discrimination: to our knowledge there is no other model with this property.

[2]It is indeed easy to visualize a version of our model where agents choose different types of human capital investment that enhances the productivity in different types of jobs. In such a model, discrimination may be efficiency enhancing. However, contrary to our framework such a model may also have the property that discrimination is voluntary in the sense that it may be incentive compatible to truthfully announce group identity if this would be unobservable.

Introducing an affirmative action policy consisting of an employment quota in the model we find that we in general cannot rule out the possibility of discriminatory equilibria. Hence the policy does not guarantee equal treatment across groups in equilibrium. However, this result alone should not be interpreted as a "failure" of affirmative action. While the ultimate goal of equality between groups is not guaranteed by the policy, it may still be that the policy is successful in the sense that the inequality is reduced. Indeed, we get some results in this direction.

In our model, the "direct effect" of affirmative action, i.e. the effect on the benefits of investment in human capital assuming that investment behavior is unchanged, is typically to increase the returns of investment for the discriminated group and to decrease them for the other group. However, we do not have a theory that predicts what particular equilibrium will occur after the introduction of the policy. Due to multiplicity of equilibria with and without the policy we must compare the full set of equilibria with and without affirmative action. The only thing that can be said in general is that the returns to investment and consequently also the fraction of agents who invest in the most discriminatory equilibrium without the policy is lower than in any equilibrium with affirmative action. The welfare effects are inconclusive. Output may decrease or increase as a result of the policy and by example we show that even if the starting point is the most discriminatory equilibrium, it is possible that the discriminated group is worse off with the policy.

The rest of this chapter is structured as follows. Section 2.2 contains the description of the one-group model and section 2.3 characterizes the equilibria of this model. In section 2.4 we extend the model by introducing two identical groups of workers and in section 2.5 we analyze the consequences of affirmative action. The discussion in section 2.6 concludes. Most proofs can be found in Section 2.7.

2.2. The Model

We assume that firms need to employ workers performing two different tasks to generate output. These tasks will be referred to as the *complex task* and the *simple task* respectively. On the labor market there are workers of two different types. Some workers, called *qualified* workers, are able to perform the complex task and others are not. Let C be the effective input of labor in the *complex task* for the firm, i.e. C equals the number of qualified workers employed in the skilled task. By S we denote the number of workers employed in the *simple task*. The output of the firm is then given by $y(C,S)$ where $y : R_+^2 \rightarrow R_+$ satisfies the standard neoclassical assumptions, i.e. it is a twice continuously differentiable function, strictly concave in both arguments, and:

A1 $y(\cdot,\cdot)$ is homogeneous of degree one [constant returns to scale]

A2 $\lim_{C\to 0} y_1(C,S) = \infty$ for any $S > 0$ and $\lim_{S\to 0} y_2(C,S) = \infty$ for any $C > 0$[3] [boundary behavior]

A3 $y(0,S) = y(C,0) = 0$ [both factors essential]

Since we make the extreme assumption that the additional output generated by *unqualified* workers in the *complex task* is zero only *qualified* workers would be hired for this task in a perfect information environment. However, in the model there will be some mismatch due to uncertainty about worker quality.

2.2.1. The Game

The timing of events is as follows: In **Stage 1** individual workers decide whether to invest or not in their human capital. After the investment decisions (**Stage 2**) each worker is assigned a signal θ by nature. In **Stage 3** firms simultaneously announce wage schedules (i.e. wages as functions of the signal) and in **Stage 4** workers choose which firm to work for. Finally, in **Stage 5** firms decide how to allocate the available workers between the two tasks.

For tractability we do not want the behavior of any individual worker to have any effect on aggregate behavior so we will assume that the population of workers is large, represented by a continuum.

The model will now be described in detail.

Stage 1. There is a continuum of agents with heterogeneous costs of investment. Each agent c has to choose an action $e \in \{e_q, e_u\}$, where $e = e_q$ means that the agent undertakes an investment in his human capital (and becomes a *qualified* worker) and $e = e_u$ that he does not. If agent c undertakes the investment he incurs a cost of c while no cost is incurred if the investment is not undertaken. The agents are distributed on the interval $[\underline{c}, \overline{c}] \subseteq R$ according to the continuous and strictly increasing distribution function G. We assume that $\underline{c} \leq 0$ and $\overline{c} > 0$.

[3]Subscripts are used to denote partial derivatives.

Stage 2. Each worker is assigned a noisy signal $\theta \in [0,1]$. The signal θ is distributed according to density f_q for workers who invested in Stage 1 and f_u for workers who did not invest. It is assumed that f_q and f_u are continuously differentiable, bounded away from zero and satisfies:

A4 $\frac{f_q(\theta)}{f_u(\theta)} > \frac{f_q(\theta')}{f_u(\theta')}$ if $\theta > \theta'$ [strictly monotone likelihood ratio property]

This assumption implies that qualified workers are more likely to get higher values of θ than unqualified workers. We let F_q and F_u denote the associated cumulative distributions.

Stage 3. There are two firms, $i = 1, 2$. The firms simultaneously announce wage schedules. We allow wages to be dependent on the signal so that a (pure) action of firm i in stage 3 is a measurable function $w_i : [0,1] \to R_+$. We assume that the firms cannot observe the distribution of signals when announcing wages[4].

Stage 4. The workers observe w_1 and w_2 and decide which firm to work for.

Stage 5. In the final stage of the game the firms allocate the available workers by using a task assignment rule which is a measurable function $t_i : [0,1] \to \{0,1\}$[5]. The interpretation is that $t_i(\theta) = 1$ (0) means that firm i assigns all workers with signal θ to the complex (simple) task.

We assume that the risk neutral workers' payoffs are additively separable in money income and the cost of investment and that workers do not care directly in which task they are employed. Thus, once the investment cost is sunk, the worker will rationally choose the firm that offers the higher wage for his particular realization of θ. To save on notation we immediately impose optimal behavior by workers in Stage 4 and write payoffs as

$$E_\theta \left[\max \{ w_1(\theta), w_2(\theta) \} | e \right] - c(e), \tag{2.1}$$

where $c(e_q) = c$ and $c(e_u) = 0$.

Next we want to express the firms' profits as a function of the actions and to do this we need frequency distributions over realized values of the signals. Intuitively one would want to appeal to

[4]The only role of this assumption is that it simplifies the description of the strategy sets. See the discussion in the end of Section 2.3.

[5]This is not the most general way to describe a pure action, but it will be sufficient for our purposes. See footnote 7.

the strong law of large numbers and take these to be given by F_q and F_u, but as noted by Judd [16] and Feldman and Gilles [14] this is problematic with a continuum of random variables. Feldman and Gilles [14] discusses alternative ways to ensure that the individuals' probability distribution and the frequency distribution coincides almost surely. The analysis in this paper relies only on this property and not the particular way we make sure that the property holds. The simplest solution is to use "aggregate shocks" rather than to assume that the signals are i.i.d. draws from F_q and F_u. The investment decisions by the agents induce distributions of qualified workers and unqualified workers on $[\underline{c}, \overline{c}]$. Call these distributions H_q and H_u. Now let the random variable x be uniformly distributed on $[0, 1]$ and let $\theta_c(x)$ denote the test-score for a qualified agent c, where

$$\theta_c(x) = \begin{cases} F_q^{-1}(H_q(c) + x) & \text{if } H_q(c) + x \leq 1 \\ F_q^{-1}(H_q(c) + x - 1) & \text{if } H_q(c) + x > 1 \end{cases} \tag{2.2}$$

It is straightforward, but somewhat tedious to verify that $\Pr[\theta_c(x) \leq \theta \mid e_q] = F_q(\theta)$ for all $c \in [\underline{c}, \overline{c}]$ and all $\theta \in [0, 1]$ and that $\int_{c \in A(x,\theta)} dH_q(c) = F_q(\theta)$ $x \in [0, 1]$ and all $\theta' \in [0, 1]$, where $A(x, \theta) = \{c \in [\underline{c}, \overline{c}] \mid \theta_c(x) > \theta\}$. Clearly the construction can be applied to the unqualified agents as well.

A single firm does not care directly about the realized frequency distributions for the whole population, but rather about the particular workers the firm has available, which depends on the decisions of the workers in Stage 4. Thus, to evaluate the profits of a single firm we need to aggregate the behavior of the workers in some way. In principle, we could derive the distributions for a firm given by arbitrary actions by workers, but we will immediately impose optimal behavior by workers in Stage 4. To capture this we define

$$I^1_{\langle w_1, w_2 \rangle}(\theta) = \begin{cases} 1 & \text{if } w_1(\theta) > w_2(\theta) \\ \frac{1}{2} & \text{if } w_1(\theta) = w_2(\theta) \\ 0 & \text{if } w_1(\theta) < w_2(\theta) \end{cases} \tag{2.3}$$

and let $I^2_{\langle w_1, w_2 \rangle}$ be defined symmetrically. The interpretation is that $I^i_{\langle w_1, w_2 \rangle}(\theta) = 1$ means that all workers with signal θ choose to work for firm i. Besides the fact that the tie-breaking rule is arbitrary[6] these functions aggregates the workers optimal responses to $\langle w_1, w_2 \rangle$ in the obvious way.

Given that a fraction π of the workers invests and wage schedules $\langle w_1, w_2 \rangle$ we can now compute the number of *qualified* workers in firm i with a signal $\theta \leq \overline{\theta}$ as $\int_0^{\overline{\theta}} I^i_{\langle w_1, w_2 \rangle}(\theta) \pi f_q(\theta) d\theta$ and the number of *unqualified* workers can be computed symmetrically. The effective input of labor in the

[6]However, one can show that there are no additional equilibrium outcomes that can be supported by changing the tie-breaking rule.

two tasks given a pair of wage schedules $\langle w_1, w_2\rangle$ and task assignment rule t_i are then given by

$$\begin{aligned} C_i(w_1,w_2,t_i) &= \int \Gamma^i_{\langle w_1,w_2\rangle}(\theta)\, t_i(\theta)\, \pi f_q(\theta)\, d\theta \\ S_i(w_1,w_2,t_i) &= \int \Gamma^i_{\langle w_1,w_2\rangle}(\theta)\,(1-t_i(\theta))\,(\pi f_q(\theta)+(1-\pi) f_u(\theta))\, d\theta \end{aligned} \tag{2.4}$$

respectively. Letting $f_\pi(\theta)$ be shorthand notation for $\pi f_q(\theta)+(1-\pi) f_u(\theta)$ we can express the profits of firm i can then be expressed as

$$\Pi^i(\cdot) = y\left(C_i(w_1,w_2,t_i), S_i(w_1,w_2,t_i)\right) - \int \Gamma^i_{\langle w_1,w_2\rangle}(\theta)\, w_i(\theta)\, f_\pi(\theta)\, d\theta. \tag{2.5}$$

After the workers' decisions in Stage 4 have been replaced by the sequentially rational allocation rule (2.3) a pure strategy for a worker is simply to decide to invest or not. We will summarize the behavior of all workers as a map $i : [\underline{c}, \overline{c}] \to \{e_q, e_u\}$.[7] A pure strategy for a firm is a pair $\langle w_i, \xi_i\rangle$ where w_i is a measurable function from $[0,1]$ into R_+, $\xi_i : M \times M \to T$, M denotes the set of measurable functions from $[0,1]$ into R_+ and T denotes the set of measurable functions from $[0,1]$ into $\{0,1\}$. The interpretation is that if $\xi_i(w_1,w_2)(\theta) = 1$ then firm i assigns workers with signal θ to the complex task given that the pair of wage schedules offered in Stage 3 is (w_1, w_2).

2.3. Characterization of Equilibria

In this section we characterize the set of equilibrium outcomes of the model. Although the model is dynamic, standard refinements such as *perfect Bayesian equilibrium* will not give any sharper predictions than *Nash Equilibrium* in terms of equilibrium *outcomes*[8]. Therefore we take *Nash equilibrium* as our solution concept.

To find the equilibria of the game we will first characterize the firms' equilibrium responses given any investment behavior by the workers. These responses determine a unique wage schedule consistent with any investment behavior by the workers and when we impose optimal behavior by the workers in the initial stage we get a simple fixed point equation that characterizes the set of equilibrium outcomes.

[7]This assumes that all workers with the same investment costs choose the same strategy. More generally one could model a pure strategy profile in analogy with a "distributional strategy" in the sense of Milgrom and Weber[22], i.e. as a joint distribution over $[\underline{c}, \overline{c}] \times \{e_q, e_u\}$. In our model this generality is not necessary since if the best response of agent c is to invest and $c' < c$, then the unique best response of agent c' is to invest.

[8]The reasons why any *Nash equilibrium outcome* can be supported as a *perfect Bayesian equilibrium* will be briefly discussed towards the end of the section.

First we will argue that the wage schedules offered by the firms must be identical almost everywhere in any Nash equilibrium of the game. The reason is simple: If one firm would offer a higher wage than the other to a set of workers with positive mass it could decrease the wage bill by lowering wages slightly for all these workers. If the cut in wages is small enough the firm still has the same distribution of workers available and by keeping the task assignment rule on the outcome path as before the deviation profits would increase[9].

Next we consider the decision problem for the firm in the final stage after a history when an arbitrary fraction of agents $\pi \in (0,1]$[10] has chosen to invest and the firms have offered wage schedules $\langle w_1, w_2 \rangle$ with $w_1(\theta) = w_2(\theta)$ for almost all $\theta \in [0,1]$. For an arbitrary θ, the quantity of qualified workers available for firm i with realized signal less than θ is then simply $\frac{1}{2}F_q(\theta)$ and the quantity of unqualified workers with signal less than θ is symmetrically $\frac{1}{2}F_u(\theta)$. By the monotone likelihood ratio assumption any optimal task assignment rule for firm i must be a cutoff-rule[11] of the form

$$t_i(\theta) = \begin{cases} 0 & \text{if } \theta < \widetilde{\theta}_i \\ 1 & \text{if } \theta \geq \widetilde{\theta}_i \end{cases}. \tag{2.6}$$

Using the identities in (2.4) and observing that the firms can do nothing about their wage costs in the final stage of the game the problem to maximize output can be written as

$$\begin{aligned} & \max_{\theta_i, C_i, S_i} y(C_i, S_i) \qquad (2.7) \\ \text{subj. to } 2C_i & \leq \pi(1 - F_q(\theta_i)) \\ 2S_i & \leq \pi F_q(\theta_i) + (1-\pi) F_u(\theta_i). \end{aligned}$$

Eliminating θ_i from the problem it is easy to show that the monotone likelihood ratio implies that the constraint set is convex[12] . Furthermore, y is strictly increasing in both arguments so the constraint must bind with equality and strict quasi-concavity of y guarantees that there is a unique solution to (2.7). Finally, the boundary condition **A2** guarantees that any solution must be interior, so the problem can be depicted graphically as in Figure 2.1.

Since both firms face a symmetric problem with a unique solution we drop the indices from now on. Eliminating C and S from the problem and using constant returns to scale we can after

[9] See Lemma 12 in the appendix for a more formal argument.

[10] Given "worker strategy profile" i this is computed as $\pi = \int_{c \in i^{-1}(\{e_q\})} dG(c)$.

[11] See Lemma 13.

[12] See Lemma 14.

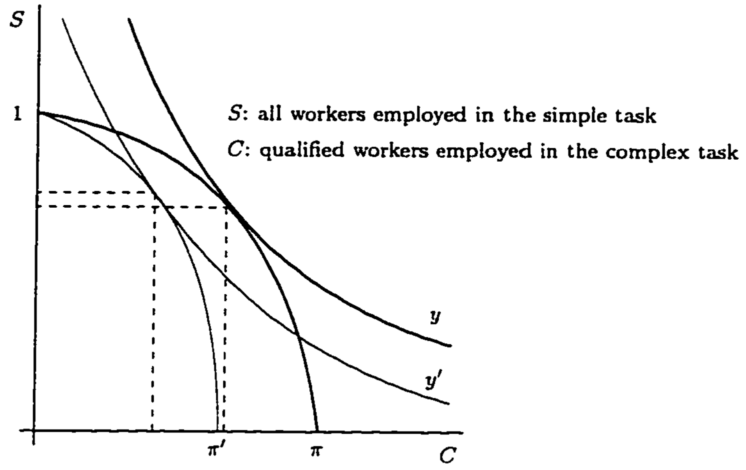

Figure 2.1: The task-assignment problem

some algebra write the first order condition for (2.7) as

$$p(\theta,\pi)\, y_1(\pi(1-F_q(\theta)), F_\pi(\theta)) = y_2(\pi(1-F_q(\theta)), F_\pi(\theta)), \tag{2.8}$$

where $F_\pi(\theta)$ is shorthand notation for $\pi F_q(\theta) + (1-\pi) F_u(\theta)$ and

$$p(\theta,\pi) = \frac{\pi f_q(\theta)}{\pi f_q(\theta) + (1-\pi) f_u(\theta)} \equiv \Pr[e_q|\theta] \tag{2.9}$$

denotes the posterior probability that a randomly drawn agent with test score θ is *qualified* given prior probability π. The economic interpretation is that an agent with signal equal to the optimal cutoff point determined by (2.8) has the same expected marginal productivity in both tasks. All agents with a lower realization of the signal are more productive in the simple task and agents with higher signals are more productive in the complex task.

For each $\pi > 0$ we let the unique solution to (2.8) be denoted by $\widetilde{\theta}(\pi)$. It is shown in Appendix 2.7.3 that $\widetilde{\theta}$ is continuously differentiable and therefore continuous on the open unit interval. To save on notation we also define

$$r(\pi) = \frac{\pi\left(1 - F_q\left(\widetilde{\theta}(\pi)\right)\right)}{\pi F_q\left(\widetilde{\theta}(\pi)\right) + (1-\pi) F_u\left(\widetilde{\theta}(\pi)\right)}, \tag{2.10}$$

which, for an arbitrary π is the ratio of effective units of complex labor over units of simple labor implied by the equilibrium task assignment rule (2.6).

Our next task is to determine the equilibrium wage schedules. The most natural guess is that all agents are paid according to their respective expected marginal productivity in the task where they are employed. Thus, given a fraction π of investors we take the candidate "labor market equilibrium" wage function to be given by $w : [0, 1] \to R_+$ defined as

$$w(\theta) = \begin{cases} y_2(r(\pi), 1) & \text{for } \theta < \widetilde{\theta}(\pi) \\ p(\theta, \pi) y_1(r(\pi), 1) & \text{for } \theta \geq \widetilde{\theta}(\pi) \end{cases}, \tag{2.11}$$

where we have used the assumption of constant returns to scale[13]. Observe that the assumption of (strict) monotone likelihood ratio implies that $p(\theta, \pi)$ is strictly increasing in θ given any $\pi > 0$. Hence the proposed wage schedule is strictly increasing on $\left[\widetilde{\theta}(\pi), 1\right]$.

Indeed, it can be shown that the intuition from Bertrand competition with constant returns to scale carries over to our model. Formally:

Proposition 4. *Let the fraction of agents who invest be given by π, let $t : [0, 1] \to \{0, 1\}$ be a cutoff rule with critical point $\widetilde{\theta}(\pi)$ determined by (2.8) and let $w : [0, 1] \to R$ be given by (2.11). Furthermore, for any firm strategy profile $\langle w_i, \xi_i \rangle_{i=1,2}$ let $t_i : [0, 1] \to \{0, 1\}$ be the task assignment rule on the outcome path, i.e. $t_i(\theta) = \xi_i(w_i, w_j)(\theta)$ for all θ. Then both firms are playing best responses if and only if $w_i(\theta) = w(\theta)$ and $t_i(\theta) = t(\theta)$ for $i = 1, 2$ and for all almost all $\theta \in [0, 1]$.*

The proof is in the appendix i Section 2.7.1. The intuition for the sufficiency part is that any deviating firm has to pay at least the expected marginal productivity (given task assignments according to candidate equilibrium) for all workers and strictly more if it wants to attract any additional workers. The only way this could be a profitable deviation in a constant returns to scale environment is if the deviating firm could allocate the workers more efficiently between tasks, which is impossible since the original task assignment rule maximizes output. Deviations on sets of measure zero will clearly have no effect on profits and it has been argued above that both firms must choose task assignment rules identical to t almost everywhere.

So far we have considered the firms' equilibrium responses for any fixed investment behavior by the workers. In a *Nash Equilibrium* of the full model the additional condition that each worker maximizes (2.1) given the wage schedules must hold as well. The workers only care about the *maximal* offer for each realization of θ , so we let $w'(\theta) = \max\{w_1'(\theta), w_2'(\theta)\}$ for any pair of

[13]I.e. $y_i\left(\pi\left(1 - F_q\left(\widetilde{\theta}(\pi)\right)\right), F_\pi\left(\widetilde{\theta}(\pi)\right)\right) = y_i(r(\pi), 1)$

wage schedules $\langle w_1', w_2' \rangle$ and write the set of pure best responses for agent $c \in [\underline{c}, \bar{c}]$ as

$$\beta_c(w') = \begin{cases} e_q & \text{if } \int w'(\theta) f_q(\theta) d\theta - c > \int w'(\theta) f_u(\theta) d\theta \\ \{e_q, e_u\} & \text{if } \int w'(\theta) f_q(\theta) d\theta - c = \int w'(\theta) f_u(\theta) d\theta \\ e_u & \text{if } \int w'(\theta) f_q(\theta) d\theta - c < \int w'(\theta) f_u(\theta) d\theta \end{cases} . \tag{2.12}$$

The unique fraction of investors consistent with all workers playing best responses is thus given by

$$\pi = G\left(\int w'(\theta) f_q(\theta) d\theta - \int w'(\theta) f_u(\theta) d\theta\right). \tag{2.13}$$

Since Proposition 4 guarantees that wages must be given by marginal productivities in any Nash equilibrium of the full model we can substitute the wage schedule (2.11) into (2.13) to obtain a fixed point equation in π. We denote by $H(\pi)$ the gross benefits of investment, i.e. the difference between the expected earnings for an agent who invests and the expected earnings for an agent who does not invest. By simple substitution we find that $H(\pi)$ can be written as

$$y_2(r(\pi), 1)\left(F_q\left(\widetilde{\theta}(\pi)\right) - F_u\left(\widetilde{\theta}(\pi)\right)\right) + y_1(r(\pi), 1)\int_{\widetilde{\theta}(\pi)}^{1} p(\theta, \pi)(f_q(\theta) - f_u(\theta))\, d\theta. \tag{2.14}$$

Since $y_1(r(\pi), 1)\, p\left(\widetilde{\theta}(\pi), \pi\right) = y_2(r(\pi), 1)$ it is easy to see that $H(\pi) > 0$ for all π in the interior of the unit interval. As discussed above optimal behavior of the workers implies that an agent should invest if and only if the cost of doing so is less than the expected benefits. The equilibria of the model are thus fully characterized by the solutions to the equation

$$\pi = G(H(\pi)) \tag{2.15}$$

Summing up these observations we have:

Proposition 5. *Consider a strategy profile* $\left\{i, \langle w_i, \xi_i \rangle_{i=1,2}\right\}$ *and let* π^* *be the fraction of investors implied by worker strategy profile* i[14]. *Furthermore, let* w, t *and* t_i *be defined as in Proposition 4 (with* $\pi = \pi^*$*). Then* $\left\{i, \langle w_i, \xi_i \rangle_{i=1,2}\right\}$ *is a Nash equilibrium if and only if* π^* *solves (2.15),* $i(c) = e_q$ *for all* $c < G^{-1}(\pi^*)$ *and* $i(c) = e_u$ *for all* $c > G^{-1}(\pi^*)$ *and* $(w_i(\theta), t_i(\theta)) = (w(\theta), t(\theta))$ *for* $i = 1, 2$ *and for almost all* $\theta \in [0, 1]$.

Note that Proposition 5 implies that the question of existence of equilibria reduces to the question of existence of a fixed point of the map $G \circ H$. This gives us a relatively easy proof of existence of equilibria, which is the next result:

[14]That is, $\pi^* = \int_{c \in i^{-1}(\{e_q\})} dG(c)$.

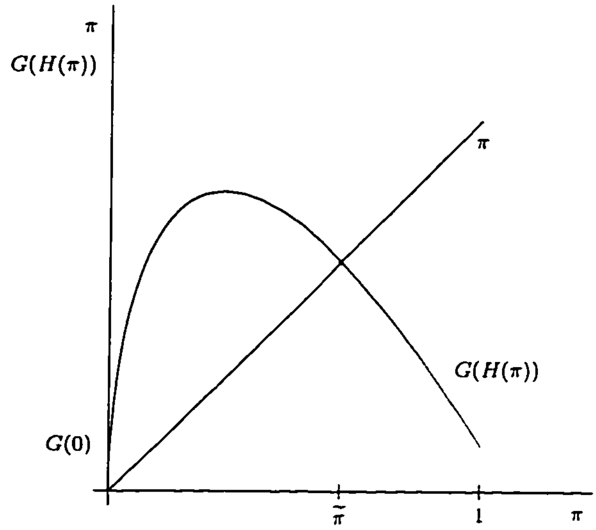

Figure 2.2: An example with a unique interior equilibrium

Proposition 6. *If* $G(0) > 0$[15]*, then there exists a non-trivial equilibrium of the model.*

The proof is the appendix in Section 2.7.2, but the idea can be understood from Figure 2.2. While (2.8) is not defined at $\pi = 0$ it follows directly from the constraint that the only *feasible* input of complex labor is zero when nobody invests. Since by assumption **A3** both tasks are needed in production, output must be zero. Hence $w(\theta) = 0$ for all θ which implies $H(0) = 0$. When $\pi = 1$ we have that $p(\theta, 1) = 1$ for all θ and since this means that the signal does not provide any additional information the wage schedule is a constant function of θ, so $H(1) = 0$. As has already been argued $H(\pi) > 0$ for all intermediate values of π and after verifying that $H(\pi)$ is continuous on $[0, 1]$ existence is established by use of the intermediate value theorem.

2.3.1. Why Use a Strategic Model?

In our model, it is very important that workers respond to wage schedules when choosing what firm to work for. We capture this by the assumption that workers are allocated between firms in

[15]Under the additional assumption that $\lim_{C \to 0} \frac{y_{11}(C,S)y(C,S)}{(y_1(C,S))^2} > -\infty$ the result holds even when $G(0) = 0$ or, equivalently, when the lower bound on the support of the cost distribution $\underline{c} = 0$. In this case the model also has a trivial equilibrium where nobody invests, which is not the case when $\underline{c} < 0$. While not easy to interpret, the condition holds for several common parametric production functions (for example in the Cobb-Douglas case). The proof, in which it is shown that the slope of $G \circ H$ is unbouded at $\pi = 0$ is available on request from the authors.

accordance to (2.3). This makes the model extremely "competitive" in a strategic sense and, as we have seen, the equilibrium conditions have an obvious flavor of competitive equilibrium. One may therefore conjecture that the model could be formalized as a model with price taking agents, but doing this one runs into technical as well as conceptual difficulties[16]. The game-theoretic modelling helps overcome these difficulties and also makes the policy analysis easier to handle.

Note that the "reduced game" obtained by assuming that each firm's distribution of available workers is determined by (2.3) is still a dynamic game and we have nevertheless been able to ignore unreached information sets: propositions 1 and 2 do not even specify what firms are supposed to do when choosing task assignment rules at information sets where the wage schedules differ on a set of points with positive measure. The reader may therefore be worried that we consider equilibria supported by non-credible threats off the equilibrium path, but this is not the case. The intuitive explanation is that the last stage of the game is non strategic, in the sense that the task assignments by the other firm has no impact on the best responses in Stage 3. Hence it is impossible to enlarge the set of equilibrium *outcomes* compared to the set of *perfect Bayesian equilibrium outcomes* by committing to task assignment rules that are suboptimal off the equilibrium path[17].

In the discussion above we derived the fixed point equation (2.15) by working from the end of the game as if we were using backwards induction. However, since we are working with a continuum of workers no single worker can affect aggregate variables and the function (3.46) therefore plays the same role in the analysis as the best response correspondence in a static game.

2.4. The Model with Two Identifiable Groups of Workers

We now extend the basic model and assume that each worker belongs to one of two identifiable groups, indexed by a and b respectively. The purpose of the section is to demonstrate the existence of equilibria with discrimination. Under the assumption that not too many agents will invest when

[16]The problem is that when maximizing over "quantities" the decision variable of the firm is to choose a distribution on the support of the noisy signal. In order to write down sensible market clearing conditions it turns out that a strong law of large numbers is needed. The technical problem is that to guarantee such a strong law of large numbers in an environment with uncountably many independent random variables one has to rely on somewhat arbitrary probability measures (see Judd [16] and Feldman and Gillles [14]). While we also have to deal with this problem in our model it is much easier to circumvent in our framework. The conceptual problem is that even with such a strong law of large numbers it is not clear how the firms should evaluate profits out of equilibrium.

[17]Since it would not affect the set of equilibrium outcomes the reader may wonder why we did not model the wage and task assignment decisions as simultaneous. The reason is that when introducing the affirmative action policy we need task assignments to be done after the firm knows what distribution of workers it has available.

there are no monetary incentives, we show existence of discriminatory equilibria by construction.

2.4.1. The Extended Model

We now assume that a fraction λ^a of the workers belongs to group a and a fraction $\lambda^b = 1 - \lambda^a$ to group b. It is assumed that the distribution of investment costs is given by G in each group and that the probability density over signals is given by f_q for any worker (from group a or b) who invests and by f_u for any worker who does not. These two assumptions means that the groups are *ex ante identical* in terms of investment costs and that the signals are *unbiased.*

As in the single group model we assume that the realized frequency distributions of signals coincides with the probability distributions F_q and F_u. This can be *derived* using the obvious generalization of the exact stochastic model described in Section 2.2.1.

The game is the same as in Section 2.2.1, except that wage schedules are now allowed to depend on "group identity". Hence, a strategy for firm i is a quadruple $\langle w_i^a, w_i^b, \xi_i^a, \xi_i^b \rangle$ where $w_i^j : [0,1] \rightarrow R_+$ is the wage schedule and ξ_i^j maps pairs of wage schedules to task assignment rules for group j. For the same reason as in the single group model, the specification of the task assignment rule off the equilibrium path will be irrelevant and we can think of the firm as choosing a pair of wage schedules and task assignment rules $t_i^j : [0,1] \rightarrow \{0,1\}$.

We maintain the assumption that workers from each group allocate themselves between firms according to (2.3), now evaluated using the relevant wage schedules for each group. Hence if the fractions of agents who invest are given by $\pi = (\pi^a, \pi^b)$, the effective input of labor in respective task in firm i is given by

$$\begin{aligned} C_i(\cdot;\pi) &= \sum_{j=a,b} \lambda^j \int I^i_{\langle w_1^j, w_2^j \rangle}(\theta)\, t_i^j(\theta)\, \pi^j f_q(\theta)\, d\theta \\ S_i(\cdot;\pi) &= \sum_{j=a,b} \lambda^j \int I^i_{\langle w_1^j, w_2^j \rangle}(\theta) \left(1 - t_i^j(\theta)\right) \left(\pi^j f_q(\theta) + \left(1-\pi^j\right) f_u(\theta)\right) d\theta \end{aligned} \tag{2.16}$$

It should be clear how the firms' profit functions should be generalized from (2.5).

2.4.2. Equilibrium in the Extended Model

Analogous to the procedure in Section 2.3 we begin by considering the problem of maximizing output over task-assignment rules with the cutoff-property. This problem can be written as;

$$\max_{(\theta^a,\theta^b)\in[0,1]^2} y\left(\sum_{j=a,b} \lambda^j \pi^j \left(1 - F_q\left(\theta^j\right)\right), \sum_{j=a,b} \lambda^j F_{\pi^j}\left(\theta^j\right)\right). \tag{2.17}$$

The program is qualitatively very similar to (2.7), the task assignment problem in the basic model. but (partial) corner solutions may now be possible. By similar arguments as in the single group model one shows that there exists a unique solution to (2.17) for any $\pi \neq (0,0)$ and that the solution satisfies the Kuhn-Tucker conditions. We let γ_j be the multiplier associated with the constraint $\theta^j \geq 0$ and η_j be associated with $1-\theta^j \geq 0$. The Kuhn-Tucker conditions are, after some rearranging, given by

$$-p\left(\theta^j,\pi^j\right) y_1(\cdot) + y_2(\cdot) + \frac{\gamma_j - \eta_j}{f_{\pi^j}\left(\theta^j\right)} = 0, \text{ for } j = a,b \tag{2.18}$$

together with the complementary slackness conditions. We let the solution be denoted by $\widetilde{\theta}(\pi) = \left(\widetilde{\theta}^a(\pi), \widetilde{\theta}^b(\pi)\right)$. As in the basic model, continuity of $\widetilde{\theta}$ follows from the implicit function theorem. To economize on notation we will let $\widetilde{r}(\pi)$ denote the factor ratio implied by $\widetilde{\theta}(\pi)$, that is

$$\widetilde{r}(\pi) = \frac{\sum_{j=a,b} \lambda^j \pi^j \left(1 - F_q\left(\widetilde{\theta}(\pi)\right)\right)}{\sum_{j=a,b} \lambda^j \left(\pi^j F_q\left(\widetilde{\theta}(\pi)\right) + (1-\pi^j) F_u\left(\widetilde{\theta}(\pi)\right)\right)} \tag{2.19}$$

As in the basic model each worker is paid according to his expected marginal productivity in equilibrium, and by the assumption of constant returns the wage schedules can be written as:

$$w^j(\theta) = \begin{cases} y_2\left(\widetilde{r}(\pi), 1\right) & \text{for } \theta < \widetilde{\theta}^j(\pi) \\ p\left(\theta,\pi^j\right) y_1\left(\widetilde{r}(\pi), 1\right) & \text{for } \theta \geq \widetilde{\theta}^j(\pi) \end{cases} \tag{2.20}$$

Finally, the fraction of agents in group j who invest can be found in the same way as before. Given a wage schedule w^j for group j the fraction of agents who optimally decides to invest is given by (2.13). The gross benefits of investment given any investment behavior π is

$$\begin{aligned} B^j(\pi) &= y_2\left(\widetilde{r}(\pi),1\right)\left(F_q\left(\widetilde{\theta}^j(\pi)\right) - F_u\left(\widetilde{\theta}^j(\pi)\right)\right) + \\ &\quad + y_1\left(\widetilde{r}(\pi),1\right) \int_{\widetilde{\theta}^j(\pi)}^{1} p\left(\theta,\pi^j\right)\left(f_q(\theta) - f_u(\theta)\right) d\theta \end{aligned} \tag{2.21}$$

and the relevant system of fixed-point equations is

$$\pi^j = G\left(B^j(\pi)\right), \text{ for } j = a,b. \tag{2.22}$$

The characterization results from the single group model generalize in a straightforward way, so the equilibrium set will be fully characterized as the solutions to (2.22). For expositional convenience we will therefore work directly with the reduced form equations. Given any solution to (2.22) we can always use (2.17) and (2.20) to construct the implied equilibrium wage schedules and task assignment rules.

We will say that an equilibrium is *discriminatory* if $\widetilde{\theta}^a(\pi) \neq \widetilde{\theta}^b(\pi)$ or $w^a \neq w^b$ and *non-discriminatory* otherwise. Here it is important to realize that $\widetilde{\theta}^j(\pi)$ as well as the wage schedule for group j depends on investment behavior in both groups. The reason is that the fraction of investors in the other group affects how scarce a resource qualified workers are and firms will therefore take investment behavior in both groups into consideration when deciding on the task assignments for any of the groups. This implies that the fraction of investors in the other group affects the benefits of investors, both by the effect on the cutoff signal and by affecting the factor ratio. In fact one can show that an increase in the fraction of investors in the other group monotonically decreases the incentives to invest, so investment in the two groups are "aggregate strategic substitutes". As a consequence of these interdependencies the set of equilibria of the extended model will **not** be the set of possible permutations of the equilibria of the single group model.

The set of *non-discriminatory* equilibria corresponds one to one with the set of equilibria of the single group model: the equilibrium conditions of the extended model reduces to the equilibrium conditions of the single group model when it is imposed that both groups are treated symmetrically. Combining this simple observation with Proposition 6, it follows that there exists at least one non-trivial non-discriminatory equilibrium in the extended model.

We are mainly interested in discriminatory equilibria and as the next proposition shows at least one such equilibrium will exist under the assumption that not too many workers derive positive utility from investment in human capital.

Proposition 7. *Let y be a given production function and let f_q, f_u be some fixed densities, where y, f_q and f_u satisfies the assumptions stated in Section 2.2.1 and let $(\lambda^a, \lambda^b) \in int(\Delta^2)$. Then, there exists $\overline{G}_0 > 0$ such that if $G(0) \leq \overline{G}_0$, then there exists an equilibrium where no workers from one of the groups are assigned to the complex task and a positive fraction of the agents from the other group are assigned to the complex task. Moreover, in this equilibrium the wage schedule for the group where all workers are assigned to the simple task is uniformly below the wage schedule for the other group.*

The construction is in the appendix in Section 2.7.3. Assuming that all agents in, say, group a, are assigned to the simple task the equilibrium conditions for the other group are qualitatively as in the single group model. Hence, applying the same steps as in the proof of Proposition 6 we have that there is an equilibrium where a fraction $\pi^b > G(0)$ of the agents in the other group invests and a positive fraction of these are assigned to the complex task, *assuming* that no agents from the discriminated group are assigned to the complex task. Let the implied cutoff for group b be given by $\tilde{\theta}^b$. In order to check that this is an equilibrium of the model with two groups we just have to check that $\left(1, \tilde{\theta}^b\right)$satisfies the Kuhn-Tucker conditions for the problem (2.17) when π^b is given as above and $\pi^a = G(0)$. This is indeed the case given that a sufficiently small fraction of agents have negative costs of investment.

Observe that for $\theta^a = 1$ to satisfy (2.18) it must be that $p(1, \pi^a) \leq y_2(\cdot)/y_1(\cdot)$. Since an increase in λ^a decreases the factor ratio the right hand of this inequality is decreasing in λ^a and it follows that $\overline{G}_0$ is strictly decreasing in λ^a. Thus, the larger the group is the more difficult it is to sustain this extreme form of discrimination against its members.

An alternative sufficient condition for existence of discriminatory equilibria is existence of multiple equilibria in the single group model. As a general property of discriminatory equilibria, it can be shown that there exists *some* equilibrium in the single group model such that any agent is better off than an agent in the disadvantaged group with the same investment costs c. The discriminated group is thus always better off in *some* "autarchic equilibrium".

2.5. Affirmative Action

In this section we will use our framework to analyze the effects of affirmative action, which we model as a quota forcing the employers to fulfill certain requirements on the representation of workers from the disadvantaged group in both tasks.

A more natural intervention would perhaps be an equal opportunities law requiring firms to offer wages that do not depend on group identity. In our simple framework this would mean that the firms would be constrained to offer identical wage schedules to both groups. Since the incentives to invest would be the same for both groups this would eliminate discrimination in our model.

The problem with this type of equal opportunity law is that the regulator must observe all information the employer has available in order to implement such a policy. In a more realistic setup where there are other variables than a one-dimensional signal, this type of policy would also be possible to evade by using other variables as proxies for group identity. Moreover, in reality hiring decisions are based on several factors that may or may not be an indicator of the expected productivity of the worker. In particular if there are other variables correlated both with group identity and intrinsic productivity it may be impossible for the regulatory authorities to disentangle what part of the correlation is "real" and what is due to statistical discrimination[18].

Also observe that if the principle of "equal pay for equal work" is interpreted to mean that the average wage for workers *performing a particular task* cannot differ across groups then the discriminatory equilibrium constructed in the proof of Proposition 9 satisfies this principle. All workers in the simple task are paid the same wage and no workers in the discriminated group are assigned to the complex task, so "wage equality" in what seems to us to be the standard operational sense holds.

2.5.1. The Model with Affirmative Action

For simplicity we will model affirmative action as a requirement for each firm to hire workers for each task in accordance with the population fractions[19]. The quantity of workers (qualified and unqualified) from group j employed in the complex task by firm i given that a fraction π^j has invested and the actions $A = \langle w_i^a, w_i^b, t_i^a, t_i^b \rangle_{i=1,2}$ is given by

$$\Psi_i^j(A) = \int I^1_{\left\langle w_i^j, w_{i'}^j \right\rangle}(\theta)\, t_i^j(\theta)\, \lambda^j f_{\pi^j}(\theta)\, d\theta \tag{2.23}$$

There is no distinction between the quantity of $j-$workers and the input of labor in the simple task. This quantity, $S_i^j(A)$, is therefore computed according to equation (2.16) in Section 2.4 and the affirmative action requirement is

$$\frac{\Psi_i^a(A)}{\Psi_i^b(A)} = \frac{\lambda^a}{\lambda^b} \text{ and } \frac{S_i^a(A)}{S_i^b(A)} = \frac{\lambda^a}{\lambda^b} \tag{2.24}$$

The payoffs as functions of actions and the timing of the actions are as before and the only difference compared to the model in Section 2.4 is that the task assignment rule chosen in the final stage of the game must satisfy the affirmative action constraint[20]. Since we think of the affirmative action policy as a *constraint* on the available actions we should in principle allow the task assignments to be contingent on π^a and π^b and adjust the strategy sets accordingly. However, for the same reasons

[18] See Lundberg [18] for a discussion on these issues and some interesting suggestions on statistical procedures the regulator could use in order to implement equal opportunities laws when there is asymmetric information between the firms and the regulator

[19] We can handle quotas with other numerical goals, but as we will discuss later it is important to have a quota for both tasks.

[20] One could alternatively keep the strategy sets as before and impose affirmative action by *charging penalties* to any firm that violates the numerical goals on employment stipulated by the policymaker. If the penalty for a violation is sufficiently costly the two approaches are equivalent.

as earlier we can focus on *Nash equilibria* without any risk of analyzing equilibria supported by non credible threats. In particular, since no worker can affect the fraction of investors it is immaterial if we view the "equilibrium responses" as dynamic reactions to the behavior of the workers or fictitious best responses.

As in the single group model the first step in the equilibrium characterization is to note that both firms must offer wage schedules that are identical almost everywhere in any equilibrium of the model. By the monotone likelihood ratio property, the task assignment rules must be a pair of cutoff rules. Using this fact we can characterize the optimal task assignment rule after any history where both firms have offered (essentially) the same wage schedules by solving the problem

$$\begin{array}{ll} \max_{\theta^a,\theta^b} y\left(\sum_{j=a,b} \lambda^j \pi^j \left(1 - F_q\left(\theta^j\right)\right), \sum_{j=a,b} \lambda^j \left(\pi^j F_q\left(\theta^j\right) + \left(1-\pi^j\right) F_u\left(\theta^j\right)\right)\right) & . \\ \text{s.t} \quad \pi^a F_q\left(\theta^a\right) + \left(1-\pi^a\right) F_u\left(\theta^a\right) = \pi^b F_q\left(\theta^b\right) + \left(1-\pi^b\right) F_u\left(\theta^b\right) & \end{array} \tag{2.25}$$

This problem is just adding a constraint to the task assignment problem (2.17) in Section 2.4. As is easily verified this constraint is the affirmative action requirement (2.24) for the special case when the wage schedules are the same and the task-assignment rules are taken to have the cutoff property, which as we have argued must be properties of equilibrium.

The first-order conditions[21] for this problem can after some rearranging be written as[22];

$$\begin{array}{c} -y_1(\cdot)\, p\left(\theta^a, \pi^a\right) + y_2(\cdot) - \frac{\mu}{\lambda^a} = 0 \\ -y_1(\cdot)\, p\left(\theta^b, \pi^b\right) + y_2(\cdot) + \frac{\mu}{\lambda^b} = 0 \end{array} \tag{2.26}$$

where $\mu > 0$ if $\pi^a < \pi^b$ (see Footnote 22). Using the constraint, the multiplier and one of the decision variables can be eliminated and the remaining equation has all the qualitative properties of (2.8). By arguments more or less identical to the ones used in the single group model one can show that for each $\pi = (\pi^a, \pi^b)$ such that either π^a or π^b is strictly positive there is a unique $\widehat{\theta}(\pi) = \left(\widehat{\theta}^a(\pi^a,\pi^b), \widehat{\theta}^b(\pi^a,\pi^b)\right) >> 0$ that solves (2.25) and that the implicit function theorem applies[23]. The solution will consequently be a smooth function of π. To write things down more compactly below will introduce one additional piece of notation. For each π, we denote by $\widehat{r}(\pi)$

[21]It is straightforward to show that the first-order conditions are sufficient and that there is a unique solution to the program (2.25).

[22]It is useful to note that the solution(s) to the problem with the affirmative action restriction formulated as an equality constraint must also be solution(s) to the problem with the same objective and the inequality constraint $\pi^a F_q(\theta^a) + (1-\pi^a) F_u(\theta^a) \leq \pi^b F_q(\theta^b) + (1-\pi^b) F_u(\theta^b)$ if $\pi^a < \pi^b$ and the reverse inequality if $\pi^a > \pi^b$. Hence, if the multiplier is taken to be positive it enters with a negative sign for the group with the smaller number of investors and a positive sign for the other group.

[23]The proofs follows the proofs of Lemma I and Lemma II in appendix B step by step.

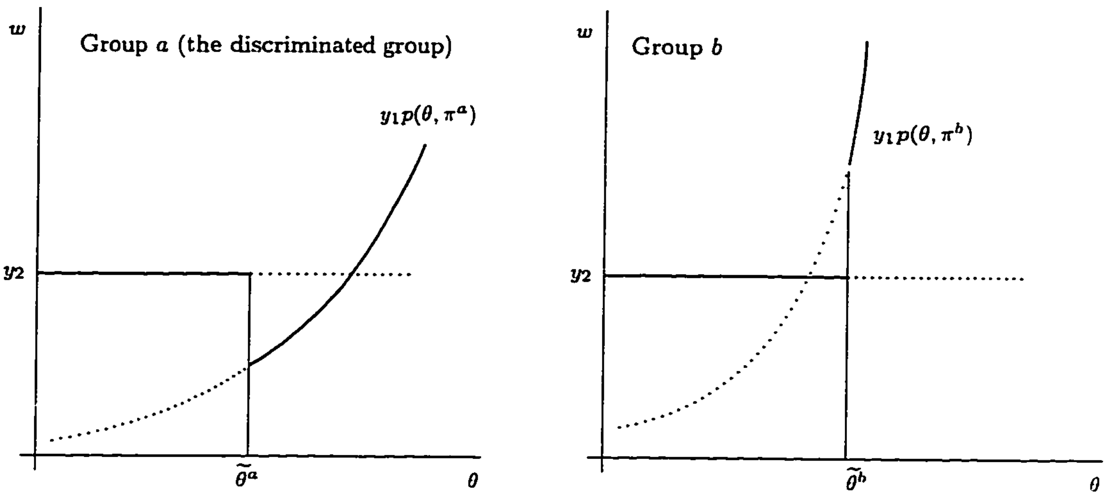

Figure 2.3: Not an equilibrium

the unique factor ratio implied by the firms optimal choice of task assignment rules, that is

$$\hat{r}(\pi) = \frac{\sum_{j=a,b} \lambda^j \pi^j \left(1 - F_q\left(\hat{\theta}^j(\pi)\right)\right)}{\sum_{j=a,b} \lambda^j \left(\pi^j F_q\left(\hat{\theta}^j(\pi)\right) + (1-\pi^j) F_u\left(\hat{\theta}(\pi)\right)\right)}. \tag{2.27}$$

While the characterization of equilibrium task-assignment rules is not significantly harder than in the single group model the determination of wages is somewhat counter-intuitive. It is tempting to guess that wages still are given by expected marginal productivities, that is to take (2.20) as the candidate equilibrium wage function, using the unique cut-off points determined above. This is however not consistent with equilibrium since (assuming $\pi^a \neq \pi^b$) some agents of the discriminated group employed in the complex tasks would be paid less than other agents from the same group who are in the simple task (see Figure 2.3). Hence a firm could deviate and attract all these workers and replace some of the workers previously in the simple task by the additional workers the firm attracts. If the affirmative action constraint was satisfied prior to the deviation it will be satisfied after the deviation as well and output is unchanged. Since the wage bill has decreased this is a profitable deviation.

The unique wage schedules consistent with equilibrium are depicted in Figure 2.4, which is drawn under the assumption that $\pi^a < \pi^b$. As can be seen from the graph the wage in the simple task is now determined by the marginal agent's productivity in the complex task rather than the

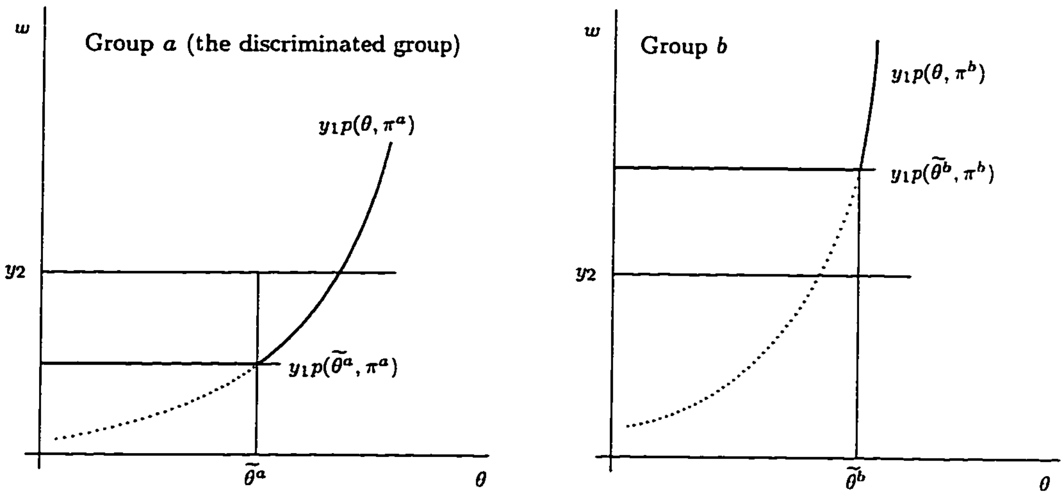

Figure 2.4: Equilibrium wage schedules under affirmative action

productivity in the simple task. Intuitively, there is no incentive to deviate in order to change the allocation of workers in the complex task for any of the firms: all workers already employed in the complex task are paid their expected marginal productivities and all workers employed in the simple task are paid more than their expected marginal productivity would be if in the complex task. Furthermore, by the affirmative action constraint the firms cannot change the ratio of $a-$workers to $b-$workers in the simple task. This means that what is left to show in order to demonstrate that the proposed wage schedules are consistent with equilibrium is that the *average wage* in the simple task equals the marginal productivity. To realize this it is useful to note that by combining the two equations in (2.26) we see that for the optimal choice of $\widehat{\theta}$ the weighted average of the expected marginal productivities in the complex task for the critical agents' in respective group equals the marginal productivity in the simple task. That is, eliminating the multiplier from (2.26) we have

$$y_1\left(\widehat{r}\left(\pi\right),1\right)\sum_{j=a,b}\lambda^j p\left(\widehat{\theta}^j\left(\pi\right),\pi^j\right)=y_2\left(\widehat{r}\left(\pi\right),1\right) \tag{2.28}$$

At this point it is simply to note that $F_{\pi^a}\left(\widehat{\theta}^a\left(\pi\right)\right) = F_{\pi^b}\left(\widehat{\theta}^b\left(\pi\right)\right)$ by the affirmative action constraint, so the left hand side of (2.28) is also the average wage in the simple task, which gives the result.

In terms of the notation introduced above we can write the proposed labor market equilibrium wage schedules depicted in Figure 2.4 as

$$w^j(\theta) = \begin{cases} y_1(\widehat{r}(\pi), 1)\, p\left(\widehat{\theta}^j(\pi), \pi^j\right) & \text{for } \theta < \widehat{\theta}^j(\pi) \\ y_1(\widehat{r}(\pi), 1)\, p(\theta, \pi^j) & \text{for } \theta \geq \widehat{\theta}^j(\pi) \end{cases} . \tag{2.29}$$

for $j = a, b$. Note that if $\pi^a < \pi^b$ then $\widehat{\theta}^a(\pi) < \widehat{\theta}^b(\pi)$, since otherwise the affirmative action constraint can not be satisfied. Hence $p\left(\widehat{\theta}^a(\pi), \pi^a\right) < p\left(\widehat{\theta}^b(\pi), \pi^b\right)$ and by inspection of (2.26) we see that $w^a(\theta) < y_2(\widehat{r}(\pi), 1)$ for $\theta < \widehat{\theta}^a(\pi)$ and $w^b(\theta) > y_2(\widehat{r}(\pi), 1)$ for $\theta < \widehat{\theta}^b(\pi)$. That is, workers in the simple task from group a are paid less than their marginal productivity in the task and workers from group b are paid more.

Summing up the discussion above as a proposition:

Proposition 8. *Let the fractions of agents who invest in each group be given by $\pi = (\pi^a, \pi^b)$ and let $\widehat{\theta}(\pi) = \left(\widehat{\theta}^a(\pi), \widehat{\theta}^b(\pi)\right)$ be the unique solution to (2.26). Furthermore, for $j = a, b$ let $w^j(\cdot)$ be given by (2.29) and $t^j(\cdot)$ be the cutoff task assignment rule with critical value $\widehat{\theta}^j(\pi)$. Finally, for an arbitrary firm strategy profile $\langle w_i^a, w_i^b, \xi_i^a, \xi_i^b \rangle_{i=a,b}$ let t_i^a and t_i^b be the task assignment rules on the outcome path for firm $i = 1, 2$. Then, both firms are playing best responses if and only if $w_i^j(\theta) = w^j(\theta)$ and $t_i^j(\theta) = t^j(\theta)$ for $i = 1, 2$ and $j = a, b$ and for almost all $\theta \in [0, 1]$.*

The proof in Section 2.7.4 fills in some of the details missing in the paragraphs above

One may believe that when we impose affirmative action as a constraint in both tasks, one constraint is really redundant. The reason for this would be that if the affirmative action constraint is satisfied in one of the tasks and if the market clears, which must be the case in equilibrium, then the affirmative action constraint must be satisfied in the other task as well. This is indeed true and it is also true that in order to characterize the equilibrium task assignment rules we need only one of the constraints. The problem is that if there is affirmative action in the complex task only and if the groups behave differently there is no "labor market equilibrium" in the continuation game. To see this it useful to consider Figure 2.4, where it is easy to see that any deviation where a firm reduces the number of workers from group b is profitable. Hence, since only non-discriminatory (Nash) equilibria remains in the full game one could in principle interpret this as saying that affirmative action works. However, we think that this is taking the notion of equilibrium too far[24].

Since for each π there is a unique wage function consistent with the firms playing mutual best responses we can proceed as in the single group model and characterize the equilibrium set as fixed

[24] It has been suggested to us to consider a quota in the complex task together with a "civil rights law" prohibiting wage discrimination in the simple task. However, the same nonexistence problem remains under this policy.

points of a function from $[0,1]^2$ to $[0,1]^2$. The interpretation of this function is as simple as in the single group model. The function computes the fractions of agents in each group who invests as a best response to the wage function implied by any given investment behavior.

Using the wage schedules (2.29) we can express the expected gross benefits from undertaking the investment for an agent in group j when $\pi = (\pi^a, \pi^b)$ as

$$\begin{aligned} H^j(\pi) &= y_1(\widehat{r}(\pi), 1)\, p\left(\widehat{\theta}^j(\pi), \pi^j\right)\left(F_q\left(\widehat{\theta}^j(\pi)\right) - F_u\left(\widehat{\theta}^j(\pi)\right)\right) \\ &\quad + y_1(\widehat{r}(\pi), 1)\int_{\widehat{\theta}^j(\pi)}^1 p\left(\theta, \pi^j\right)\left(f_q(\theta) - f_u(\theta)\right) d\theta. \end{aligned} \tag{2.30}$$

Arguing as in the single group model we see easily that $\pi = (\pi^a, \pi^b)$ is an equilibrium if and only if

$$\pi^j = G\left(H^j(\pi)\right) \text{ for } j = 1, 2 \tag{2.31}$$

From these expressions it is easily seen that *any non-discriminatory equilibrium in the extended model is an equilibrium under affirmative action.* This should be fairly obvious since if the groups behave the same way then the employers voluntarily treat the groups identically. To see it formally we observe that the multiplier in the conditions (2.26) is zero when $\pi^a = \pi^b$, so both equations in (2.31) reduces to the fixed point equation for the single group model.

If there are asymmetric equilibria under affirmative action, inspection of (2.29) reveals that the wage schedule for the group with the lower fraction of agents who invests will be uniformly below the wage schedule for the other group. Hence, wage discrimination persists in our model unless the policy forces the economy to an equilibrium where the fractions of agents who invest are the same in both groups. For this reason we will say that an equilibrium is *discriminatory* unless $\pi^a = \pi^b$ and although one could potentially think of alternatives, we say that the *most discriminatory equilibrium* is the equilibrium for which the difference $|\pi^a - \pi^b|$ is the largest. We observe:

Observation *If $0 < G(0) < \overline{G}_0$ (with $\overline{G}_0$ defined as in Proposition 7) so that the most discriminatory equilibrium (π^a, π^b) is such that $\pi^a = G(0)$ and if $(\widehat{\pi}^a, \widehat{\pi}^b)$ is the most discriminatory equilibrium under affirmative action with $\widehat{\pi}^a < \widehat{\pi}^b$. Then $\widehat{\pi}^a > \pi^a$.*

To see this, observe that if $G(0) > 0$ there must clearly be a positive fraction of agents from both groups who invest in any equilibrium. Hence the wage schedule for both groups will be *strictly* increasing and there will consequently be some (possibly very small) monetary incentives to invest for agents in both groups in any equilibrium under affirmative action.

The fact that affirmative action provides a lower bound on the fraction of investors in the discriminated group that is higher than in the "worst" equilibrium without affirmative action does not help us if we want to analyze the consequences of introducing affirmative action in general. In particular, it provides no guidance at all if we want to say something about the effects of affirmative action starting from a situation where there is discrimination, but where some agents from both groups are employed in both tasks.

If all equilibria under affirmative action would be non-discriminatory the situation would be different. Hence, it would be desirable to have some sufficient conditions under which affirmative action always eliminates the possibility of discriminatory equilibria. Intuitively, affirmative action makes it harder to sustain discrimination since it pushes up wages in the simple task for workers from the group with the higher fraction of investors and pushes down wages in the simple task for the discriminated group. However, as we show next, any such sufficient conditions must involve stringent restrictions on the distribution of investment costs.

Proposition 9. *Let y be a given production function and let f_q, f_u be some fixed densities, where y, f_q and f_u satisfies the assumptions stated in Section 2.2.1. Furthermore fix any $(\lambda^a, \lambda^b) \in int(\Delta^2)$. Then there exists some strictly increasing distribution function G with $G(0) > 0$ such that the model with affirmative action has an equilibrium $(\widehat{\pi}^a, \widehat{\pi}^b)$ with $\widehat{\pi}^a < \widehat{\pi}^b$.*

Proof. Suppose $\pi^a = 0$ and $0 < \pi^b < 1$. Then optimality conditions for the problem (2.25) can in this case be written as

$$\begin{aligned} y_2(\cdot) - \tfrac{\mu}{\lambda^a} &= 0 \\ -y_1(\cdot)\, p(\theta^b, \pi^b) + y_2(\cdot) + \tfrac{\mu}{\lambda^b} &= 0 \end{aligned} \tag{2.32}$$

we observe that the unique solution $\left(\widehat{\theta}^a, \widehat{\theta}^b\right)$ must still be interior. Using (2.30) we also see that $H^a(0, \pi^b) = 0 < H^b(0, \pi^b)$. It is straightforward to verify that H^j is continuous at $(0, \pi^b)$ for $j = a, b$ and it follows that there must exist some $(\widehat{\pi}^a, \widehat{\pi}^b)$ where $0 < \widehat{\pi}^a < \widehat{\pi}^b$ and, since H^a is initially increasing, $0 < H^a(\widehat{\pi}^a, \widehat{\pi}^b) < H^b(\widehat{\pi}^a, \widehat{\pi}^b)$. There must therefore exist some strictly increasing function G such that $G(0) > 0$, $G(H^a(\widehat{\pi}^a, \widehat{\pi}^b)) = \widehat{\pi}^a$ and $G(H^b(\widehat{\pi}^a, \widehat{\pi}^b)) = \widehat{\pi}^b$, i.e. $(\widehat{\pi}^a, \widehat{\pi}^b)$ is an equilibrium in the economy with fundamentals $\{y, f_q, f_u, (\lambda^a, \lambda^b), G\}$. ■

The result can be strengthened in several directions. First, it should be clear that we get multiplicity for a generic set of distribution functions. To see this one notes that we can always find an open set U containing $(\widehat{\pi}^a, \widehat{\pi}^b)$ such that the expected benefits of investment for members in group b exceeds the benefits for members in group a for all $\pi \in U$. Also, the argument only relies on existence of a function G that takes on particular values at a few points which means that assumptions about its curvature will not be enough to get any sufficient conditions for ruling out discriminatory equilibria. For example, one can show that there always exist uniform distributions such that there are discriminatory equilibria in the model. The idea should be clear from the proof above, but a slightly more complicated argument is needed to assure that $G(0) \geq 0$.

2.5.2. Welfare Effects of Affirmative Action

The purpose of this section is to illustrate that even if starting from the most extreme form of discrimination, the disadvantaged group may or may not be hurt by affirmative action. To show this we will consider simple distribution functions of the form:

$$G(c) = \begin{cases} \pi^a & \text{if } c \leq \bar{c} \\ \pi^b & \text{if } \bar{c} < c \leq \bar{\bar{c}} \\ 1 & \text{if } c > \bar{\bar{c}} \end{cases}, \tag{2.33}$$

where $\bar{\bar{c}}$ is some large cost. It is not difficult to extend the examples to strictly increasing distribution functions, but for analytical simplicity we will not do so.

In the first example it is demonstrated that even if we start from the most extreme form of discrimination introduction of affirmative action may make the disadvantaged group may be worse off. The idea is to choose (π^a, π^b) so that an there is an equilibrium with nobody in group a employed in the complex task. Using the notation from Section 2.4 we let $B^j(\pi)$ denote the expected gross benefits of investment in the model without affirmative action. Now, if we fix $\pi^b > 0$ and π^a is sufficiently small the solution to the task assignment problem for the firm is to assign all workers from group a to the simple task, while some workers from group b will be assigned to the complex task. Hence $B^a(\pi^a, \pi^b) = 0 < B^b(\pi^a, \pi^b)$ for π^a sufficiently small. Thus, if $\bar{c}$ in (2.33) is in between 0 and $B^b(\pi^a, \pi^b)$ we have that (π^a, π^b) is an equilibrium in the model without affirmative action. Next we proceed in the spirit of the argument of the proof of Proposition 9 and argue that for π^a small enough and for the right choice of $\bar{c}$ this will also be an equilibrium with affirmative action. To see this we note that $H^a(\pi^a, \pi^b) \to 0$ as $\pi^a \to 0$ while $H^b(\pi^a, \pi^b) \to H^b(0, \pi^b) > 0$ as $\pi^a \to 0$. Hence there exists some $\pi^a > 0$ such that

$$\min\left(B^b(\pi^a, \pi^b), H^b(\pi^a, \pi^b)\right) > H^a(\pi^a, \pi^b) > B^a(\pi^a, \pi^b) = 0. \tag{2.34}$$

The desired result follows: choosing $\bar{c}$ in (2.33) between $H^a(\pi^a, \pi^b)$ and the minimum of $B^b(\pi^a, \pi^b)$ and $H^b(\pi^a, \pi^b)$ we have that (π^a, π^b) is an equilibrium both with and without the policy. It is now obvious that *output decreases when affirmative action is introduced* in this case. This follows since the set of feasible production plans with the policy is a strict subset of the feasible plans without the policy and the unique solution without the restriction is not in this subset.

The change in expected utility for an agent of group j who invests[25] is given by the difference in expectation of the wage schedules with respect to f_q:

$$\begin{aligned} \Delta W^j_{INV} &= y_2\left(\widetilde{r}(\pi),1\right)F_q\left(\widetilde{\theta^j}(\pi)\right)+\int_{\widetilde{\theta^j}(\pi)}^{1} y_1\left(\widetilde{r}(\pi),1\right)p\left(\theta,\pi^j\right)f_q(\theta)\,d\theta - \\ &\quad -y_1\left(\widehat{r}(\pi),1\right)\left(p\left(\widehat{\theta^j}(\pi),\pi^j\right)F_q\left(\widehat{\theta}(\pi)\right)+\int_{\widehat{\theta^j}(\pi)}^{1} p\left(\theta,\pi^j\right)f_q(\theta)\,d\theta\right) \end{aligned} \quad (2.35)$$

The change in expected utility for agents who do not invest is derived symmetrically. Note that the factor ratio in general changes when the policy is introduced. This effect may go either way depending on the choice of production function and of distribution functions for θ. However, given the way this example is constructed this creates no difficulties. If the factor ratio increases when the policy is introduced we can use (2.35) directly to show that the expected benefits for *any* agent in group a decreases with the policy. On the other hand, if the factor ratio decreases we cannot say anything in general since the wage under affirmative action for high realizations of θ now may be higher than the (constant) wage in the original discriminatory equilibrium. However, no matter what happens to the factor ratio we can always rely on the fact that under affirmative action $w^a(\theta) \leq y_1\left(\widehat{r}(\pi),1\right)p\left(1,\pi^a\right)$ for all θ, so for π^a small enough the policy must decrease the expected utility for all agents in group a. The welfare effect for the other group is ambiguous.

It is also possible to construct examples where the discriminated group gains even if an equilibrium occurs that leaves the group discriminated. To illustrate the point we want to make it is however more straightforward to show that introduction of affirmative action may imply that a symmetric equilibrium is the only possibility.

Consider the benefits of investment if a fraction π^b would invest in *both groups*, $B^j\left(\pi^b,\pi^b\right)$ in the model without the quota and $H^j\left(\pi^b,\pi^b\right)$ in the model with, both which equals the benefits of investment in the single group model if a fraction π^b invests, $H\left(\pi^b\right)$. It is straightforward to show that the benefits of investment for agents in one of the groups is monotonically decreasing in the fraction of investors in the other group, so

$$B^a\left(\pi^a,\pi^b\right) < B^a\left(\pi^b,\pi^b\right) = H\left(\pi^b\right) < B^b\left(\pi^a,\pi^b\right) \quad (2.36)$$

[25]Since we constructed the equilibria so that the fraction of investors remains the same we do not need to worry about agents who change their behaviour when the policy is introduced.

As we argued above $B^a(\pi^a, \pi^b) = 0$ for some small enough $\pi^a > 0$, so for $\bar{c}$ chosen in between 0 and $B^b(\pi^a, \pi^b)$ and π^a small enough (π^a, π^b) is an equilibrium. But if $\bar{c}$ is chosen in the interval $(0, H^a(\pi^a, \pi^b))$ then (π^a, π^b) is not an equilibrium when affirmative action is introduced. Thus (assuming a is the smaller group we cannot switch to discrimination of the other group) the only remaining equilibrium candidate is one where a fraction of π^b invests in both groups. Since $\bar{c}$ can be chosen so that it is smaller than $H(\pi^b)$ there is a range of parameter values so that this is indeed an equilibrium. One can show that the wage for agents employed in the simple task will be higher in the symmetric equilibrium and this means that all agents in group a benefits from the policy. Since production increases the other group may or may not be made worse off.

By relatively standard continuity arguments both examples can be extended to some strictly increasing distribution functions as well.

We also conjecture that there are circumstances where affirmative action is necessarily a Pareto improvement and other circumstances where removing affirmative action is a Pareto improvement, but we have not been able to show this yet.

2.6. Discussion

We believe that our model captures important aspects of how discrimination may be sustained in the real world: when few workers of a particular group invest in their skills, the firms will tend not to promote these workers to higher paid more qualified jobs. This in turn suggests that the incentives to invest in human capital should be lower for agents from a group where few workers invest in their skills than for agents from a group where more workers invest. Hence discrimination as a consequence of self-fulfilling expectations seems like a plausible explanation for differences in labor market performance between groups.

Multiple equilibrium explanations of discrimination (as well as of other economic phenomena) are often criticized on the grounds that the model gives no prediction. Our model is also vulnerable to this type of criticism since it does not give a unique prediction for any fixed fundamentals. However, when we combine the logic of self-confirming expectations with factor complementarity this problem becomes less severe since the model has some implications about the relation between relative group size and possibilities for discrimination. In this context we again want to stress that since group size matters in the determination of discriminatory equilibria we cannot take an arbitrary discriminatory equilibrium and construct a new equilibrium by reversing the roles of the groups.

Since we are explicitly taking competitive forces on the labor market into consideration our model is a natural framework to analyze the consequences of anti-discriminatory policies. In this paper, we focus on affirmative action. The specific way we model it is subject to criticism since one would more naturally require quotas only on the skilled job, rather than in both jobs. Unfortunately, a quota in the skilled job only implies non-existence of "labor market equilibria" in continuation games where the two groups behave differently. Besides this technical aspect, we do not think that modelling affirmative action as a quota is particularly problematic. We already pointed out that we could as well impose a penalty on employers not conforming to a specific requirement and that a sufficiently stiff penalty would give the same results as our policy gives. The quotas we are considering also have the attractive feature that they are possible to implement under rather weak assumptions about what is observable to the policymaker.

One can think of alternative anti-discriminatory policies, such as different kinds of subsidies, that are feasible under the same informational assumptions that are needed in order to implement employment quotas. In a sequel to this paper we intend to compare the effects of quotas of the type considered in this paper and different subsidies and to analyze optimal policies under different informational assumptions.

2.7. Appendix

2.7.1. Proof of Proposition 4

Proof. *(sufficiency)* Suppose that one of the firms would deviate from the proposed equilibrium strategies and play some arbitrary strategy $\langle w_i', \xi_i' \rangle$ so that the actions on the implied *outcome path* $\langle w_i', t_i' \rangle$ is different from $\langle w, t \rangle$ given by (2.11) and (2.6) on a set of positive measure (in principle both firms could be playing according to the characterization and one firm could deviate by offering a wage schedule different only on a finite set of points and this way trigger the other firm to react by changing the task assignment rule. However, such a deviation would not change profits for the deviator, which is why we without loss of generality can assume that the *actions* by the deviator is changed on a set of positive measure). Define the following sets: $\Theta^h = \{\theta : w'(\theta) > w(\theta)\}$, $\Theta^l = \{\theta : w'(\theta) < w(\theta)\}$, $\Theta^e = \{\theta : w'(\theta) = w(\theta)\}$. For ease of notation let C' and S' denote the implied factor inputs for the deviating firm given that the other firm plays according to the proposed equilibrium strategies. Using (2.4) we see that these quantities can be expressed as:

$$C' = \int_{\theta\in\Theta^h} t'(\theta)\,\pi f_q(\theta)\,d\theta + \frac{1}{2}\int_{\theta\in\Theta^c} t'(\theta)\,\pi f_q(\theta)\,d\theta \tag{2.37}$$

$$S' = \int_{\theta\in\Theta^h} (1-t'(\theta))\, f_\pi(\theta)\,d\theta + \frac{1}{2}\int_{\theta\in\Theta^c} (1-t'(\theta))\, f_\pi(\theta)\,d\theta$$

where $f_\pi(\cdot)$ denotes the density where $f_\pi(\theta) = \pi f_q(\theta) + (1-\pi) f_u(\theta)$ for each $\theta \in [0,1]$. Using the definition of the profit function (2.5) and the allocation rule (2.3) we can express the profits for the deviating firm as;

$$\Pi^i_{dev} = y(C', S') - \int_{\theta\in\Theta^h} w'(\theta)\, f_\pi(\theta)\,d\theta - \frac{1}{2}\int_{\theta\in\Theta^c} w(\theta)\, f_\pi(\theta)\,d\theta \tag{2.38}$$

Let $C, S > 0$ be the implied factor inputs if both firms are playing according to the equilibrium strategies. By concavity of y and Euler's theorem it follows that

$$\Pi^i_{dev} \leq y_1(C,S)\,C' + y_2(C,S)\,S' - \int_{\theta\in\Theta^h} w'(\theta)\, f_\pi(\theta)\,d\theta - \frac{1}{2}\int_{\theta\in\Theta^c} w(\theta)\, f_\pi(\theta)\,d\theta \tag{2.39}$$

From the definition of $p(\theta,\pi)$ we have that $\pi f_q(\theta) = p(\theta,\pi)\, f_\pi(\theta)$ and from the proposed equilibrium wage schedule (2.11) we get $w(\theta) = \max\{y_1(C,S)\,p(\theta,\pi), y_2(C,S)\}$. Some algebraic manipulations using these equalities gives;

$$y_1(C,S)\,C' \leq \int_{\theta\in\Theta^h} t'(\theta)\, y_1(C,S)\, p(\theta,\pi)\, f_\pi(\theta)\,d\theta + \frac{1}{2}\int_{\theta\in\Theta^c} t'(\theta)\, y_1(C,S)\, p(\theta,\pi)\, f_\pi(\theta) \tag{2.40}$$

$$y_2(C,S)\,S' \leq \int_{\theta\in\Theta^h} (1-t'(\theta))\, w(\theta)\, f_\pi(\theta)\,d\theta + \frac{1}{2}\int_{\theta\in\Theta^c} (1-t'(\theta))\, w(\theta)\, f_\pi(\theta)\,d\theta$$

Summing over these inequalities we see that

$$y_1(C,S)\,C' + y_2(C,S)\,S' \leq \int_{\theta\in\Theta^h} w(\theta)\, f_\pi(\theta)\,d\theta + \frac{1}{2}\int_{\theta\in\Theta^c} w(\theta)\, f_\pi(\theta)\,d\theta, \tag{2.41}$$

and by substituting this into (2.39) we get

$$\Pi^i_{dev} \leq \int_{\theta\in\Theta^h} (w(\theta) - w'(\theta))\, f_\pi(\theta)\,d\theta, \tag{2.42}$$

and since $w'(\theta) > w(\theta)$ for all $\theta \in \Theta^h$ this means that no deviation earns positive profits (we can not conclude that a deviation leaves the deviator strictly worse off since there are deviations that "scale down" production that also gives zero profits). Since the deviation was arbitrary this completes the proof of the sufficiency part of Proposition 4. ■

The necessity part of Proposition 4 will be proved by using a sequence of intermediate results:

Lemma 12. *Suppose $(w_i, \xi_i)_{i=1,2}$ is a pair of best responses. Then $w_1(\theta) = w_2(\theta)$ for almost all $\theta \in [0,1]$*

Proof. Suppose $\langle w_i, \xi_i \rangle_{i=1,2}$ are best responses and that there exists a set $\Theta \subseteq [0,1]$ positive measure such that $w_i(\theta) - w_j(\theta) > 0$ for all $\theta \in \Theta$. Consider a deviation $\langle w_i', \xi_i' \rangle$ such that $\xi_i'(w_i', w_j)(\theta) = \xi_i(w_i, w_j)(\theta)$ for all θ, $w_i'(\theta) = w_i(\theta)$ for $\theta \in [0,1] \backslash \Theta$ and $w_i'(\theta) = (w_i(\theta) + w_j(\theta))/2$ for $\theta \in \Theta$. We notice that the deviation leaves the distribution of available workers unchanged and since the task assignment rule on the outcome path also is unchanged this means that output is unchanged. The difference in expected profits is then simply the difference in the total wage bill, i.e.

$$\Delta\Pi_{dev}^i = \int_{\theta\in\Theta} \frac{w_i(\theta) - w_j(\theta)}{2} (\pi f_q(\theta) + (1-\pi) f_u(\theta)) \, d\theta > 0 \tag{2.43}$$

which contradicts the hypothesis that $\langle w_i, \xi_i \rangle_{i=1,2}$ is a pair of best responses. ∎

Lemma 13. *Let t_i denote the implied task assignment rule on the equilibrium path for firms $i = 1,2$. Then there exists some $\widetilde{\theta}^i \in (0,1)$ such that $t_i(\theta) = 1$ for almost all $\theta > \widetilde{\theta}^i$ and $t_i(\theta) = 0$ for almost all $\theta < \widetilde{\theta}^i$ and for $i = 1,2$.*

Proof. By Lemma 12, $w_i(\theta) = w_j(\theta)$ for almost all θ, so $I_{\langle w_1, w_2\rangle}^i(\theta) = \frac{1}{2}$ for almost all θ. It follows that if t_i must satisfy:

$$\begin{aligned} t_i(\cdot) \in \; & \arg\max_{t(\cdot)} y(c_i, s_i) \\ \text{subj. to} \quad C_i = \; & \int t(\theta) \pi f_q(\theta) \, d\theta \\ S_i = \; & \int (1 - t(\theta)) f_\pi(\theta) \, d\theta \end{aligned} \tag{2.44}$$

For a contradiction, suppose that the claim is false. Then there are sets $\Theta^h, \Theta^l \subseteq [0,1]$ with positive measure such that $\theta^h > \theta^l$ for all $\theta^h, \theta^l \in \Theta^h \times \Theta^l$, but $t_i(\theta^h) = 0$ for all $\theta^h \in \Theta^h$ and $t_i(\theta^l) = 1$ for all $\theta^l \in \Theta^l$. Since f_q and f_u are continuous the mixture f_π is continuous as well and we may therefore without loss of generality assume that $\int_{\theta\in\Theta^h} f_\pi(\theta) \, d\theta = \int_{\theta\in\Theta^l} f_\pi(\theta) \, d\theta > 0$. Consider the alternative task-assignment rule,

$$t_i^a(\theta) = \begin{cases} 1 & \text{if } \theta \in \Theta^h \\ 0 & \text{if } \theta \in \Theta^l \\ t_i(\theta) & \text{otherwise} \end{cases} \tag{2.45}$$

Let S_i^a and C_i^a be the factor inputs implied by t_i^a and let S_i and C_i be the inputs given the rule t_i. Since $\int_{\theta\in\Theta^h} f_\pi(\theta) \, d\theta = \int_{\theta\in\Theta^l} f_\pi(\theta) \, d\theta$ it follows that $S_i^a = S_i$, the input of simple labor is unchanged. Since the deviation assigns to the complex task workers who are productive with a higher probability it is rather obvious that $C_i^a > C_i$. To see this formally we note that

$$C_i^a = \frac{\pi}{2}\left[\int_{\theta\in\Theta^h} f_q(\theta)\,d\theta - \int_{\theta\in\Theta^l} f_q(\theta)\,d\theta\right] + C_i. \tag{2.46}$$

Suppose $C_i^a \leq C_i$, which by (2.46) implies $\int_{\theta\in\Theta^h} f_q(\theta)\,d\theta \leq \int_{\theta\in\Theta^l} f_q(\theta)\,d\theta$. Let $l(\theta)$ denote the likelihood ratio $f_q(\theta)/f_u(\theta)$ and rewrite this inequality as

$$\int_{\theta\in\Theta^h} l(\theta)\, f_u(\theta)\,d\theta \leq \int_{\theta\in\Theta^l} l(\theta)\, f_u(\theta)\,d\theta, \tag{2.47}$$

which since $l(\theta^h) > l(\theta^l)$ for all θ^h, θ^l implies that $\int_{\theta\in\Theta^h} f_u(\theta)\,d\theta \leq \int_{\theta\in\Theta^l} f_u(\theta)\,d\theta$. But then

$$\begin{aligned}\int_{\theta\in\Theta^h} f_\pi(\theta)\,d\theta &= \pi\int_{\theta\in\Theta^h} f_q(\theta)\,d\theta + (1-\pi)\int_{\theta\in\Theta^h} f_u(\theta)\,d\theta < \\ &< \pi\int_{\theta\in\Theta^l} f_q(\theta)\,d\theta + (1-\pi)\int_{\theta\in\Theta^l} f_u(\theta)\,d\theta = \int_{\theta\in\Theta^l} f_\pi(\theta)\,d\theta\end{aligned} \tag{2.48}$$

which contradicts our original assumption. Hence $C_i^a > C_i$ and since $S_i^a = S_i$ this means that output is higher under t_i^a, so t_i could not solve (2.44). ■

Lemma 14. *Let t_i denote the implied task assignment rule on the equilibrium path for firms $i = 1, 2$ and let t be defined by (2.6). Then $t_i(\theta) = t(\theta)$ almost everywhere.*

Proof. By Lemma 13 the problem of finding an optimal task assignment rule reduces to finding an optimal solution to the programing problem (2.7) in the main text. Since firms are facing symmetric problems we drop indices and perform a change in variables by defining $C = \pi(1 - F_q(\theta))$ and $S = \pi F_q(\theta) + (1-\pi) F_u(\theta)$. The problem can then be restated as

$$\begin{aligned}&\max_{C,S}\, y(C,S) \\ &\text{subj. to } g(C,S) \equiv \pi - C - S + (1-\pi) F_u\left(F_q^{-1}\left(\tfrac{\pi - C}{\pi}\right)\right) \geq 0\end{aligned} \tag{2.49}$$

One verifies that

$$\frac{\partial g(C,S)}{\partial C} = -1 - \frac{1-\pi}{\pi}\frac{f_u\left(F_q^{-1}\left(\frac{\pi-C}{\pi}\right)\right)}{f_q\left(F_q^{-1}\left(\frac{\pi-C}{\pi}\right)\right)} = -1 - \frac{1-\pi}{\pi}\frac{1}{\psi\left(F_q^{-1}\left(\frac{\pi-C}{\pi}\right)\right)} \tag{2.50}$$

and taking second derivatives we find that $\partial^2 g/\partial C^2 < 0$ while all other elements of the Hessian matrix is zero by the linearity in S. Hence g is concave (one can actually see that $\partial^2 g/\partial C^2 < 0$ without explicitly performing the differentiation since ψ and F_q^{-1} are both strictly increasing). Since y is concave the Kuhn-Tucker conditions are sufficient conditions for a solution to (2.49) and necessity follows since concavity of g is sufficient for constraint qualification. Invoking the boundary conditions we easily see that any solution to the Kuhn-Tucker conditions must be interior. Since the programs (2.8) and (2.49) are equivalent this completes the proof. ■

Lemma 15. *Suppose* $\langle w_1, w_2 \rangle$ *is a pair of equilibrium wage schedules and let* $\tilde{\theta}(\pi)$ *be the solution to (2.6). Then there is a pair* (k_s, k_c) *such that* $w_i(\theta) = k_s$ *for* $i = 1, 2$ *and for almost all* $\theta < \tilde{\theta}(\pi)$ *and* $w_i(\theta) = p(\theta, \pi) k_c$ *for* $i = 1, 2$ *and for almost all* $\theta > \tilde{\theta}(\pi)$.

Proof. We will begin by showing that $w_i(\theta) = k_s$ for almost all $\theta < \tilde{\theta}$. For contradiction assume that there exists sets $\Theta^a, \Theta^b \subseteq \left[0, \tilde{\theta}(\pi)\right]$ with strictly positive measure such that $w_i(\theta) < k$ for all $\theta \in \Theta^a$ and $w_i(\theta) \geq k$ for all $\theta \in \Theta^b$. By continuity of f_π we may without loss of generality assume that $\int_{\theta \in \Theta^a} f_\pi(\theta) d\theta = \int_{\theta \in \Theta^b} f_\pi(\theta) d\theta > 0$. To show that this is inconsistent with equilibrium we will construct a deviation where the firm replace get rid of some workers that are paid above k and attracts some workers that are being paid a lower wage. Intuitively it is rather clear that this deviation will be profitable as long as the total input of workers in both tasks constant. To show this formally consider the following deviation by firm i

$$w_i'(\theta) = \begin{cases} w_i(\theta) + \epsilon & \text{for } \theta \in \Theta^a \\ 0 & \text{for } \theta \in \Theta^b \\ w_i(\theta) & \text{otherwise} \end{cases} \tag{2.51}$$

Since input of both factors remains constant under the deviation (given that the task assignment rule is unchanged, which we assume) the difference in payoffs for the deviating firm is just the difference in wage payments, i.e.

$$\Delta(\epsilon) = \frac{1}{2}\left[\int_{\theta \in \Theta^b} w_i(\theta) f_\pi(\theta) d\theta - \int_{\theta \in \Theta^a} (w_i(\theta) + \epsilon) f_\pi(\theta) d\theta\right] \tag{2.52}$$

Since $\lim_{\epsilon \to 0} \Delta(\epsilon) > 0$ there exists $\epsilon > 0$ such that $\Delta(\epsilon) > 0$. Hence, for ϵ small enough the deviation is profitable.

Symmetrically, suppose there are sets $\Theta^a, \Theta^b \subseteq \left[\tilde{\theta}(\pi), 1\right]$ with strictly positive measure (where we again w.l.g may assume $\int_{\theta \in \Theta^a} f_q(\theta) d\theta = \int_{\theta \in \Theta^b} f_q(\theta) d\theta$) such that $(w_i(\theta) / p(\theta, \pi)) < k$ for all $\theta \in \Theta^a$ and $(w_i(\theta) / p(\theta, \pi)) \geq k$ for all $\theta \in \Theta^b$. Again we consider a deviation according to (2.51). Noting that $p(\theta, \pi) = (\pi f_q(\theta) / f_\pi(\theta))$ we find that output is unchanged in this case as well and it is easy to verify by a similar argument as above that the deviation is profitable for ϵ small enough. ■

We now collect the pieces together and prove the necessity part of Proposition 4:

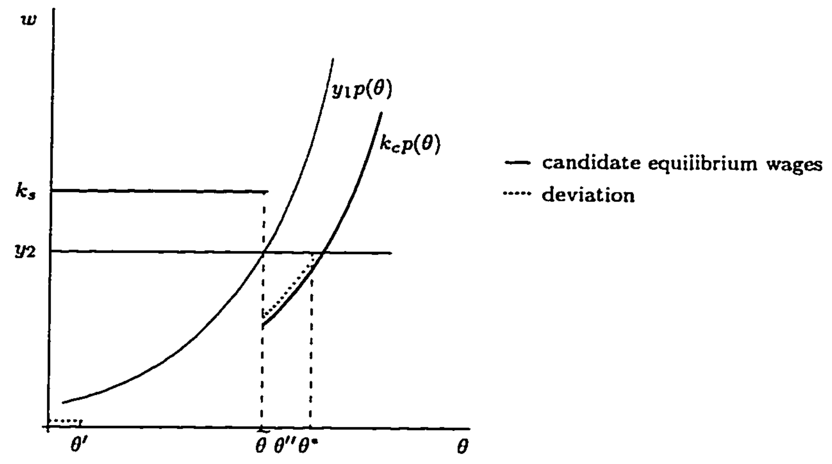

Figure 2.5: A profitable deviation

Proof. (necessity). By Lemma 14 it follows that the task assignment rule on any equilibrium path must satisfy $t_i(\theta) = t(\theta)$ for $i = 1, 2$ and almost all θ. Using the notation from the main text and constant returns to scale we have from Lemma 15 that in any equilibrium of the model both firms must offer almost identical wage schedules (disregarding sets of measure zero) of the form:

$$w(\theta) = \begin{cases} k_s \\ p(\theta, \pi) k_c \end{cases} \tag{2.53}$$

It remains to be shown that $k_s = y_2(C, S)$ and $k_c = y_1(C, S)$. By straightforward calculations it can be shown that if $k_s < y_2(C, S)$ and $k_c < y_1(C, S)$ then both firms are making positive profits and a uniform deviation where firm i offers $w_i'(\theta) = w_i(\theta) + \epsilon$ for all θ would be profitable for ϵ small enough. Also, if both inequalities would go the other way and the wages would be uniformly above the candidate equilibrium wage schedule both firms would make strictly negative profits and a deviation to $w_i(\theta) = 0$ for all θ would be profitable. The cases that requires a little work are when the inequalities work in opposite directions.

The two cases can be taken care of with perfectly symmetric arguments we will only consider the case with $k_s > y_2(C, S)$ and $k_c < y_1(C, S)$. The idea behind the construction is illustrated in Figure 6.

Recall that $y_1(C, S)\, p\left(\widetilde{\theta}(\pi), \pi\right) = y_2(C, S)$ by (2.8). Hence if $k_s > y_2(C, S)$ and $k_c < y_1(C.S)$ then $k_c p\left(\widetilde{\theta}(\pi), \pi\right) < k_s$ and there is an interval $\left(\widetilde{\theta}(\pi), \theta^*\right)$ such that $w_i(\theta) = p(\theta, \pi) k_c < k_s$ for all θ in this interval. The idea behind the construction (see Figure 6) is now simply to demonstrate that it is better to dispose of some of the workers being paid k_s and replace them by cheaper workers from $\left(\widetilde{\theta}(\pi), \theta^*\right)$. While this logic is perfectly simple the formal argument below is rather messy. The reason for this being that we need to keep track of the changes in the effective factor inputs as well as changes in the wage bill.

Let θ' solve $F_\pi(\theta') = F_\pi(\theta^*) - F_\pi\left(\widetilde{\theta}(\pi)\right)$ and define θ'' as the solution to $F_\pi(\theta'') - F_\pi\left(\widetilde{\theta}(\pi)\right) = (F_\pi(\theta^*) - F_\pi(\widetilde{\theta}(\pi)))/2$. Consider the following deviation for firm i :

$$w_i'(\theta) = \begin{cases} 0 & \text{for } \theta \in [0, \theta') \\ w_i(\theta) + \epsilon & \text{for } \theta \in [\widetilde{\theta}(\pi), \theta^*) \\ w_i(\theta) & \text{for } \theta \in [\theta^*, 1) \cup [\theta', \widetilde{\theta}) \end{cases} \tag{2.54}$$

$$t_i'(\theta) = \begin{cases} 0 & \text{for } \theta \in [0, \theta'') \\ 1 & \text{for } \theta \in [\theta'', 1) \end{cases} \tag{2.55}$$

By construction, the input of simple labor is unchanged (i.e. $1/2F_\pi\left(\widetilde{\theta}(\pi)\right)$ is the input of simple labor before the deviation and since the workers on $[0, \theta')$ will go to the other firm and since all workers on $[\widetilde{\theta}(\pi), \theta'')$ will be in thee firm after the deviation a quantity of $F_\pi(\theta'') - F_\pi\left(\widetilde{\theta}(\pi)\right) + 1/2\left(F_\pi\left(\widetilde{\theta}(\pi)\right) - F_\pi(\theta')\right)$ will be in the simple task after the deviation). The change in effective units of complex labor is given by $C' - C = \frac{\pi}{2}\left(F_q(\theta^*) - 2F_q(\theta'') + F_q\left(\widetilde{\theta}\right)\right)$ and using that $F_\pi(\theta^*) - 2F_\pi(\theta'') + F_\pi\left(\widetilde{\theta}\right) = 0$ it is not hard to show that $C' - C > 0$. This should be fairly obvious since the mass of workers assigned to the complex task is unchanged but the average value of θ has increased (formal argument will is similar to the one used in Lemma 13). Thus, output increases under the deviation so the difference in payoffs must be larger than the difference in the wage bill, that is:

$$\begin{aligned} \Delta(\epsilon) &> \tfrac{1}{2}\int_0^1 w_i(\theta) f_\pi(\theta)\, d\theta - \tfrac{1}{2}\int_{\theta'}^{\widetilde{\theta}} w_i(\theta) f_\pi(\theta)\, d\theta - \int_{\widetilde{\theta}}^{\theta^*} (w_i(\theta) - \epsilon) f_\pi(\theta)\, d\theta - \tfrac{1}{2}\int_{\theta^*}^1 w_i(\theta) f_\pi(\theta)\, d\theta = \\ &= \tfrac{1}{2}k_s F_\pi(\theta') - \tfrac{1}{2}\int_{\widetilde{\theta}}^{\theta^*} p(\theta, \pi) k_c f_\pi(\theta)\, d\theta - \int_{\widetilde{\theta}}^{\theta^*} \epsilon f_\pi(\theta)\, d\theta \end{aligned} \tag{2.56}$$

Recall that $p(\theta, \pi) k_c < k_s$ for $\theta \in \left(\widetilde{\theta}, \theta^*\right)$ and that $F_\pi(\theta') = F_\pi(\theta^*) - F_\pi\left(\widetilde{\theta}\right)$ so that:

$$\frac{1}{2}k_s F_\pi(\theta') = \frac{1}{2}\int_{\widetilde{\theta}}^{\theta^*} k_s f_\pi(\theta)\, d\theta > \frac{1}{2}\int_{\widetilde{\theta}}^{\theta^*} p(\theta, \pi) k_c f_\pi(\theta)\, d\theta. \tag{2.57}$$

Hence, $\lim_{\epsilon \to 0} \Delta(\epsilon) > 0$ and there exists $\epsilon > 0$ such that the deviation is profitable. The case with $k_s < y_2(C, S)$ and $k_c > y_1(C, S)$ can be treated symmetrically and Proposition 2 follows. ■

2.7.2. Proof of Proposition 6

Lemma 16. *Suppose that $y : R_+^2 \to R$ is strictly concave in both arguments and homogenous of degree 1. Then for each $\pi \in (0,1]$ there exists a unique $\widetilde{\theta}(\pi) \in (0,1)$ such that (2.8) is satisfied.*

Proof. Define $\rho : (0,1) \times (0,1] \to R_+$ by $\rho(\theta,\pi) = \frac{\pi(1-F_q(\theta))}{\pi F_q(\theta)+(1-\pi)F_u(\theta)}$ for all $(\theta,\pi) \in (0,1) \times (0,1]$ and let the function $D : (0,1) \times (0,1] \to R$ be defined as

$$D(\theta,\pi) = p(\theta,\pi) - \frac{y_2(\rho(\theta,\pi),1)}{y_1(\rho(\theta,\pi),1)}. \tag{2.58}$$

Since y is homogenous of degree one $\widetilde{\theta}(\pi)$ solves the first order condition for the task assignment problem (equation (2.8) in Section 2.3) if and only if $D(\widetilde{\theta}(\pi),\pi) = 0$. It follows the (strict) monotone likelihood ratio property that $p(\theta,\pi)$ is strictly increasing in θ for any $\pi > 0$. Fixing S , $y_1(C,S)$ is strictly decreasing in C while $y_2(C,S)$ is strictly increasing in C. Since F_q and F_u are strictly increasing $\rho(\theta,\pi)$ is strictly decreasing in θ. Consequently $y_1(\rho(\theta,\pi),1)$ is strictly increasing and $y_2(\rho(\theta,\pi),1)$ is strictly decreasing in θ. Hence, the ratio $y_2(\rho(\theta,\pi),1)/y_1(\rho(\theta,\pi),1)$ is strictly decreasing which implies that $D(\theta,\pi)$ is strictly increasing in θ. Thus, there can be at most one solution $D(\widetilde{\theta}(\pi),\pi) = 0$ and the next task is to show that a solution exists for any $\pi > 0$. We note that $0 < p(0,\pi) < p(1,\pi) < 1$ for any $\pi > 0$. Since F_q and F_u are c.d.fs it is easy to check that $\lim_{\pi\to 0}\rho(\theta,\pi) = \infty$ and $\lim_{\pi\to 1}\rho(\theta,\pi) = 0$. Using the boundary conditions A2, constant returns to scale and standard limit laws

$$\lim_{\theta\to 0}\frac{y_2(\rho(\theta,\pi),1)}{y_1(\rho(\theta,\pi),1)} = \lim_{\theta\to 0}\frac{y_2\left(1,\frac{1}{\rho(\theta,\pi)}\right)}{y_1(\rho(\theta,\pi),1)} = \frac{\lim_{x\to 0} y_2(1,x)}{\lim_{z\to\infty} y_1(z,1)} = \infty. \tag{2.59}$$

Symmetrically $\lim_{\theta\to 1} y_2(\rho(\theta,\pi),1)/y_1(\rho(\theta,\pi),1) = 0$. It follows that $\lim_{\theta\to 0} D(\theta,\pi) = -\infty$ and $\lim_{\theta\to 1} D(\theta,\pi) = p(1,\pi) > 0$. Hence there exists a unique $\widetilde{\theta}(\pi) \in (0,1)$ satisfying $D(\widetilde{\theta}(\pi),\pi) = 0$ for each $\pi > 0$ ■

Lemma 17. *$\widetilde{\theta}$ and r satisfy the following properties:*

i $\widetilde{\theta}$ is continuously differentiable on $(0,1)$.

ii $\lim_{\pi\to 0+} r(\pi) = 0$

iii r is monotonically increasing in π

Proof. i) It is easy to check that $D_1(\theta,\pi) > 0$ for each $\pi > 0$ which means that the hypotheses of the implicit function theorem is satisfied for each $\widetilde{\theta}(\pi)$. ii) The equation $D\left(\widetilde{\theta}(\pi),\pi\right)$

$= p\left(\widetilde{\theta}(\pi),\pi\right) - \frac{y_2(r(\pi),1)}{y_1(r(\pi),1)} = 0$ must be satisfied for each $\pi > 0$. But $\lim_{\pi\to 0} p\left(\widetilde{\theta}(\pi).\pi\right) \leq \lim_{\pi\to 0} p(1,\pi) = 0$ so $\lim_{\pi\to 0}\frac{y_2(r(\pi),1)}{y_1(r(\pi),1)} = 0$, which implies that $\lim_{\pi\to 0} r(\pi) = 0$. iii) For a contradiction suppose $r(\pi) < r(\pi')$ for $\pi > \pi'$. Since (2.8) must be satisfied for both π and π' and since the first derivatives of y are decreasing in its own argument it follows that $p\left(\widetilde{\theta}(\pi),\pi\right) < p\left(\widetilde{\theta}(\pi'),\pi'\right)$. But since $p(\theta,\pi)$ is increasing in the second argument (that is, the posterior is increasing in the prior) this means that $\widetilde{\theta}(\pi) < \widetilde{\theta}(\pi')$. Plugging this into the definition of r it follows that $r(\pi) > r(\pi')$ which is a contradiction. ■

Proof. (Proposition 6) By Proposition 2 the set of all equilibria of the model are fully characterized as fixed points of the map $G \circ H : [0,1] \to [0,1]$, where H is defined by (2.14) in Section 2.3 (Proposition 2 says how equilibrium strategies consistent with a particular fixed point can be constructed).

By Lemma 17, $\widetilde{\theta}$ is continuously differentiable on $(0,1)$ and since it follows that r and H are compositions of continuously differentiable functions these are also continuously differentiable on $(0,1)$. This implies that H is a continuous function of π on $(0,1)$. By constant returns to scale the first derivatives of y are homogenous of degree 0 and since $\widetilde{\theta}(\pi)$ must satisfy the first order condition (2.8) we have that $\widetilde{\theta}(\pi), r(\pi)$ must satisfy

$$p\left(\widetilde{\theta}(\pi),\pi\right) y_1(r(\pi),1) = y_2(r(\pi),1) \tag{2.60}$$

for every $\pi \in (0,1)$. For $\pi = 1$ we have that $p(\theta,1) = 1$ for all θ, so it does not really matter what workers are assigned to the respective tasks. However, $r(1)$ must nevertheless satisfy $y_1(r(1),1) = y_2(r(\pi),1)$ (one particular way of achieving this is by a cutoff rule). Since the workers are all equally productive in both tasks we get that $w(\theta) = y_1(r(\pi),1) = y_2(r(\pi),1)$ for all θ and it follows that the benefits of investment is given by $H(1) = 0$. It is easy to verify that $\lim_{\pi\to 1} r(\pi) = r(1)$ by use of (2.60) and using (2.14) it follows that $\lim_{\pi\to 1} H(\pi) = H(1)$. The case with $\pi = 0$ is taken care of in the same way. No matter how workers are allocated between tasks output is zero, which implies that $w(\theta) = 0$ for all θ. Hence $H(0) = 0$. Furthermore from (2.60) we have that

$$0 = \lim_{\pi\to 0} p\left(\widetilde{\theta}(\pi),\pi\right) = \lim_{\pi\to 0}\frac{y_2(r(\pi),1)}{y_1(r(\pi),1)}. \tag{2.61}$$

and using the boundary conditions we see that the only possibility for this to be satisfied is if $\lim_{\pi\to 0} r(\pi) = 0$, $\lim_{\pi\to 0} y_2(r(\pi),1) = 0$ and $\lim_{\pi\to 0} y_1(r(\pi),1)p(\widetilde{\theta}(\pi),\pi) = 0$. Since $F_q(\theta) -$

$F_u(\theta)$ and $\int_{\widetilde{\theta}(\pi)}^{1} p(\theta,\pi)(f_q(\theta) - f_u(\theta))\,d\theta$ are bounded below and above it follows from (2.14)that $\lim_{\pi \to 0} H(\pi) = 0 = H(0)$ establishing continuity of $G \circ H$ on the whole interval $[0,1]$.

Consider any $\pi \in (0,1)$. Inspection of (2.60) shows that $0 < r(\pi) < \overline{r}$ where $\overline{r}$ is the unique value satisfying $y_1(\overline{r},1) = y_2(\overline{r},1)$. Hence $0 < \widetilde{\theta}(\pi) < 1$ and since $p(\theta,\pi)$ is strictly increasing in θ for any $0 < \pi < 1$ by the assumption of strictly monotone likelihood ratio it follows that

$$\int_{\widetilde{\theta}(\pi)}^{1} p(\theta,\pi)(f_q(\theta) - f_u(\theta))\,d\theta > p\left(\widetilde{\theta}(\pi),\pi\right)\left[F_u\left(\widetilde{\theta}(\pi)\right) - F_q\left(\widetilde{\theta}(\pi)\right)\right]. \tag{2.62}$$

Hence $H(\pi) > 0$ for all $\pi \in (0,1)$. Since $G(0) > 0$ it follows as a simple application of the intermediate value theorem that there exists at least one fixed point of $G \circ H$. It follows directly from the assumption that $G(0) > 0$ that any fixed point must be in the open interval $(0,1)$. Hence there exists at least one non trivial equilibrium ■

2.7.3. Proof of Proposition 7

Assume that there is an equilibrium such that all agents in group a are assigned to the complex task. Such an equilibrium exist if and only if there is a (π^a,π^b) solving $\pi^j = G(H^j(\pi^a,\pi^b))$ for $j = a,b$ such that $\widetilde{\theta}^a(\pi) = \widetilde{\theta}^a(\pi^a,\pi^b) = 1$. Note that in any such equilibrium we have that $w(\theta) = y_2(r(\pi),1)$ for all θ, implying that $\pi^a = G(0)$. The cutoff rules for the task assignments must satisfy the Kuhn-Tucker conditions for the problem (2.17), in particular

$$p\left(\widetilde{\theta}^b(\pi),\pi^b\right) y_1(r(\pi),1) = y_2(r(\pi),1) + (\gamma_b - \eta_b)/f_{\pi^b}\left(\widetilde{\theta}^b(\pi)\right), \tag{2.63}$$

where γ_b is the multiplier associated with the constraint that $\theta^b \geq 0$ and η_b with the constraint that $1-\theta^b \geq 0$. Note that $\widetilde{\theta}^b(\pi) < 1$ if $\widetilde{\theta}^a(\pi)$ since it follows from the definition of r that $r(\pi) = 0$ if $\left(\widetilde{\theta}^a(\pi),\widetilde{\theta}^b(\pi)\right) = (1,1)$. But then the left hand side of (2.63) is less than or equal to zero and the right hand side is unbounded, so this could not be the case in equilibrium.

Imposing $\widetilde{\theta}^a(\pi) = 1$ the condition (2.63) together with the complementary slackness conditions is indeed necessary and sufficient conditions for optimality. Proceeding step by step as in the proof of Lemma 16 in Section 2.7.2 one shows that there is a unique solution to these conditions (which may or may not involve assigning all workers in group b to the complex task) for any $\pi^b \in (0,1]$. Let this solution be given by $\theta^b(\pi^b)$ (observe that so for all we know so far this need not coincide with $\widetilde{\theta}^b(G(0),\pi^b)$). Fixing exogenously both the investment behavior and the task assignments for group a the model is qualitatively the same as the single group model (with some unexplained input of labor in the simple task) and we can establish that $\theta^b(\pi^b)$ is continuous in π^b by use of the implicit function theorem, exactly as in the single group model (the possibility of corner solutions where all b workers are assigned to the complex task does not create any discontinuities). Define $r^d : (0,1] \to R_+$ as the ratio of factor inputs function of π^b, assuming that all agents from group a are assigned to the simple task, i.e.:

$$r^d(\pi^b) = \frac{\lambda^b \pi^b \left(1 - F_q\left(\widetilde{\theta}^b(\pi^b)\right)\right)}{\lambda^a + \lambda^b \pi^b F_{\pi^b}\left(\widetilde{\theta}^b\right)} \tag{2.64}$$

Note that $\pi^b = G\left(H^b\left(G(0), \pi^b\right)\right)$ in order for π^b to be consistent with equilibrium. But this is equivalent to finding a fixed point of (2.15) with $H(\pi)$ defined as in (2.14) but with $\widetilde{\theta}$ replaced by θ^b and r replaced by r^d. It is easily checked that θ^b and r^d has all the properties of $\widetilde{\theta}$ and r that were used in the proof of Proposition 6 (i.e. Lemma 16 and Lemma 17I still holds). It then remains to check that it is optimal for the firms to assign all workers in group a to the simple task. From the (full) Kuhn-Tucker conditions it follows that this is the case if and only if

$$p(1, G(0)) \leq \frac{y_2\left(r^d(\pi^b), 1\right)}{y_1\left(r^d(\pi^b), 1\right)} = p\left(\theta^b(\pi^b), \pi^b\right) \equiv p(1, \overline{G}_0) \tag{2.65}$$

But $\lim_{G(0)\to 0} p(1, G(0)) = 0$ and $y_2\left(r^d(\pi^b), 1\right)/y_1\left(r^d(\pi^b), 1\right) > 0$. This implies that the cutoffs $\left(1, \theta^b(\pi^b)\right)$ indeed solves (2.17) if $\pi = \left(G(0), \pi^b\right)$ if $G(0)$ is small enough. Hence $r^d(\pi) = r(\pi)$ and $\left(G(0), \pi^b\right)$ is an equilibrium of the model. The fact that all agents in group b are paid a higher wage than any agent from group a follows directly from the wage schedules. ∎

2.7.4. Proof of Proposition 8

The proof of the sufficiency part of the proposition uses the following intermediate result.

Lemma 18. *Suppose both firms choose the proposed equilibrium strategy. Then both firms are earning zero profits.*

Proof. Using the proposed wage schedules (2.29) the total wage bill can be expressed as:

$$\begin{aligned} W &= \sum_{j=a,b} \lambda^j \int w^j(\theta) f_{\pi^j}(\theta)\, d\theta = \\ &= y_1 \left(\sum_{j=a,b} \lambda^j p\left(\widehat{\theta}^j, \pi^j\right) F_{\pi^j}\left(\widehat{\theta}^j\right) \sum_{j=a,b} \lambda^j \pi^j \left[1 - F_q\left(\widehat{\theta}^j\right)\right] \right) \end{aligned} \tag{2.66}$$

where the missing arguments of y_1 are the implied factor inputs. But $F_{\pi^a}\left(\widehat{\theta}^a\right) = F_{\pi^b}\left(\widehat{\theta}^b\right)$ due to the affirmative action constraint and using the fact that the weighted average of the marginal productivities in the complex task for the critical workers equals the marginal productivity in the simple task (equation 2.28) we get

$$W = y_2(\cdot) F_{\pi^a}\left(\widehat{\theta}^a\right) + \sum_{j=a,b} \lambda^j \pi^j \left[1 - F_q\left(\widehat{\theta}^j\right)\right] y_1(\cdot) \tag{2.67}$$

Since $\sum_{j=a,b} \lambda^j \pi^j \left[1 - F_q\left(\widehat{\theta}^j\right)\right] = C$ and $F_{\pi^a}\left(\widehat{\theta}^a\right) = \lambda^a F_{\pi^a}\left(\widehat{\theta}^a\right) + \lambda^b F_{\pi^b}\left(\widehat{\theta}^b\right) = S$ (this comes directly from the constraint imposed by affirmative action) the result follows from (2.67) by using Euler's theorem. ■

Proof of Proposition 8 (sufficiency). The proof parallels the proof of Proposition 4, but since the affirmative action constraint has to be used in a non-obvious way we give a rather detailed version of the proof. Suppose one firm should deviate from the candidate equilibrium strategies and play $\{w^a_{dev}, w^b_{dev}, \xi^a_{dev}, \xi^b_{dev}\}$ so that the actions implied on the outcome path are $\langle w^a_{dev}, w^b_{dev}, t^a_{dev}, t^b_{dev}\rangle$.

Define : $\Theta^h_j = \{\theta : \widehat{w}^a(\cdot) > w^a(\cdot)\}, \Theta^l_j = \{\theta : \widehat{w}^a(\cdot) < w^a(\cdot)\}, \Theta^e_j = \{\theta : \widehat{w}^a(\cdot) = w^a(\cdot)\}$ for $j = a, b$. Let C and S the implied factor inputs employed in the candidate equilibrium and C_{dev}, S_{dev} be the implied factor inputs for the deviating firm i given that the other firm still plays according to the proposed equilibrium strategies. The profits for the deviating firm, Π^i_{dev}, can be expressed as ;

$$\Pi^i_{dev} = y(C_{dev}, S_{dev}) - \sum_{j=a,b} \lambda^j \left[\int_{\theta\in\Theta^h_j} w^j_{dev}(\theta) f_\pi(\theta)\, d\theta + \frac{1}{2} \int_{\theta\in\Theta^e_j} w^j(\theta) f_\pi(\theta)\, d\theta \right] \tag{2.68}$$

Using concavity and constant returns to scale as in the in the proof of Proposition 4:

$$\begin{aligned}\Pi^i_{dev} \leq\ & y_1(C,S)\, C_{dev} + y_2(C,S)\, S_{dev} \\ & - \sum_{j=a,b} \left[\int_{\theta\in\Theta^h_j} w^j_{dev}(\theta) f_\pi(\theta)\, d\theta + \frac{1}{2} \int_{\theta\in\Theta^e_j} w^j(\theta) f_\pi(\theta)\, d\theta \right]\end{aligned} \tag{2.69}$$

Using $\pi^j f_q(\theta) = p(\theta, \pi^j) f_{\pi^j}(\theta)$, and manipulating, we have:

$$\begin{aligned} y_1(C,S)\, C_{dev} =\ & \sum_{j=a,b} \lambda^j \int_{\theta\in\Theta^h_j} t^j_{dev}(\theta)\, y_1(C,S)\, p\left(\theta, \pi^j\right) f_{\pi^j}(\theta)\, d\theta \\ & + \frac{1}{2} \sum_{j=a,b} \lambda^j \int_{\theta\in\Theta^e_j} t^j_{dev}(\theta)\, y_1(C,S)\, p\left(\theta, \pi^j\right) f_{\pi^j}(\theta)\, d\theta \end{aligned} \tag{2.70}$$

But by definition of (2.29) we have that $w^j(\theta) = y_1(C,S)\,p(\theta,\pi^j)$ for $\theta \geq \widehat{\theta}^j$ and $w^j(\theta) > y_1(C,S)\,p(\theta,\pi^j)$ for $\theta < \widehat{\theta}^j$.Hence

$$y_1(C,S)\,C_{dev} \leq \sum_{j=a,b} \lambda^j \left[\int_{\theta\in\Theta_j^h} t^j_{dev}(\theta)\, w^j(\theta)\, f_{\pi^j}(\theta)\,d\theta + \frac{1}{2} \int_{\theta\in\Theta_j^c} t^j_{dev}(\theta)\, w^j(\theta)\, f_{\pi^j}(\theta)\,d\theta \right] \tag{2.71}$$

Symmetrically, note that $w^j(\theta) = y_1(C,S)\,p\left(\widehat{\theta}^j,\pi^j\right)$and $w^j(\theta) > y_1(C,S)\,p\left(\widehat{\theta}^j,\pi^j\right)$ for $\theta > \widehat{\theta}^j$. Note that

$$\begin{aligned} y_1(C,S)\,p\left(\widehat{\theta}^j,\pi^j\right)\frac{S^j_{dev}}{\lambda^j} &\leq \int_{\theta\in\Theta_j^h} w^j(\theta)\left(1-t^j_{dev}(\theta)\right) f_{\pi^j}(\theta)\,d\theta \\ &\quad + \frac{1}{2}\int_{\theta\in\Theta_j^c} w^j(\theta)\left(1-t^j_{dev}(\theta)\right) f_{\pi^j}(\theta)\,d\theta \end{aligned} \tag{2.72}$$

Making use of affirmative action constraint it follows that:

$$y_2(C,S)\,S_{dev} = y_2(C,S)\left(\frac{\lambda^a S^a_{dev}}{\lambda^a} + \frac{\lambda^b S^b_{dev}}{\lambda^b}\right) = y_2(C,S)\left(\lambda^a+\lambda^b\right)\frac{S^b_{dev}}{\lambda^b} \tag{2.73}$$

and combining with (2.72) and (2.28) we get:

$$y_2(C,S)\,S_{dev} \leq \sum_{j=a,b} \lambda^j \int_{\theta\in\Theta_j^h} w^j(\theta)\left(1-t^j_{dev}(\theta)\right) f_{\pi^j}(\theta)\,d\theta \tag{2.74}$$

$$+\frac{1}{2}\sum_{j=a,b} \lambda^j \int_{\theta\in\Theta_j^c} w^j(\theta)\left(1-t^j_{dev}(\theta)\right) f_{\pi^j}(\theta)\,d\theta \tag{2.75}$$

Summing over (2.71) and (2.74) we get:

$$y_1(C,S)\,C_{dev} + y_2(C,S)\,S_{dev} \leq \sum_{j=a,b} \lambda^j \left[\int_{\theta\in\Theta_j^h} w^j(\theta)\, f_{\pi^j}(\theta)\,d\theta + \frac{1}{2}\int_{\theta\in\Theta_j^c} w^j(\theta)\, f_{\pi^j}(\theta)\,d\theta \right] \tag{2.76}$$

The last steps of the argument is exactly as in the proof of Proposition 4. Substituting (2.76) into the expression for the profits and noting that the deviator must pay higher wages than the candidate equilibrium wages over the relevant ranges gives the result. ■

The necessity part is proved using the following steps.

Lemma 19. *Suppose* $\langle w_i^a, w_i^b, \xi_i^a, \xi_i^b\rangle_{i=1,2}$ *is a pair of best responses. Then, (1)* $w_1^j(\theta) = w_2^j(\theta)$ *for almost all* $\theta \in [0,1]$, $j = a,b$. *(2) Firms earn zero profits. (3)* $\xi_1^j(w_1^j, w_1^j) = \xi_2^j(w_2^j, w_2^j) = t^j(\theta)$, $j = a,b$, *for almost all* $\theta \in [0,1]$, *where* $t^j(\cdot)$ *is the cutoff task assignment rule with critical value*

$\widehat{\theta}^j$. *(3) Let* t_i^j *denote the task assignment rule on the equilibrium path for firm group* $j = a, b$ *and firm* $i = 1, 2$. *Then, there exists* $\widehat{\theta}^a$ *and* $\widehat{\theta}^b$ *such that the optimal task assignment rule for group* i *has the cutoff property with critical value* $\widehat{\theta}^i$.

Proof. (1) Equality of wages follows easily observing that if one firm offer higher wages to a positive mass of workers, it could profitably deviate by reducing it. (2) Given constant returns to scale, if firms earned positive profits, one could profitably deviate by reducing the entire wage schedule by a small amount so as to capture the entire labor supply and double profits. (3) Finally, given that firms offer the same wage schedule, an argument similar to the one used in the proof of Proposition 1, Lemma 13 and 14 shows that (2.26) has a unique solution $\widehat{\theta} = (\widehat{\theta}^a, \widehat{\theta}^b)$, and the optimal task assignment rule is the cutoff rule $t^j(\cdot)$ with critical value $\widehat{\theta}^j$, $j = a, b$. ■

Lemma 20. *Suppose* $\left\langle w_1^j, w_2^j \right\rangle_{j=a,b}$ *is a pair of equilibrium wage schedules and* $\widehat{\theta} = (\widehat{\theta}^a, \widehat{\theta}^b)$ *is the solution to (2.26). Then there is a pair* $\langle k_c^j,\ k_s^j \rangle$ *for each group* $j = a, b$ *such that* $w_i^j(\theta) = k_s^j$ *for almost all* $\theta \leq \widehat{\theta}^j$ *and* $w_i^j(\theta) = k_c^j p(\theta, \pi^j)$ *for almost all* $\theta > \widehat{\theta}^j$, $j = a, b$.

Proof. $w_i^j(\theta) = k_s^j$ for almost all $\theta \leq \widehat{\theta}^j$, $j = a, b$ follows from the same argument used in Lemma 15. To prove that $w_i^j(\theta) = k_c^j p(\theta, \pi^j)$ for almost all $\theta > \widehat{\theta}^j$ the argument is more cumbersome: we now have to make sure that deviations don't affect the affirmative action constraint, which complicates the arguments. Suppose for contradiction that in the candidate best response wages the ratio $w_i^j(\theta)/p(\theta, \pi^j)$ is not constant in θ for one group which we w.l.o.g take to be group a. Then we can find a positive measure set $\Theta^a \subset [\widehat{\theta}^a, 1]$ such that $w_i^a(\theta)/p(\theta, \pi^a) > w_i^a(\theta')/p(\theta', \pi^a)$ for all $\theta \in \Theta^a$, $\theta' \in [\widehat{\theta}^a, 1] \backslash \Theta^a$. It is always possible to choose Θ^a small enough so that there exists $\Theta^b \subset [\widehat{\theta}^b, 1]$ such that $\lambda^a \int_{\theta \in \Theta^a} f_{\pi^a}(\theta) d\theta = \lambda^b \int_{\theta \in \Theta^b} f_{\pi^b}(\theta) d\theta$ (i.e. the mass of workers in the two sets is the same) and $w_i^b(\theta)/p(\theta, \pi^b) \geq w_i^b(\theta')/p(\theta', \pi^b)$ for almost all $\theta \in \Theta^b$, $\theta' \in [\widehat{\theta}^b, 1] \backslash \Theta^b$. Now, suppose one firm posts zero wage to workers belonging to sets Θ^a and Θ^b. By construction, affirmative action constraint remains satisfied and qualified workers have been reduced by $R_C = \sum_{i=a,b} \lambda^i \int_{\theta \in \Theta^i} \pi^i f_q(\theta) d\theta$. The proposed deviation consists in "firing" workers belonging to sets Θ^a and Θ^b while reducing proportionally workers in the simple task to keep the factor ratio at the same level of the candidate equilibrium. Any candidate equilibrium must involve zero profits, but since wage bill per unit of production is lower, profits must be positive after the deviation so that the deviation is profitable. Formally, let C and S be the total factor inputs respectively in the complex and simple task. To keep the factor ratio constant, the deviation must reduce workers in the simple task by $R_S = S \cdot R_c/C$. Because of the affirmative action constraint, reduction of workers in the simple task must be proportionally distributed between groups. Compute then θ^a_{dev} and θ^b_{dev} to satisfy $\int_{[0,\theta^j_{dev}]} f_{\pi^j}(\theta)d\theta = \lambda^j R_s$, $j = a, b$ (we also have to make sure that $\theta^j_{dev} < \widehat{\theta}^j$. $j = a, b$ which is guaranteed by choosing Θ^a small enough). Consider the following deviation from the candidate equilibrium wage profile $w^j(\theta)$, $j = a, b$:

$$w^j_{dev}(\theta) = \begin{cases} 0 & \text{if } \theta \in \Theta^j \cup [0, \theta^j_{dev}], \\ w^j(\theta) & \text{otherwise} \end{cases} \quad j = a, b \tag{2.77}$$

By construction, the factor ratio remains constant. Using constant returns to scale, production decreases by $R_c/C = R_s/S$. It is now intuitive but cumbersome to show that average wage per unit of production decrease, so that the deviation is profitable. In the simple task average wage per worker is constant by the first part of this lemma. In the complex task, define the average wage per qualified worker in the candidate equilibrium as $\overline{k}^j_c = \int_{\theta\in[\widehat{\theta}^j,1]} w(\theta)f_{\pi^j}(\theta)d\theta/[\pi^j(1 - F_q(\widehat{\theta}^j)]$.

Similarly, define $\overline{k}^j_{dev} = \int_{\theta\in[\widehat{\theta}^j,1]\setminus\Theta^j} w^j_{dev}(\theta)f_{\pi^j}(\theta)d\theta / \int_{\theta\in[\widehat{\theta}^j,1]\setminus\Theta^i} \pi^j f_q(\theta)d\theta$, the average wage under the proposed deviation. Using $w(\theta)f_{\pi^i}(\theta) = \pi^i f_q(\theta) \cdot w(\theta)/p(\theta, \pi^j)$ and $w^j_{dev}(\cdot) = w^j_i(\cdot)$ for $\theta \in [\widehat{\theta}^j, 1]\setminus\Theta^j$, $j = a, b$ we can rewrite average wages and derive the following inequality from the fact that the ratio $w(\cdot)/p(\cdot, \pi^j)$ is higher for $\theta \in \Theta^j$, $j = a, b$

$$\frac{\int_{\theta\in[\widehat{\theta}^j,1]\setminus\Theta^j} \pi^j f_q(\theta)\frac{w(\theta)}{p(\theta)}d\theta + \int_{\theta\in\Theta^j} \pi^j f_q(\theta)\frac{w(\theta)}{p(\theta)}d\theta}{\int_{\theta\in[\widehat{\theta}^j,1]\setminus\Theta^j} \pi^j f_q(\theta)d\theta + \int_{\theta\in\Theta^j} \pi^j f_q(\theta)d\theta} > \frac{\int_{\theta\in[\widehat{\theta}^j,1]\setminus\Theta^j} \pi^j f_q(\theta)\frac{w(\theta)}{p(\theta)}d\theta}{\int_{\theta\in[\widehat{\theta}^j,1]\setminus\Theta^j} \pi^j f_q(\theta)d\theta} \tag{2.78}$$

i.e. $\overline{k}^j > \overline{k}^j_{dev}$, $j = a, b$. But then the total wage bill payed to the complex task workers is equal to $W_{dev} = \sum_i \overline{k}^i_{dev} \int_{\theta\in[\widehat{\theta}^i,1]\setminus\Theta^i} \pi f_q(\theta)d\theta < (1 - R_c)\sum_i \overline{k}^i \int_{\theta\in[\widehat{\theta}^i,1]} \pi f_q(\theta)d\theta$. Since wages decrease proportionally more than production, $w^j_{dev}(\cdot)$ implies positive profits, and the deviation is profitable. ■

Lemma 21. *The equilibrium wage schedules $\left\langle w^j_1, w^j_2 \right\rangle_{j=a,b}$ are continuous at almost all $\theta \in [0, 1]$.*

Proof. In Lemma 20 we established that $w^j_i(\theta) = k^j_s$ for almost all $\theta \leq \widehat{\theta}^j$ and $w^j_i(\theta) = k^j_c p(\theta, \pi^j)$ for almost all $\theta > \widehat{\theta}^j$, $j = a, b$. All we need to show is that $k^j_s = k^j_c p(\widehat{\theta}^j, \pi^j)$, $j = a, b$. The proof is by contradiction and consists of two parts: either there is a group with $k^j_s > k^j_c p(\widehat{\theta}^j . \pi^j)$, or $k_s \leq k^j_c p(\widehat{\theta}^j, \pi^j)$ for both groups, with strict inequality for at least one group.

Consider the first case. Then, there exists a positive measure interval $[\widehat{\theta}^j, \theta^*]$ such that $k_s > k^j_c p(\theta, \pi^j) = w^j_i(\theta)$ for almost all $\theta \in [\widehat{\theta}^j, \theta^*]$. In the proof of the necessity part of Proposition 4 we have already shown that there exists a profitable deviation from such a wage schedule. The deviation consists in offering a slightly higher wage to workers in the interval $[\widehat{\theta}^j, \theta^*]$, and use them to replace an equal mass of workers in the simple task who receive an higher wage k_s. Notice that his deviation does not affect the affirmative action constraint since the deviation firm is replacing expensive simple task workers with an equal mass cheaper workers stolen from the other firm.

Suppose instead $k_s^j \leq k_c^j p(\widehat{\theta}^j, \pi^j)$ for $j = a, b$ with strict inequality for at least one group, say group a. We propose a deviation on wages of both groups that keeps production constant and maintains the affirmative action constraint satisfied. Take and define $\theta^{a''} \in (\theta^{a'}, \widehat{\theta}^a)$ as the value that divides the mass of workers with $\theta \in [\theta^{a'}, \widehat{\theta}^a]$ in two equal parts, i. e. $\theta^{a''}$ is the solution of the following equation: $[F_{\pi^a}(\widehat{\theta}^a) - F_{\pi^a}(\theta^{a'})]/2 = [F_{\pi^a}(\theta^{a''}) - F_{\pi^a}(\theta^{a'})]$. In the proposed deviation, workers with $\theta \in [\theta^{a'}, \theta^{a''})$ will be assigned to the simple task, and workers with $\theta \in [\theta^{a''}, \widehat{\theta}^a]$ to the complex task. We want to keep mass of productive workers in the complex task constant. For this purpose, compute $\theta^{a*} > \widehat{\theta}^a$ so that $\int_{\theta^{a''}}^{\widehat{\theta}^a} f_q(\theta)d\theta = \left(\int_{\widehat{\theta}^a}^{\theta^{a*}} f_q(\theta)d\theta\right)/2$ and consider the following deviation:

$$w_{dev}^a(\theta) = \begin{cases} k_s^a + \epsilon & \text{for } \theta \in [\theta^{a'}, \widehat{\theta}^a] \\ 0 & \text{for } \theta \in [\widehat{\theta}^a, \theta^{a*}] \\ w_i^a & \text{otherwise} \end{cases} \tag{2.79}$$

$$t_{dev}^a(\theta) = \begin{cases} 0 & \text{for } \theta \in [0, \theta^{a''}] \\ 1 & \text{for } \theta \in (\theta^{a''}, 1] \end{cases}$$

By construction, the mass of workers employed in the simple task and the mass of qualified workers employed in the complex task are unchanged. On the other hand, if we consider this deviation alone, the affirmative action constraint will not be satisfied because the mass of workers employed in the complex task is increased by the following amount (a formal argument is omitted, but it will be symmetric to the argument used in Lemma 13 showing that employing the same mass of workers with higher average θ increases the mass of qualified workers):

$$\Psi' - \Psi = \lambda^a \left(\int_{\theta^{a''}}^{\widehat{\theta}^a} f_{\pi^a}(\theta)d\theta - \frac{\int_{\widehat{\theta}^a}^{\theta^{a*}} f_{\pi^a}(\theta)d\theta}{2} \right) \tag{2.80}$$

To keep the affirmative action constraint unchanged we have to deviate also on wages offered to group b. The idea is to compute $\theta^{b'}$ and θ^{b*} with $\widehat{\theta}^b < \theta^{b'} < \theta^{b*} < 1$ to attract from the other firm workers with $\theta \in [\widehat{\theta}^b, \theta^{b'})$, get rid of workers with $\theta \in [\theta^{b*}, 1]$ so as to keep the mass of productive workers employed in the complex task constant (equation (2.81)) and to satisfy the affirmative action constraint (equation (2.82)). Formally, compute $\theta^{b'}$ and θ^{b*} in order to satisfy the following set of equations:

$$\int_{\widehat{\theta}^b}^{\theta^{b\prime}} f_q(\theta)d\theta = \int_{\theta^{b*}}^{1} f_q(\theta)d\theta \tag{2.81}$$

$$\int_{\widehat{\theta}^b}^{\theta^{b\prime}} f_{\pi^b}(\theta)d\theta - \int_{\theta^{b*}}^{1} f_{\pi^b}(\theta)d\theta = \frac{\Psi' - \Psi}{\lambda^a} \tag{2.82}$$

(choosing $\theta^{a\prime}$ close enough to $\widehat{\theta}^a$ guarantees existence of $\theta^{b\prime}$ and θ^{b*} solving the system of equations). Consider deviation $\langle w^a_{dev}, t^a_{dev}\rangle$ together with the following deviation:

$$w^b_{dev}(\theta) = \begin{cases} k^b_s + \epsilon & \text{for } \theta \in [\widehat{\theta}^b, \theta^{b\prime}] \\ 0 & \text{for } \theta \in [\theta^{b*}, 1] \\ w^b_i & \text{otherwise} \end{cases} \tag{2.83}$$

Since the proposed deviations make sure that production remain constant, change in profits will depend only on change in wages. Letting ϵ terms go to zero, we have:

$$\begin{aligned} \lim_{\epsilon \downarrow 0} \Delta w_{dev} &= \frac{1}{2}\int_{\theta^{a\prime}}^{\widehat{\theta}^a} w^a_i(\theta) f_{\pi^a}(\theta)d\theta - \frac{1}{2}\int_{\widehat{\theta}^a}^{\theta^{a*}} w^a_i(\theta) f_{\pi^a}(\theta)d\theta \\ &+ \frac{1}{2}\int_{\widehat{\theta}^b}^{\theta^{b\prime}} w^b_i(\theta) f_{\pi^b}(\theta)d\theta - \frac{1}{2}\int_{\theta^{b*}}^{1} w^b_i(\theta) f_{\pi^b}(\theta)d\theta \end{aligned} \tag{2.84}$$

Observe now that if $\theta^{a\prime}$ is close enough to $\widehat{\theta}^a$ there is $h < k^a_c$ satisfying $k^a_s + \epsilon < hp(\theta, \pi^a)$ for every $\theta \in [\theta^{a\prime}, \widehat{\theta}^a]$. We can then conclude using the usual relation $f_{\pi^j}(\cdot) = \pi^j f_q(\cdot)$:

$$\lim_{\epsilon \downarrow 0} \Delta w_{dev}(\epsilon) < \frac{1}{2}\int_{\theta^{a\prime}}^{\widehat{\theta}^a} h\pi^a f_q(\theta)d\theta - \frac{1}{2}\int_{\widehat{\theta}^a}^{\theta^{a*}} k^a_c \pi^a f_q(\theta)d\theta + \frac{1}{2}k^b_c \pi^b \left(\int_{\widehat{\theta}^b}^{\theta^{b\prime}} f_q(\theta)d\theta - \int_{\theta^{b*}}^{1} f_q(\theta)d\theta\right) \tag{2.85}$$

The last term on the right hand side is equal to zero by construction, so the expression reduces to $2\lim_{\epsilon\downarrow 0} \Delta w_{dev}(\epsilon) < \pi^a(h - k^a_c)\int_{\theta^{a\prime}}^{\widehat{\theta}^a} f_q(\theta)d\theta < 0$

We can then choose ϵ small enough so that the change in wages is negative and the deviation is profitable. ■

Proof of Proposition 8 (necessity) Using the result shown in Lemma 21, the total wage bill paid to workers of group j is equal to $k^j_c = [p(\widehat{\theta}^j, \pi^j)F_{\pi^j}(\widehat{\theta}^j) + 1 - F_q(\widehat{\theta}^j)]$ and is strict increasing in k^j_c. As shown in the proof of sufficiency part of this proposition, $k^a_c = k^b_c = y_1(C, S)$ implies zero profits. To show that a wage schedule with $k^j_c \neq y_1(C, S)$ cannot be an equilibrium, suppose for example that $k^a_c > y_1(C, S)$. Then zero profits condition implies $k^b_c < y_1(C, S)$. Then we can construct a profitable deviation that deals only with workers assigned to the complex task. The idea of the deviation comes from the observation that the cost of labor per productive worker is higher in group a than in group b. We substitute high test result workers in group a with low test result workers of the same group, taking care of keeping the number of workers employed in the skilled task constant this will reduce the number of qualified workers of group a; if we construct a symmetric deviation for group b in order to restore the original mass of qualified workers, then the total wage bill will be lower than in the candidate equilibrium without changing total production and keeping the affirmative action constraint satisfied. Formally, define $\theta^{a\prime}$ and θ^{a*} with $\widehat{\theta}^a < \theta^{a\prime} < \theta^{a*} < 1$ as the solution of the following equation:

$$F_{\pi^a}(\theta^{a\prime}) - F_{\pi^a}(\widehat{\theta}^a) = 1 - F_{\pi^a}(\theta^{a*}) \tag{2.86}$$

Consider the following deviation of the wage function for group a:

$$w^a_{dev}(\theta) = \begin{cases} w^a_i(\theta) + \epsilon & \text{for } \theta \in [\widehat{\theta}^a, \theta^{a\prime}] \\ 0 & \text{for } \theta \in [\theta^{a*}, 1] \\ w^a_i(\theta) & \text{otherwise} \end{cases} \tag{2.87}$$

The mass of qualified workers of workers that belong to group a decrease since we substitute workers with high test result with an equal mass of workers with low test result (the formal proof is similar to the one used in Lemma 13). We can quantify the loss in productive workers employed in the complex task as $C' - C = \frac{\lambda^a}{2}\left(\int_{\theta^{a*}}^{1} \pi^a f_q(\theta) - \int_{\widehat{\theta}^a}^{\theta^{a\prime}} \pi^a f_q(\theta)\right)$: Next, consider the following deviation for group b:

$$w^b_{dev}(\theta) = \begin{cases} 0 & \text{for } \theta \in [\widehat{\theta}^b, \theta^{b\prime}] \\ w^b_i(\theta) + \epsilon & \text{for } \theta \in [\theta^{b*}, 1] \\ w^b_i(\theta) & \text{otherwise} \end{cases} \tag{2.88}$$

With $\theta^{b\prime}$ and θ^{b*} satisfying the following two equations:

$$F_{\pi^b}(\theta^{b\prime}) - F_{\pi^b}(\widehat{\theta}^b) = 1 - F_{\pi^b}(\theta^{b*}) \tag{2.89}$$

$$\int_{\theta^{b*}}^{1} \pi^b f_q(\theta) - \int_{\widehat{\theta}^b}^{\theta^{a\prime}} \pi^b f_q(\theta) = \frac{2}{\lambda^b}(C' - C) \tag{2.90}$$

Notice that if $\theta^{a\prime}$ is chosen to be close enough to $\widehat{\theta}^a$ (which implies $C' - C$ close enough to zero) then a solution to the system of equations specified above exists). Equation (2.89) guarantees that the number of employed workers remains constant, and (2.90) that the gain in productive workers employed in the complex task obtained with deviation $w^{b\prime}_i(\cdot)$ is equal to the loss due to $w^{a\prime}_i(\cdot)$.

Under the proposed deviations, productions remains constant, so difference in profits depends uniquely on difference in wage bill:

$$\begin{aligned} w(\epsilon) - w &= \frac{\lambda^a}{2}\int_{\widehat{\theta^a}}^{\theta^{a\prime}} \left(w_i^a(\theta) + 2\epsilon\right) f_{\pi^a}(\theta)d\theta - \frac{\lambda^a}{2}\int_{\theta^{a*}}^{1} w_i^a(\theta) f_{\pi^a}(\theta)d\theta \\ &\quad -\frac{\lambda^b}{2}\int_{\widehat{\theta^b}}^{\theta^{b\prime}} w_i^b(\theta) f_{\pi^b}(\theta)d\theta + \frac{\lambda^b}{2}\int_{\theta^{b*}}^{1} \left(w_i^b(\theta) + 2\epsilon\right) f_{\pi^b}(\theta)d\theta \\ &= k_c^a(C' - C) - k_c^b(C' - C) + \epsilon\left[\lambda^a \int_{\widehat{\theta^a}}^{\theta^{a\prime}} \pi^a f_q(\theta)d\theta + \lambda^b \int_{\theta^{b*}}^{1} \pi^b f_q(\theta)d\theta\right] \end{aligned} \tag{2.91}$$

Since $k_c^b < y_1(C, S) < k_c^a$ and by construction $C' - C < 0$, then there is an ϵ small enough such that the difference in wage bill is negative and the deviation is profitable. ■

3. STATISTICAL DISCRIMINATION AND EFFICIENCY

Peter Norman*

University of Pennsylvania

July 16, 1997

Abstract

This paper seeks to investigate if there is an efficiency rationale for policies aimed at statistical discrimination. To do this we consider a model of statistical discrimination with imperfectly observable human capital investments by the workers. In this setup we contrast the equilibria of the competitive model with the allocation chosen by a utilitarian social planner. The planner works under the same informational constraints as the firms in the competitive model and is furthermore constrained to use reward schemes that satisfy incentive compatibility, individual rationality and budget balance.

Any equilibrium is constrained sub-optimal. Hence, there is an efficiency rationale for intervention, no matter whether the equilibrium under consideration involves discrimination or not. However, since this inefficiency arises because a "free riding" problem occurs when human capital investments are imperfectly observable this does not mean that statistical discrimination is inefficient. On the contrary, the planner may actually want to discriminate between two groups of ex ante identical workers. The reason is that by designating one group to be investors and increasing the fraction of investors within this group compared to the best non-discriminatory allocation the "mismatch" is reduced: fewer workers with the investment end up performing jobs where the investment is unnecessary and fewer workers without the investment is allocated to tasks where

[0]I thank Stephen Coate, Stephen Morris, Andrea Moro and Andrew Postlewaite and seminar participants at the Summer School in Tel-Aviv for comments, discussions and encouragements. The usual disclaimer applies.

the investment is needed. This positive effect on social surplus has to be contrasted with the fact that discrimination creates inefficiencies in investment behavior. It is shown by example that it is quite possible that the positive effect dominates, so the solution to the planning problem actually involves discrimination.

3.1. Introduction

In this paper I seek to investigate if there is an efficiency rationale to intervene in an economy with statistical discrimination. Although this question seems to be of importance for more concrete policy analysis, such as evaluation of affirmative action programs, there are few attempts in the literature to analyze the question in detail.

In the earliest theories of discrimination, building on the seminal contribution by Becker[9], discrimination is explained from preferences. While a planner may potentially disagree with the individual agents' racist/sexist preferences there is nothing inefficient with discrimination in Becker's model.

These preference based models of discrimination were challenged in the seventies by theories of *statistical discrimination,* which can explain discrimination without resorting to an assumption of prejudice. This literature was started by Phelps[24] and Arrow [4] and further developed by Aigner and Cain [3] and others. The main idea in these models is that workers' abilities are imperfectly observable, so firms have to make inferences based on whatever observables they have available. Now if, as Phelps [24] and Aigner and Cain [3] assumes, traditional measures of productivity are less informative for minorities this means that relatively few minority workers are in the top end of the wage distribution, even if the distributions of abilities are identical for the two groups.

Models of statistical discrimination with exogenously determined ability distributions can explain differential treatment in the labor market even when firms maximize profits and there are no differences in abilities across groups. However, this does not mean that discrimination is a market failure. On the contrary, firms simply respond to the differences in screening technologies for the different groups. Given that there are allocative benefits to have higher skilled workers in more advanced jobs, statistical discrimination improves allocative efficiency.

If skills are acquired by the workers through costly human capital investment, equilibria with statistical discrimination may, as Lundberg and Startz [19] observed, be inefficient. The potential inefficiency is that low cost agents from the discriminated group may choose not to invest while high cost agents from the other group invest. Hence it may be efficiency enhancing to enforce an equal opportunities policy where firms are not allowed to make wages dependent on group identity.

A similar exercise can be found in Schwab [27] who demonstrates that an equal opportunities policy may result in an increase in social surplus even if agents from one group are on average more productive than the other. From this Schwab draws the conclusion that there are circumstances under which statistical discrimination is inefficient. However, it is not clear whether this is the most natural interpretation. Schwab uses a model similar to Akerlof's [1] market for lemons and in the equilibrium when firms are allowed to practice statistical discrimination too few workers from both groups enter the "standardized" labor market and too many workers from both groups stay in the "individualized" market. When firms are not allowed to treat the groups differently this inefficiency is increased for one group and decreased for the other. The net effect on social surplus is ambiguous in general and with the right choice of supply elasticities the policy may increase surplus. Still, it's hard to see why one should interpret this as saying that statistical discrimination is a market failure when what is driving the result is that the usual lemons problem is less severe in one group than the other. Also, for statistical discrimination to arise in Schwabs model there must be intrinsic productivity differences between groups. But this has the consequence that groups are treated differently also in the surplus maximizing allocation.

Underlying the analysis in this paper is the view that for any given positive model of the labor market there is a natural welfare analysis based on a planner with no more information than the firms on the market. This view leads us to compare equilibrium outcomes of a decentralized model with the solution to a corresponding planning problem rather than comparing equilibrium outcomes under two different regimes, one where statistical discrimination is allowed and one where it is not. While arguments of course always can be made about the planners objective, informational assumptions etcetera, this approach has certain advantages. In particular, the problem with pairwise comparisons of different regimes is usually that there may be other, better policies that are "as easy to implement" as the ones under consideration. This problem is mitigated by using the methodology in this paper since the planners' feasible set is determined by more primitive assumptions.

The decentralized model is the same as in Moro and Norman [23]. In this model human capital investments increase productivity only for workers that are employed to perform a qualified task (management), while workers in the other task (manual labor) are equally productive irrespective of whether they invested in their human capital or not. The investment decision is unobservable for the firms, but firms observe a noisy signal with some information value and make wage offers on the basis of this signal and, potentially, group identity. No asymmetries between groups are assumed: the model abstracts from differences between workers in intrinsic abilities and the distributions of investment costs and the signaling technology are the same for both groups. Still, discrimination may occur in equilibrium. The reason for this is that when the fraction of investors is increased in one of the groups, firms will respond by assigning more workers to the qualified task. Given that the tasks are complementary in production, this decreases the marginal productivity in the qualified task relative the other. In equilibrium wages are given by expected marginal productivities, so the consequence for a worker from a group where investment behavior is unchanged is unambiguously to decrease the incentives to invest. Hence, the complemetarities in production creates *strategic* complemetarities between groups that work "in favor of" discrimination.

This decentralized model is contrasted with a command economy where a utilitarian social planner has available the same production and screening technology as the firms in the decentralized model. Since human capital investments are unobservable, the planner must resort to using reward schemes depending on for the planner observable variables: the noisy signal and group identity. Besides the usual incentive compatibility constraints we also assume that the planner has to respect an individual rationality constraint and budget balance.

It is shown that there are circumstances where the solution to this planning problem involves discrimination. The intuitive reason for this is that there is an element of economics of specialization involved. By designating one group as investors and the other group as non-investors the planner is able to reduce the matching problem. That is, fewer investors will be assigned to the unskilled task where the investment is not needed and fewer non-investors will be assigned to the skilled task. However, this is not without a cost. Given that the distributions of investment costs are the same across groups discrimination raises the investment cost of the average investor if the total quantity of investors is kept constant. Hence, the planner faces a trade-off between the informational gains and the losses in terms of increased investment costs and examples can be constructed where the net effect goes either way.

Comparing the decentralized model with the planning problem we find that there is underinvestment in equilibrium. This is hardly surprising: since human capital investments are imperfectly observable a "free riding" problem occurs. The more people that invest, the higher is the probability that any worker with a particular signal has invested. Hence, non-investors will be more likely to pass as "probable investors" when the fraction of investors increase. In fact, all equilibria are constrained sub-optimal and can be improved upon by a small increase in the fraction of investors for at least one of the groups. The only case when a small increase in the fraction of investors does not increase surplus is if the fraction of investors in a group where all workers are assigned to the unqualified job is increased. In this case the surplus would actually be reduced since the increased investment costs would not be offset by an increase in productivity when these workers all are assigned to the job where the investment is not needed. For all interior equilibria however, with or without discrimination, surplus is always increased when the fraction of investors in any of the groups increases.

The model is admittedly very stylized. However, neither the "informational gains" of discrimination nor the underinvestment seem to be artifacts of any simplifying assumptions. If this is true, it means that normative analysis in models of statistical discrimination with endogenous human capital formation will be difficult for two reasons. First of all, discrimination may actually improve allocative efficiency and even if the distributional consequences are undesirable, it is not inconceivable that this may be solved by transfers. Secondly and, maybe, more important, even if the constrained optimal allocation is non-discriminatory we still have to handle the fact that there is underinvestment in equilibrium. Thus, while we would want to pose the question "which is the best way to eliminate or limit discrimination" we cannot separate policies aimed at discrimination and policies aimed at the underinvestment problem.

There are, of course, good reasons to be careful when interpreting the results. Nevertheless, the very same assumptions that make discrimination of ex ante identical groups a possible equilibrium outcome also create potential efficiency gains of discrimination.

Since there is no correlation between group identity and investment costs, the model does not offer any prediction about which group should specialize in human capital investment. This is clearly an idealization and if there would be the slightest correlation, the planner would no longer be indifferent. However, this wouldn't change the conclusion that specialization in human capital acquisition has potential welfare benefits.

The rest of the paper is organized as follows: Section 2 describes the model and in section 3 we consider a planning problem and characterize the solutions to this problem. In section 4 the surplus-maximizing allocations are contrasted with the equilibria of the "laissez faire" model and in section 5 we consider simple transfer instruments that are implementable under weaker informational assumptions than these needed in order to implement the solutions to the planning problem. Section 6, finally, contains a discussion.

3.2. The Model

We will work with the same production and screening technology as in Moro and Norman [23] and the decentralized model is identical to the model considered there. For expositional purposes we will begin with a somewhat informal review of the model. Details can be found in Moro and Norman [23].

3.2.1. The production technology

Output is produced from labor input in two jobs, which we refer to as the *complex task* and the *simple task*. The effective input of labor in these task will be denoted by C and S respectively and we assume that output is produced from these inputs according to a production function $y : R^2_+ \rightarrow R$, which satisfies all standard neoclassical assumptions: y is twice continuously differentiable, strictly quasi-concave, exhibits constant returns to scale, both factors are essential and the Inada conditions are satisfied.

The crucial assumption is that *only those workers who made an investment in their human capital are able to perform the complex task satisfactorily.* Specifically, we assume that a worker without the human capital investment does not contribute at all to the effective input of labor in the complex task while all workers with the investment gives equal contributions, normalized to unity. The effective input of labor in the complex task will thus be given by the quantity of workers employed in the task who invested. In the simple task the investment does not matter for productivity, so here we simply add up the number of workers to get the effective input of labor. All this can be relaxed considerably without changing anything qualitative. What is needed is really only that the investment is more important for one of the tasks than the other. At the cost of some additional complexity we can also handle heterogeneities between workers in intrinsic abilities to perform the complex task.

3.2.2. Human capital investments and the screening technology

There are two groups of workers, indexed by a and b respectively. Each group consists of a continuum of workers with heterogenous costs of investment in their human capital. An agent is thus characterized by her (unobservable) cost of investment, denoted by c, and her (observable) group identity, which is either a or b. A worker who undertakes the investment incurs the cost c, but becomes able to perform the complex task. Workers who don't invest incur no cost, but are unable to perform the qualified task satisfactorily. We denote by λ^a the fraction of agents with observable characteristic a and by $\lambda^b = 1 - \lambda^a$ the fraction of agents with characteristic b. For both groups a and b, investment costs are distributed according to some continuous and strictly increasing distribution function G with support in the interval $[\underline{c}, \bar{c}]$[1]. We assume that $\underline{c} \leq 0$ and $\bar{c} > 0$.

Each worker's investment decision is unobservable, but there is a noisy signal $\theta \in \Theta$ that firms can use as an indicator of how likely it is that a certain worker is qualified. For most of the analysis we assume that Θ is the unit interval and that θ is distributed according to density f_q for a worker who invested in the first stage and according to density f_u for a worker who did not. These densities are assumed to be continuously differentiable and bounded away from zero we assume without loss of generality that f_u and f_q satisfies the (strict) *monotone likelihood ratio property*, i.e. $f_q(\theta) / f_u(\theta)$ is strictly increasing in θ. Thus, qualified workers are more likely to get higher values of θ than workers who are not qualified[2]. The cumulative densities are denoted by F_q and F_u respectively. It should be observed that the distribution of signals does not depend on group identity, so the screening technology is not biased in favor of any of the groups.

Since the investment decision is unobservable, wages can not be contingent on the investment decision but only on the observable signal. For simplicity we assume that the workers are risk neutral. We take the payoffs to the expected wage earnings net investment costs, which we can write as $E_{F_q}(w(\theta)) - c$ for a worker with investment cost c who invested and $E_{F_u}(w(\theta))$ for a worker who did not invest.

3.3. Optimal Policies in a Command Economy

In this section we will consider the problem of a utilitarian social planner operating in the economic environment described in Section 3.2. Since investment decisions are unobservable, the planner can control investment behavior only indirectly by designing reward schemes as a function of the noisy signal and group identity. We will also constrain the planner to use reward schemes that are non-negative and satisfy a budget balance condition.

[1] Thus, the two groups are identical in terms of intrinsic investment costs.

[2] These assumptions are without loss of generality. If the monotone likelihood ratio property would not hold the signals can always be relabeled.

3.3.1. The planning problem

In principle the planner can allocate workers between tasks using arbitrary (measurable) task assignment rules $t^j : \Theta \to \{0,1\}$. The interpretation is that $t^j(\theta) = 0$ $(= 1)$ means that a worker from group j with signal θ is assigned to the simple task (complex task). Given any pair of task-assignment rules and the fractions of investors in each group, which we denote by $\pi = (\pi^a . \pi^b)$, we can compute the input of labor in each task as[3]

$$\begin{aligned} C &= \sum_j \lambda^j \int t^j(\theta) \pi^j f_q(\theta) d\theta \\ S &= \sum_j \lambda^j \int (1 - t^j(\theta)) \left(\pi^j f_q(\theta) + (1-\pi^j) f_u(\theta)\right) d\theta. \end{aligned} \tag{3.1}$$

Using these expressions and the monotone likelihood ratio it is not difficult to prove that we may without loss of generality restrict attention to task assignment rules that satisfies a cut-off property. That is, we only need to find a "critical signal" θ^j for each group. Then, all workers with signals below will be assigned to the simple task and all workers with signals above to the complex task. The intuition behind this is simply that the higher the signal the higher is the probability that a particular worker has invested. Thus, if some low signal workers are in the complex task and some high signal workers are in the simple task we can increase the effective input of labor in the complex task while keeping the input of labor in the simple task constant by switching tasks for the right number of workers.

Taking this cutoff property of optimal task assignments as given and using $F_\pi(\theta)$ as shorthand notation for $\pi F_q(\theta) + (1-\pi) F_u$ we can write the problem for the utilitarian social planner as

$$\sup_{\{\pi^j, \theta^j, w^j(\cdot)\}} y\left(\sum_j \lambda^j \pi^j (1 - F_q(\theta^j)), \sum_j \lambda^j F_{\pi^j}(\theta^j)\right) - \sum_j \lambda^j \int_{\underline{c}}^{G^{-1}(\pi^j)} c g(c) dc \tag{3.2}$$

$$\begin{aligned} \text{s.t } \pi^j &= G\left(\int w^j(\theta)(f_q(\theta) - f_u(\theta)) d\theta\right) \quad \text{[IC]} \\ y(\cdot) &\geq \sum_j \lambda^j \left(\pi^j \int w^j(\theta) f_q(\theta) d\theta + (1-\pi^j) \int w^j(\theta) f_u(\theta) d\theta\right) \quad \text{[BB]} \\ w^j(\theta) &\geq 0 \text{ for all } \theta \quad \text{[IR]} \end{aligned}$$

[3]It should be noted that f_q and f_u are used as *frequency distributions* in the (3.1), i.e. we are *assuming* that a strong law of large numbers holds. For a simple stochastic model that has the property that the probability distributions facing the individual workers coincides with the realized frequency distributions we refer the reader to Moro and Norman [23]. More technical discussions about laws of large numbers for a continuum of random variables can be found in Judd [16] or Feldman and Gilles [14].

The reason for taking the supremum will be made clear below, but except that there are some cases when the maximum cannot be attained, the interpretation of the program (3.2) is pretty straightforward. The planner seeks to make total surplus as large as possible subject to some rather familiar looking constraints: the Incentive compatibility constraint reflects the assumption that the planner can not observe whether a particular worker has invested or not, so the fraction of investors will be determined as best responses to the wage schemes chosen by the planner[4]. The budget balance and individual rationality constraints have the obvious interpretations.

3.3.2. Characterization of the surplus maximizing plan

In characterizing the solutions to the problem (3.2) two auxiliary maximization problems will be used. Since the first of these problems is to maximize the objective in (3.2) without the constraints I will refer to this as the *unconstrained problem*. We write the problem as

$$\max_{\{\pi^j,\theta^j\}\in[0,1]^4} y\left(\sum_j \lambda^j \pi^j \left(1 - F_q\left(\theta^j\right)\right), \sum_j \lambda^j F_{\pi^j}\left(\theta^j\right)\right) - \sum_j \lambda^j \int_{\underline{c}}^{G^{-1}(\pi^j)} cg(c)\,dc. \tag{3.3}$$

While the objective function is strictly concave in θ there are no simple conditions that guarantees that the function is concave in (θ, π). Thus, all we know a priori is that a solution exists and that any solution must satisfy the first order conditions. It should be noted that the problem (3.3) is introduced purely for analythical purposes and doesn't have any natural economic interpretation[5].

Note that the constraints of (3.2) are "minimal" in the sense that by throwing any of the constraints there are always wage schemes implementing any feasible solution of (3.3). Obviously, if the constraint [IC] would be deleted, anything could be "implemented" by $w^j(\theta) = 0$ for all θ. It is also easy to check that it is possible to implement anything if the constraint [BB] is delated and the reason is simply that the benefits to investment can be made arbitrarily large by making the wage for the highest signals sufficiently high. To see that [IR] constraint is needed for the other constraints to have any bite we just consider simple wage schemes of the form

$$w^j(\theta) = \begin{cases} w_L^j & \text{for } \theta < \theta^j \\ w_H^j & \text{for } \theta \geq \theta^j \end{cases}. \tag{3.4}$$

[4]Recall that the payoff for an investor is $E_{F_q}(w(\theta)) - c$ and the payoff for a non-investor is $E_{F_u}(w(\theta))$. In order for a worker to behave rationally given wage schedule w the worker must invest if and only if $E_{F_q}(w(\theta)) - c \geq E_{F_u}(w(\theta))$. Hence $G\left(\int w(\theta)(f_q(\theta) - f_u(\theta))\,d\theta\right)$ is the fraction of workers who rationally choose to invest given wage schedule w. Now, with the different groups facing different wage schemes we get the constraints [IC] in (3.2).

[5]To make sense of the problem (3.3) we would have to justify why a planner who can control investment behavior without any incentive constraints doesn't know who the investors are when allocating workers between tasks.

Note that the benefits of investment given these wages are $\left(w_H^j - w_L^j\right)\left(F_u\left(\theta^j\right) - F_q\left(\theta^j\right)\right)$. Then if we let

$$w_L^j = -\frac{\pi^j\left(1-F_q\left(\theta^j\right)\right)+\left(1-\pi^j\right)\left(1-F_u\left(\theta^j\right)\right)}{\pi^j F_q\left(\theta^j\right)+\left(1-\pi^j\right)F_u\left(\theta^j\right)} w_H^j \tag{3.5}$$

it is easy to verify that the total wage payments are zero for any choice of w_H^j, but that the benefits of investment goes to infinity as w_H^j goes to infinity. Hence any $\left(\pi^a, \pi^b\right)$ and $\left(\theta^a, \theta^b\right)$ can be implemented if [IR] is delated from the program.

Our second auxiliary problem is

$$\begin{aligned} & \sup_{w^j} \int w^j(\theta)\left(f_q(\theta) - f_u(\theta)\right) d\theta && (3.6) \\ \text{s.t. } y \geq & \sum_{k=a,b} \lambda^k \left(\pi^k \int w^k(\theta) f_q(\theta) d\theta + \left(1-\pi^k\right)\int w^k(\theta) f_u(\theta) d\theta\right) \\ w^j(\theta) \geq & \ 0 \\ & \pi^a, \pi^b \text{ given.} \end{aligned}$$

Here we ask how large can we make the benefits of investment for one of the groups, if production, the fraction of investors and the wage scheme for the other group is held fixed. For convenience we use the notation W^j for the total wage payments for group $j = a, b$, that is

$$W^j = \pi^j \int w^j(\theta) f_q(\theta) d\theta + \left(1-\pi^j\right)\int w^j(\theta) f_u(\theta) d\theta, \tag{3.7}$$

and let

$$b^{j*} \equiv \frac{\left(y - \lambda^k W^k\right)\left(f_q(1) - f_u(1)\right)}{\lambda^j\left(\pi^j f_q(1) + \left(1-\pi^j\right) f_u(1)\right)} \tag{3.8}$$

for $j \in \{a, b\}$ and $k \neq j$.

Lemma 22. *Fix π, y and w^k. The supremum of the benefits of investment for group j over wage schemes satisfying [IR] and [BB] defined in (3.6) is given by b^{j*}*

The proof is relegated to the appendix. To understand intuitively where (3.8) comes from it is useful to consider the case with a finite set of signals, $\Theta = \{\epsilon, 2\epsilon, ..., 1\}$. In this case it should not be too surprising that the way to maximize incentives is to reward only those workers who got the highest possible signal and give nothing to the workers with lower signals. The quantity of workers from group j with $\theta = 1$ is given by the denominator in (3.8) and they share $y - \lambda^k W^k$ (in per capita terms). It follows that the corresponding benefits of investment equals the right hand side in (3.8). In the continuum model this translates to b^{j*} being the lowest upper bound on the benefits of investment.

As is easily seen from (3.8), the higher is π^j or W^k, the lower is b^{j*}. That π^j affects "maximal" incentives negatively simply reflects that the number of agents who actually get rewarded increases when π^j increases, so with a fixed pie this means that the reward decreases. The effect from W^k is even more straightforward since the total resources spent on group j decreases with W^k and, for the same reason, b^{j*} is also increasing in y.

Now suppose that π^j satisfies

$$\pi^j = G\left(\frac{\left(y - \lambda^k W^k\right)\left(f_q(1) - f_u(1)\right)}{\lambda^j\left(\pi^j f_q(1) + \left(1-\pi^j\right) f_u(1)\right)}\right). \tag{3.9}$$

Recall that if group j faces a reward scheme w^j, the fraction of workers that invests as a best response is given by $G(\int w^j(\theta)(f_q(\theta) - f_u(\theta))d\theta)$. Combining with Lemma 22 and using the fact that b^{j*} is decreasing in π^j we can then show:

Lemma 23. *There is a unique solution to (3.9) and this solution is the least upper bound on the fraction of investors in group j satisfying [IC],[IR] and [BB] for any given y and W^k .*

The proof is in the appendix. By Lemma 22, reward schemes consistent with [IR] and [IC] can be constructed generating benefits of investments arbitrarily close to the argument of the right hand side in (3.9) if a fraction π^j invests. Since G is continuous this implies that a fraction of investors arbitrarily close to π^j is consistent with [IC],[IR] and [BB]. Next, suppose that something strictly larger could be implemented. Since b^{j*} is strictly decreasing in the fraction of investors this means that in order for [IR] and [BB] to be satisfied, the benefits of investment must be strictly lower than the unique solution to (3.9), which means that the fraction of investors consistent with incentive compatibility is strictly lower that π^j, a contradiction.

We are interested in under what conditions, if any, the solution to (3.3), the *unconstrained solution,* coincides with the optimal cutoffs and fractions for the full program (3.2). The following result gives a necessary and sufficient condition for when a *feasible* plan under program (3.3) can be implemented in the full program.

Proposition 10. *Let $\left\langle \pi^a, \pi^b, \theta^a, \theta^b \right\rangle$ be any feasible solution to (3.3) and let y be the output corresponding to this plan. Then there exists a pair of wage schemes $\left\langle w^a, w^b \right\rangle$ such that these together with $\left\langle \pi^a, \pi^b, \theta^a, \theta^b \right\rangle$ is a feasible solution for the full problem (3.2) if and only if there is some $\alpha \in [0,1]$ such that $G\left(\frac{\alpha y(f_q(1)-f_u(1))}{\lambda^a f_{\pi^a}(1)}\right) > \pi^a$ and $G\left(\frac{(1-\alpha) y(f_q(1)-f_u(1))}{\lambda^b f_{\pi^b}(1)}\right) > \pi^b$.*

The proof is in the appendix in Section 3.6.3.

To fix ideas we first consider the case when there is only a single group of workers. The planning problem (3.2) then simplifies to;

$$\sup_{\{\pi,\theta,w(\cdot)\}} y\left(\pi\left(1-F_q(\theta)\right),F_\pi(\theta)\right)-\int_{\underline{c}}^{G^{-1}(\pi)} cg(c)\,dc \tag{3.10}$$

$$\begin{aligned}
\pi &= G\left(\int w(\theta)\left(f_q(\theta)-f_u(\theta)\right)d\theta\right) \quad \text{[IC]}\\
y(\cdot) &\geq \pi\int w(\theta)\,f_q(\theta)\,d\theta+(1-\pi)\int w(\theta)\,f_u(\theta)\,d\theta \quad \text{[BB]}\\
w(\theta) &\geq 0 \text{ for all } \theta \quad \text{[IR]}
\end{aligned}$$

This problem can be interpreted either as the original problem with the restriction that both groups are treated equally (a restriction that may or may not bind depending on parameters) or as the planning problem corresponding to the case when there are no observable "irrelevant" group characteristics. We first note that no constraints to the problem are directly affected by the cutoffs for the task-assignment. This implies that task assignment will always be done so as to maximize production given the particular value of π chosen in optimum. Formally

Lemma 24. *Suppose that $\pi^*,\theta^*,w^*(\cdot)$ solves (3.10). Then*

$$\theta^*=\arg\max_\theta y\left(\pi^*\left(1-F_q(\theta)\right),F_{\pi^*}(\theta)\right)$$

The proof is simple and we leave out the details. Since no constraint depends on θ, the cutoff must be chosen so as to maximize the objective given π^*. Now, for each $\pi\in(0,1]$ it can be shown that there is a unique maximizer $\widetilde{\theta}(\pi)$[6], which is a continuously differentiable function of π. We let $Y(\pi)$ denote the highest achievable output for any fraction of investors π[7],

$$\begin{aligned}
Y(\pi) &= y\left(\pi\left(1-F_q\left(\widetilde{\theta}(\pi)\right),F_\pi\left(\widetilde{\theta}(\pi)\right)\right)\right)= \\
&= \max_\theta y\left(\pi\left(1-F_q(\theta),F_\pi(\theta)\right)\right).
\end{aligned} \tag{3.11}$$

and note that applying Lemma 23 for the special case with a single group of workers we find that the least upper bound for the fraction of investment for any given output y consistent with incentive compatibility, budget balance and individual rationality is the unique fixed point to

$$\pi=G\left(y\frac{f_q(1)-f_u(1)}{\pi f_q(1)+(1-\pi)f_u(1)}\right). \tag{3.12}$$

[6]See Moro and Norman [23] for formal arguments. The idea is that variables can be changed so that the problem is rewritten as $\max y(C,S)$ subject to $C\leq\pi\left(1-F_q(\theta)\right)$ and $S\leq F_\pi(\theta)$ we have a strictly quasi-concave objective function and, as a consequence of the monotone likelihood ratio, a (strictly) convex constraint set.

[7]While $\widetilde{\theta}(0)$ is not uniquely defined, we may just take it to be $\lim_{\pi\to 0}\widetilde{\theta}(\pi)$, which is *a* maximizer for $\pi=0$.

So we can collapse [IR],[IC] and [BB] into a single constraint and rewrite (3.10) as to find

$$\begin{aligned} \pi^* &\in \arg\sup_{\pi} Y(\pi) - \int_{\underline{c}}^{G^{-1}(\pi)} cg(c)\,dc \qquad (3.13) \\ \text{subj to } \pi &< G\left(Y(\pi)\frac{f_q(1)-f_u(1)}{\pi f_q(1)+(1-\pi)f_u(1)}\right). \end{aligned}$$

Since the constraint set is open, it is more or less direct that any (local as well as global) maximum must be a local maximum of the corresponding unconstrained problem[8]. The only situation when π^* is not a local maximum of the unconstrained problem is therefore when

$$\pi^* = G\left(Y(\pi^*)\frac{f_q(1)-f_u(1)}{\pi^* f_q(1)+(1-\pi^*)f_u(1)}\right), \qquad (3.14)$$

Although these observations are rather straightforward we will summarize them as a proposition, which will be useful in Section 4 when we compare the planning model with the decentralized model.

Proposition 11. *Suppose π^* solves (3.13). Then there are two possibilities;*

1. *π^* satisfies the first order condition $Y'(\pi^*) = G^{-1}(\pi^*)$*
2. *π^* is a solution to (3.14)*

The proof is in the text above the proposition.

Before going on to the case with multiple groups, the reader may note that there are two potential analythical difficulties with problem (3.13). The first is that $Y(\pi)$ may not be concave[9]. Maybe more disturbing is that *even if Y is concave* the constraint set may not be a convex set. As will be evident in Section 4 this creates some difficulties that I have only been able to solve partially. In particular, we cannot even in general rule out the possibility that π^* solving (3.13) is *larger* than the solution to the unconstrained problem.

For the case with multiple groups we can do the same exercise as with a single group, but now

$$\begin{aligned} Y(\pi) &= Y(\pi^a,\pi^b) = \max_{\theta^a,\theta^b} y\left(\sum_j \pi\left(1-F_q(\theta^j)\right), \sum_j F_\pi(\theta^j)\right) = \qquad (3.15) \\ &= y\left(\sum_j \pi\left(1-F_q\left(\tilde{\theta}^j(\pi)\right)\right), \sum_j F_\pi\left(\tilde{\theta}^j(\pi)\right)\right), \end{aligned}$$

[8]Since we assume that $G(0) \leq 0$ we can moreover rule out $\pi^* = 0$ as a candidate solution. No assumption made so far guarantees that $\pi^* < 1$, but since this case is not very interesting we will simply assume it away. What is needed is roughly that $G(\bar{c})$ is sufficiently large.

[9]I have not been able to come up with any easily interpratable sufficient conditions for concavity.

where $\widetilde{\theta}^j(\pi)$ are defined in analogy with the single-group model as the pair of unique output maximizing cut-offs. In this case partial corner solutions are not only possible, but will actually be rather interesting. However, the possibilities can still be broken down so that either the solutions satisfies the first order conditions for the unconstrained problem (including complementary slackness conditions) or the supremum is not achievable, in which case there exists some $\alpha \in [0,1]$ such that

$$\pi^{j*} = G\left(\frac{\alpha Y(\pi^*)(f_q(1) - f_u(1))}{\lambda^j(\pi^{j*} f_q(1) + (1-\pi^{j*}) f_u(1))}\right), \tag{3.16}$$

for $j = a, b$.

While these are only necessary conditions we will see that these are actually very useful when comparing the planning problem with the "laissez faire" economy in Section 4.

3.3.3. Example: Discrimination By the Planner

The main point in this example is to demonstrate that a utilitarian social planner may want to discriminate against one of the groups, although the groups are ex ante identical. The intuition is straightforward; compare a situation where a positive fraction π invests in both groups with a situation where no worker from one of the groups and a fraction $\pi' > \pi$ in the other group invests. Moving from the symmetric to the discriminatory/specialized allocation the likelihood that a worker is assigned to the wrong task is decreased. On the other hand, to increase the fraction of investors in one of the groups the average investment cost must be increased, so the planner faces a trade-off. By discriminating the informational problem becomes less severe, but it causes inefficiencies in investment behavior. The example is rather algebra intensive and can be skipped.

The logic of the exercise is as follows: first we constrain the planner to set $\pi^a = \pi^b$ and solve for the best "color blind" allocation. We then show that there exists a feasible plan with discrimination that gives a higher value of the planners' objective and conclude that the optimal solution (which we don't solve for explicitly) must involve discrimination.

We let the set of signals be given by $\Theta = \{\theta_L, \theta_H\}$ with conditional probability distributions;

$$\begin{array}{ccc} & \theta_L & \theta_H \\ \text{not inv} & \alpha & (1-\alpha) \\ \text{inv} & (1-\alpha) & \alpha \end{array}, \tag{3.17}$$

where $\alpha > 1/2$. The production technology is of Cobb-Douglas form, $y(C,S) = C^{\beta} S^{1-\beta}$ and investment costs are distributed according to $c \sim U[0,k]$. Thus, $G^{-1}(\pi) = k\pi$ and that total investment costs are

$$\int_{\underline{c}}^{G^{-1}(\pi)} cg(c)\,dc = \int_0^{k\pi} \frac{c}{k} dc = \left[\frac{c^2}{2k}\right]_0^{k\pi} = \frac{k\pi^2}{2}. \tag{3.18}$$

Depending on parameter values the solution to the problem to maximize total surplus (under the "equal treatments constraint") can be of one of three forms;

Type 1 A worker is employed in the complex task if and only if the worker has a high signal.

Type 2 All workers with $\theta = \theta_L$ and a fraction $\sigma > 0$ of the workers with $\theta = \theta_H$ in the simple task and a fraction $(1-\sigma)$ of the workers with the high signal in the complex task.

Type 3 All workers with $\theta = \theta_H$ and a fraction $\sigma' > 0$ of the workers with $\theta = \theta_L$ in the complex task and a fraction $(1-\sigma')$ of the workers with low signals in the simple task.

If the solution is of type 1 or 2 we can compute the factor inputs as

$$\begin{aligned} C(\sigma,\pi) &= (1-\sigma)\pi\alpha \\ S(\sigma,\pi) &= \sigma(\pi\alpha + (1-\pi)(1-\alpha)) + \pi(1-\alpha) + (1-\pi)\alpha, \end{aligned} \tag{3.19}$$

and if the unconstrained solution is implementable, σ, π must solve[10]

$$\max_{\sigma,\pi} ((1-\sigma)\pi\alpha)^{\beta} (\sigma(\pi\alpha + (1-\pi)(1-\alpha)) + \pi(1-\alpha) + (1-\pi)\alpha)^{1-\beta} - \frac{k\pi^2}{2} \tag{3.20}$$

By inspection of the problem we see that $\sigma = 1$ could never be a solution ($\sigma = 0$ is however a potential solution). The first order conditions are, after some rearranging,

$$\frac{\pi\alpha}{\pi\alpha + (1-\pi)(1-\alpha)} \frac{\beta}{1-\beta} \frac{S(\sigma,\pi)}{C(\sigma,\pi)} - 1 \geq 0, \tag{3.21}$$

(where the inequality must be an inequality if $\sigma > 0$) and

$$\beta\left(\frac{C(\sigma,\pi)}{S(\sigma,\pi)}\right)^{\beta-1} \alpha(1-\sigma) + (1-\beta)\left(\frac{C(\sigma,\pi)}{S(\sigma,\pi)}\right)^{\beta} (\sigma-1)(2\alpha-1) = k\pi, \tag{3.22}$$

If, on the other hand, the solution is of type 3 we have that

$$\begin{aligned} C(\sigma',\pi) &= \pi\alpha + \sigma'\pi(1-\alpha) \\ S(\sigma',\pi) &= (1-\sigma')(\pi(1-\alpha) + (1-\pi)\alpha) \end{aligned} \tag{3.23}$$

[10]The example is constructed so that the unconstrained solution is implementable, which will be demonstrated later.

and the (unconstrained) maximization problem has first order condition[11]

$$\frac{\pi(1-\alpha)}{\pi(1-\alpha)+(1-\pi)\alpha}\frac{\beta}{1-\beta}\frac{S(\sigma',\pi)}{C(\sigma',\pi)}-1=0, \tag{3.24}$$

for σ', while the condition for π still is (3.22). Since there is a discrete jump downwards in the marginal productivity in the complex task when the first low agent with low signal is put in the complex task, we suspect that there is a range of values for π such that the optimal solution is such that the planner puts a worker in the complex task if and only if the worker has a high signal. We now set $\beta = 1/2$ and *assume* that $\pi = 1/2$ in the optimal solution (we will set k later so that this is indeed a solution). After some simple substitutions we see that the first order condition (3.21) for a type 1 or 2 solution reduces to

$$\alpha\frac{\frac{1}{2}(1+\sigma)}{\frac{1}{2}(1-\sigma)\alpha}=\frac{(1+\sigma)}{(1-\sigma)}\geq 1 \;\; (=1 \text{ if } \sigma>0) \tag{3.25}$$

and the condition (3.24) for a type 3 solution reduces to

$$(1-\alpha)\frac{\frac{1}{2}(1-\sigma')}{\frac{1}{2}\alpha+\frac{1}{2}\sigma'(1-\alpha)}=1 \tag{3.26}$$

Inspecting these conditions we find that $\sigma = 0$ is the optimal solution to the full (unconstrained). The implied factor ratio is then simply α and after simplifications we see that in order for the first order condition (3.22) to hold for $\pi = 1/2$ it must be that

$$\left(\frac{1}{\alpha}\right)^{\frac{1}{2}}\alpha-\alpha^{\frac{1}{2}}(2\alpha-1)=2\alpha^{\frac{1}{2}}(1-\alpha)=k \tag{3.27}$$

The value of the objective function for this candidate solution is $\alpha^{\frac{1}{2}}(3+\alpha)/8$[12].

We now observe that given any σ (σ') there is at most one solution to (3.22), the first order condition for π. To realize this we first note that the factor ratio (complex/simple) is monotonically increasing in the fraction of investors. To see this formally in the case when some workers with high signals are in the simple task we consider the derivatives;

$$\begin{aligned}\frac{d}{d\pi}C(\sigma,\pi) &= \frac{d}{d\pi}(1-\sigma)\pi\alpha=(1-\sigma)\alpha>0 \\ \frac{d}{d\pi}S(\sigma,\pi) &= \frac{d}{d\pi}\sigma(\pi\alpha+(1-\pi)(1-\alpha))+\pi(1-\alpha)+(1-\pi)\alpha= \\ &= (1-\sigma)(1-2\alpha)<0,\end{aligned} \tag{3.28}$$

[11]Note that the corner is already taken care of in the other case.

[12]Rather than looking at second order conditions we will simply find *all solutions to the first order conditions* and then simply compare the value of the objective functions for all these candidate solutions.

since $\alpha > 1/2$. It is easy to check that the same holds true in the case when some workers with low signals are in the complex task as well. Hence, $C(\sigma, \pi) / S(\sigma, \pi)$ is strictly increasing in π for any σ, which means that the left hand side of (3.22) is strictly decreasing in π, while $k\pi$ is strictly increasing in π. It follows that there can be at most one solution for each σ. Hence, the candidate solution above is the unique candidate of the first type.

We now specialize the example further and suppose that $\alpha = 2/3, \beta = 1/2$ and $k = (2/3)^{\frac{3}{2}}$ (which makes $\pi = 1/2, \sigma = 0$ a solution candidate). We will now try to check if there are any other solutions to the first order conditions.

Type 2 Solutions: We now check for candidate solutions where some workers with the high signal are assigned to the simple task. In this case (3.21) must hold with equality and substituting in the parameter values above the condition simplifies to,

$$\frac{2\pi}{(1+\pi)} \frac{\sigma(1+\pi)+2-\pi}{(1-\sigma)2\pi} = 1. \tag{3.29}$$

Solving gives $\sigma^*(\pi) = \frac{2\pi-1}{2(1+\pi)}$ (which of course means that $\pi > \frac{1}{2}$). We can then compute the effective input of respective factor as $S(\sigma^*(\pi), \pi) = \frac{1}{2}$ and $C(\sigma^*(\pi), \pi) = \frac{\pi}{1+\pi}$. The implied factor ratio is consequently $\frac{2\pi}{(1+\pi)}$. Plugging this into the optimality condition for π and solving we get that the unique $\pi \in [0, 1]$ that satisfying (3.22) is

$$\pi^* = -\frac{1}{2} + \sqrt{\frac{1}{4} + \left(\frac{3\sqrt{3}}{4}\right)^{\frac{2}{3}}} \approx 0.70023 \tag{3.30}$$

Thus, there is a unique type 1 solution to the first order conditions in the example of consideration and we can compute the value of the objective as

$$\left(\frac{\pi^*}{1+\pi^*}\right)^{\frac{1}{2}} \left(\frac{1}{2}\right)^{\frac{1}{2}} - \frac{(2/3)^{\frac{3}{2}}(\pi^*)^2}{2} \approx 0.32, \tag{3.31}$$

to be compared with the value in the candidate solution with $\pi = 1/2$ and $\sigma = 0$

$$\left(\frac{1}{3}\right)^{\frac{1}{2}} \left(\frac{1}{2}\right)^{\frac{1}{2}} - \frac{(2/3)^{\frac{3}{2}}(1/2)^2}{2} \approx 0.34 \tag{3.32}$$

Type 3 Solutions: We will finally check for potential solutions where all workers with high signals and some workers with low signals are in the complex task. In this case the factor inputs are (with parameters set as above);

$$\begin{aligned} C(\pi, \sigma') &= \frac{2\pi}{3} + \frac{\sigma'(1-\pi)}{3} \\ S(\pi, \sigma') &= (1-\sigma')\left(\frac{2-\pi}{3}\right) \end{aligned} \tag{3.33}$$

and (3.24), the first order condition for σ' simplifies to

$$\pi(1-\sigma') = 2\pi + \sigma'(1-\pi), \tag{3.34}$$

which since $\pi > 0$ could never be satisfied for any $\sigma' \in [0,1]$. Hence, in this parametric example, the unique solution to the problem of maximizing total surplus (treating groups identically) is to set $\pi = \frac{1}{2}$ and $\sigma = \sigma' = 0$.

Next, we will check if there are wage schemes implementing this solution, i.e. we need to check that there exists a wage scheme w such that

$$G\left(\sum_{\theta\in\{\theta_L,\theta_H\}} \Pr(\theta|\text{inv})\, w(\theta) - \sum_{\theta\in\{\theta_L,\theta_H\}} \Pr(\theta|\text{not inv})\, w(\theta)\right) = \frac{1}{2} \tag{3.35}$$

Consider the wage scheme

$$w(\theta) = \begin{cases} 0 & \text{for } \theta = \theta_L \\ \frac{\sqrt{6}}{3} & \text{for } \theta = \theta_H \end{cases} \tag{3.36}$$

We find that the benefits to invest under this wage scheme is $\frac{1}{3}\frac{\sqrt{6}}{3} = \frac{\sqrt{6}}{9}$ and the fraction of investors consistent with this wage scheme is consequently given by

$$\int_0^{\frac{\sqrt{6}}{9}} \frac{1}{k} dc = \frac{\sqrt{6}}{9k} = \frac{\sqrt{6}}{9}\left(\frac{3}{2}\right)^{\frac{3}{2}} = \frac{1}{2}, \tag{3.37}$$

which means that the incentive compatibility is satisfied. Furthermore we need to make sure that the wage scheme satisfies the budget balance constraint. This constraint becomes

$$\underbrace{\frac{1}{\sqrt{3}}\frac{1}{\sqrt{2}}}_{y} - \underbrace{\frac{\sqrt{6}}{3}}_{w(\theta_H)}\left(\underbrace{\frac{2}{3}}_{\alpha}\underbrace{\frac{1}{2}}_{\pi} + \underbrace{\frac{1}{3}}_{1-\alpha}\underbrace{\frac{1}{2}}_{1-\pi}\right) \geq 0. \tag{3.38}$$

It is straightforward to verify that (3.38) holds with equality, which means that (3.36) satisfies [IR], [IC] and [BB].

Now let λ $(1-\lambda)$ be the fraction of workers from group a (b) and suppose that $\pi^a = 1$ and $\pi^b = 0$ and that (irrespective of the signal) a worker is assigned to the simple task if and only if he belongs to group a. The value of the planners' objective is then

$$\lambda^{\frac{1}{2}}(1-\lambda)^{\frac{1}{2}} - \frac{\lambda}{2}\left(\frac{2}{3}\right)^{\frac{3}{2}}. \tag{3.39}$$

In order to maximize the incentives for group a we let

$$w^a(\theta_H) = \frac{3}{2}\frac{y}{\lambda} = \frac{3\lambda^{\frac{1}{2}}(1-\lambda)^{\frac{1}{2}}}{2\lambda} \tag{3.40}$$

and let $w^a(\theta_L) = w^b(\theta_H) = w^b(\theta_L) = 0$. One verifies that budget balance is satisfied with equality with these wages and that the benefits of investment for agents in group a are

$$w^a(\theta_H)\frac{1}{3} = \frac{\lambda^{\frac{1}{2}}(1-\lambda)^{\frac{1}{2}}}{2\lambda}. \tag{3.41}$$

In order for $\pi^a = 1$ to be incentive compatible the benefits of investment must be at least $k = (2/3)^{\frac{3}{2}}$. Setting the benefits of investment equal to $(2/3)^{\frac{3}{2}}$ we find that the unique solution to this equation is $\lambda = 27/59 \approx 0.46$ (this value is the highest value of λ for which $\pi^a = 1$ can be implemented) and one verifies that the value of the objective is

$$\lambda^{\frac{1}{2}}(1-\lambda)^{\frac{1}{2}} - \lambda(2/3)^{\frac{3}{2}}/2 = \frac{3}{4}\lambda^{\frac{1}{2}}(1-\lambda)^{\frac{1}{2}} \approx 0.37 > 0.34 \tag{3.42}$$

Thus, there is a feasible solution involving discrimination that gives higher social surplus than the best non-discriminatory plan.

3.4. The Decentralized Model

We now consider how the "competitive model" compares with the allocation that solves the planning problem. The main point is to show that there is underinvestment in equilibrium.

3.4.1. The Model

To facilitate reading we will give a brief description of the game. The timing of events is as follows. First individual workers decide whether to invest or not. Each worker is then assigned a signal by nature. Firms then simultaneously announce wage schedules that are functions on all observable variables. A pure action for firm i in this stage can thus be described as a pair of wage schedules $\langle w_i^a, w_i^b \rangle$, where $w_i^j : [0,1] \to R_+$ for $j = a, b$. Workers observe the wage schedules announced by the firms and then decide what firm to work for. Then, in the final stage each firm decides how to allocate its available workers between the two tasks. Exactly as in the planner's problem the monotone likelihood ratio property implies that task-assignments will be done according to a cut-off rule for each group.

The reader may consult Moro and Norman [23] for a formal description of the strategy sets and payoffs as function of the actions.

3.4.2. Equilibrium

With firms competing in a Bertrand fashion for workers one shows that the essentially unique wage schemes consistent with equilibrium when a fraction π of the workers invest is given by

$$w^j(\theta) = \begin{cases} y_2(r(\pi), 1) & \text{for } \theta < \widetilde{\theta}(\pi) \\ p(\theta, \pi^j)\, y_1(r(\pi), 1) & \text{for } \theta \geq \widetilde{\theta}(\pi) \end{cases} \tag{3.43}$$

where

$$\widetilde{\theta}(\pi) = \left(\widetilde{\theta^a}(\pi), \widetilde{\theta^b}(\pi)\right) = \underset{(\theta^a,\theta^b)\in[0,1]^2}{\arg\max}\; y\left(\sum_j \lambda^j \pi^j \left(1 - F_q(\theta^j)\right), \sum_j \lambda^j F_{\pi^j}(\theta^j)\right) \tag{3.44}$$

and

$$r(\pi) = \frac{\sum_j \lambda^j \pi^j \left(1 - F_q\left(\widetilde{\theta^j}(\pi)\right)\right)}{\sum_j \lambda^j F_{\pi^j}\left(\widetilde{\theta^j}(\pi)\right)} \tag{3.45}$$

For formal arguments see Moro and Norman [23]. The intuitive reasoning is as follows: Bertrand competition between firms implies that both firms must offer the same wage schemes. This means that both firms are facing the same distribution of signals and must choose the (unique) cutoff-rule with cutoff given by (3.44), which, indeed, is the same as the planner would do if he had to take π as given. Next one shows that wages must be given by expected marginal productivities (in the task a worker will actually be assigned to), which is what (3.43) says.

In equilibrium only those workers that have incentives to invest will do so. The gross benefits to invest for a worker is simply the difference between the expected wage for a worker who invests and the expected wage for a worker who does not invest and this difference can be written as

$$\begin{aligned} B^j(\pi) &= y_2(r(\pi), 1)\left(F_q\left(\widetilde{\theta^j}(\pi)\right) - F_u\left(\widetilde{\theta^j}(\pi)\right)\right) + \\ &\quad + y_1(r(\pi), 1) \int_{\widetilde{\theta^j}(\pi)}^{1} p(\theta, \pi^j)\left(f_q(\theta) - f_u(\theta)\right) d\theta \end{aligned} \tag{3.46}$$

The set of equilibria of the model are thus fully characterized by the solutions to the fixed point equations

$$\pi^j = G\left(B^j(\pi)\right) \text{ for } j = a, b \tag{3.47}$$

To focus on the extent of underinvestment in the model we will forget momentarily about the group characteristics. Thus, π will be a scalar in the following discussion and $\widetilde{\theta}(\pi)$ will now denote the (single) cutoff point when a fraction π of the workers invest. We will also use $B(\pi)$ as the gross benefits of investment in the single group case (defined in analogy with (3.46)).

3.4.3. Underinvestment in equilibrium

To fix ideas we initially consider how the set of *symmetric* equilibria (or equivalently the set of equilibria in the model where there is only a single group[13]) compares with the solutions to the planning problem considered in Section 3.3. We start with a local result.

Proposition 12. *If π^{eq} is an arbitrary equilibrium, then a sufficiently small increase in the fraction of investors increases the value of the planners' objective.*

Proof. We note that there is a neighborhood U around any equilibrium π^{eq} such that any $\pi \in U$ is implementable and that the objective function is continuously differentiable. We may therefore simply study the sign of the derivative of the planners' objective to see how surplus is affected by small changes in π. Taking the derivative we get that

$$\frac{d}{d\pi}\left[Y\left(\pi^{eq}\right)-\int_{\underline{c}}^{G^{-1}\left(\pi^{eq}\right)} cg\left(c\right)dc\right]=Y'\left(\pi^{eq}\right)-G^{-1}\left(\pi^{eq}\right)=Y'\left(\pi^{eq}\right)-B\left(\pi^{eq}\right), \qquad (3.48)$$

where the last equality follows from (3.47). But for any $\pi > 0$

$$\begin{aligned} Y'\left(\pi\right) &= y_1\left(r(\pi),1\right)\left(1-F_q\left(\widetilde{\theta}\left(\pi\right)\right)\right)+y_2\left(r\left(\pi\right),1\right)\left(F_q\left(\widetilde{\theta}\left(\pi\right)\right)-F_q\left(\widetilde{\theta}\left(\pi\right)\right)\right) > \qquad (3.49)\\ &> y_1\left(r\left(\pi\right),1\right)\left(\int_{\widetilde{\theta}(\pi)}^{1} p\left(\theta,\pi\right)\left(f_u\left(\theta\right)\right)d\theta\right)+y_2\left(r\left(\pi\right),1\right)\left(F_q\left(\widetilde{\theta}\left(\pi\right)\right)-F_q\left(\widetilde{\theta}\left(\pi\right)\right)\right)= \\ &= B\left(\pi\right) \end{aligned}$$

It follows that $Y'\left(\pi^{eq}\right)-B\left(\pi^{eq}\right)>0$. ■

Given any equilibrium where $\pi^j > G(0)$ in the model with multiple groups *exactly the same argument goes through.* The (local) underinvestment property thus carries over even to interior discriminatory equilibria (for both groups) and, in the case where one group is in the lower corner, there is always underinvestment in the dominant group, unless everybody invests.

Local results are always of limited interest, but in some nice cases we can actually get a global result as a consequence of the next proposition.

Proposition 13. *Let $\underline{\pi}^{eq}$ be the smallest (non-discriminatory) equilibrium and let (π^*, θ^*) be any solution to the planning problem (3.10). Then $\underline{\pi}^{eq} < \pi^*$.*

[13]See Moro and Norman [23] for a more detailed discussion of the model with just a single group of workers.

Proof. Let (π^*, θ^*) solve the planning problem (3.10). From Proposition 11 we know that there are two possibilities. The first possibility is that (π^*, θ^*) is indeed a maximum of (3.10), in which π^* must satisfy

$$Y'(\pi^*) = G^{-1}(\pi^*) > B(\pi^*), \tag{3.50}$$

where the inequality follows from (3.49). Thus $B(\pi^*) < Y'(\pi^*)$, which implies that $\pi^* < G(B(\pi^*))$. Given the assumption that $G(0) > 0$ it follows by the intermediate value theorem that there must exist some $\underline{\pi}^{eq} \in (0, \pi^*)$ such that $\underline{\pi}^{eq} = G(B(\underline{\pi}^{eq}))$, which proves the result in the case when (π^*, θ^*) is a maximum. Suppose next that the supremum is not attainable. This means that there exists no feasible wage scheme such that $\pi^* = G\left(\int w(\theta)(f_q(\theta) - f_u(\theta))\, d\theta\right)$. so clearly $\pi^* < G(B(\pi^*))$. The same logic applies. ■

It follows directly that;

Corollary *If there is a unique (non-discriminatory) equilibrium, then there is underinvestment in this equilibrium.*

When there are multiple equilibria one may still conjecture that there is underinvestment in *any* equilibrium even in a global sense. But here the non-convexities and the potential lack of concavity of the objective create problems and although I have not been able to set up an example, it seems that it is quite possible that there are equilibria with overinvestment, in the sense that the fraction of investors is higher than the global maximum (or sup) of the planing problem.

However, the case when Y is concave is easily handled.

Proposition 14. *Suppose Y is concave. Then there is underinvestment in any equilibrium.*

Proof. If Y is concave, the planners objective function is concave, which means that there is at most one π^* satisfying $Y'(\pi^*) = G^{-1}(\pi^*)$ and that the social surplus is strictly increasing on $[0, \pi^*]$. Now if there exists a feasible wage scheme implementing π^* we are done since concavity of the objective implies that $Y'(\pi) < G^{-1}(\pi)$ for all $\pi > \pi^*$, so $B(\pi) < G^{-1}(\pi)$ for all $\pi \geq \pi^*$, i.e. there could not be any equilibrium on $[\pi^*, 1]$. If π^* is not implementable then concavity implies that if π' is a solution to (3.10), then there exists no π in between π' and π^* that are implementable (i.e., if $\pi' < \pi^*$[14], then there exists no $\pi \in [\pi', \pi^*]$ such that π can be implemented). If there is no

[14]Since the set of implementable values of π may be non-convex we can't rule out the possibility that the actual solution to (3.10) is larger than π^* . However, all equilibria must be smaller than π^*, so this doesn't affect the argument.

implementable wage scheme there is no equilibrium wage scheme either, so any solution to (3.10) must always be larger than all equilibria of the laissez faire model. ■

While suggestive this result has the obvious drawback that Y is a derived function and not a primitive of the model. The shape of Y will depend in a rather complicated way on both the production function y and the density functions f_q and f_u.

3.5. Discussion

In this paper we have examined the efficiency properties of statistical discrimination. Unlike the earlier literature on the subject, we have done this by contrasting a competitive model where statistical discrimination may in equilibrium occur with the solutions to a planning problem.

All equilibria of the competitive model are inefficient. Hence, policy intervention can always be justified without resorting to equity arguments. Still, there may very well be "too little discrimination in equilibrium", meaning that a social planner would prefer to increase the differences between groups[15].

When Schwab [27] and Lundberg and Startz [19] investigates the efficiency properties of statistical discrimination they compare a competitive model where the firms are allowed to use all available information, including group identity, with a situation where firms are no longer allowed to use group identity when determining wages. While this approach leads to several interesting insights, it is always vulnerable to the criticism that there may be alternative policies that are as simple to implement as the "equal opportunities law" they consider[16]. In fact, in both papers cited above there are policies that implement the "first best" and since both Schwab [27] and Lundberg and Startz [19] assume that there are some intrinsic differences between groups, first best involves differential treatment of groups.

One major advantage with the approach in this paper is that it is not subject to this type of criticism. Rather than comparing the *laissez faire* equilibria with equilibria under a particular policy, the set of feasible policies is determined from more primitive assumptions on the informational technology.

[15]In the formal model of the paper we have assumed that the planer is utilitarian. If the planner also has distributional concerns this conclusion may still be valid. The reason is that the planner then may counter the undesirable distributional consequences by transfers between groups.

[16]In fact, as noted by Lundberg [18], the equal oppurtunities policy may in fact be rather difficult to implement. In the model in this paper, implementation of an equal opportunities policy requires that the policymaker observes the signal for each worker, which is exactly the informational assumption I make in the planning problem.

To keep the problem tractable many simplifying assumptions were made. However, the trade-off between the gains from discrimination in terms of reduced "mismatch" and the losses in terms of inefficient investment decisions seems robust. The key assumptions for these welfare benefits of discrimination/specialization to occur are: 1) human capital investments are imperfectly observable, 2) human capital more important in certain (qualified) jobs than other and 3) labor input in qualified jobs and unqualified jobs are complementary in the neoclassical sense. The first of these assumptions is absolutely necessary to explain discrimination without resorting to prejudice or asymmetries between groups and the other two seems rather plausible. Although I think it is important to realize that models of statistical discrimination with endogenous human capital have these potential welfare benefits of discrimination/specialization, one should clearly be careful interpreting this result. There may, for example, exist more subtle way than discrimination of ethnic groups to generate more specialized human capital investments.

The other main point of the paper, that it is difficult to separate policies aimed at discrimination and policies aimed at underinvestment in human capital, should be uncontroversial. Models of human capital formation usually have too few workers investing in equilibrium. Hence, it seems that this should be relevant for *any* model of statistical discrimination with endogenous human capital.

3.6. Appendix

3.6.1. Proof of Lemma 22

Proof. First we show that there exists a sequence of feasible wage schemes that such that the benefits of investment can be made arbitrarily close to b^{j*} . Without loss we consider only $j = a$ and consider wage schemes of the form

$$w_\epsilon^a(\theta) = \begin{cases} 0 & \text{for } \theta < 1-\epsilon \\ \frac{\overline{y}-\lambda^b W^b}{\lambda^a(1-F_{\pi^a}(1-\epsilon))} & \text{for } \theta \geq 1-\epsilon \end{cases} \tag{3.51}$$

Obviously [IR] is satisfied and one verifies that

$$\lambda^a \int w_\epsilon^a(\theta) f_{\pi^a}(\theta) d\theta = \overline{y} - \lambda^b W^b, \tag{3.52}$$

which means that [BB] is satisfied with equality for an arbitrary $\epsilon \in (0,1)$. The benefits to invest given w_ϵ^a is

$$h^a(\epsilon;\pi,\overline{y}) = \frac{\overline{y}-\lambda^b W^b}{\lambda^a(1-F_{\pi^a}(1-\epsilon))}\left(F_u(1-\epsilon) - F_q(1-\epsilon)\right), \tag{3.53}$$

where $F_{\pi}(\theta) = \pi F_q(\theta) + (1-\pi) F_u(\theta)$. Note that

$$sign\left(\frac{d}{d\epsilon}\left(\frac{F_u(1-\epsilon) - F_q(1-\epsilon)}{(1 - F_{\pi^a}(1-\epsilon))}\right)\right) = -sign\left((f_u - f_q)(1 - F_{\pi^a}) + (F_u - F_q) f_{\pi^a}\right), \quad (3.54)$$

where f_{π} is the density associated with F_{π}, i.e. $f_{\pi}(\theta) = \pi f_q(\theta) + (1-\pi) f_u(\theta)$. But

$$\begin{aligned} A(\epsilon) &= (f_u - f_q)(1 - F_{\pi^a}) + (F_u - F_q) f_{\pi^a} = \\ &= f_u\left(1 - F_{\pi^a} + (1-\pi^a) F_u - (1-\pi^a) F_q\right) - f_q\left(1 - F_{\pi^a} - \pi^a F_u + \pi^a F_q\right) = \\ &= f_u(1 - F_q) - f_q(1 - F_u) \end{aligned} \quad (3.55)$$

We now claim that $A(\epsilon) > 0$ for all $\epsilon > 0$ as a consequence of the monotone likelihood ratio property. To see this assume $A(\epsilon) \leq 0$ and let $l(\theta) = f_q(\theta)/f_u(\theta)$ denote the likelihood ratio. Then,

$$l(1-\epsilon) = \frac{f_q(1-\epsilon)}{f_u(1-\epsilon)} \geq \frac{1 - F_q(1-\epsilon)}{1 - F_u(1-\epsilon)} = \frac{\int_{1-\epsilon}^{1} f_q(\theta)\, d\theta}{\int_{1-\epsilon}^{1} f_u(\theta)\, d\theta} = \frac{\int_{1-\epsilon}^{1} l(\theta) f_u(\theta)\, d\theta}{\int_{1-\epsilon}^{1} f_u(\theta)\, d\theta} \quad (3.56)$$

But, by the monotone likelihood ratio property, $l(\theta) > l(1-\epsilon)$ for all $\theta > 1-\epsilon$. Thus

$$\int_{1-\epsilon}^{1} l(\theta) f_u(\theta)\, d\theta > l(1-\epsilon) \int_{1-\epsilon}^{1} f_u(\theta)\, d\theta, \quad (3.57)$$

which gives $l(1-\epsilon) > l(1-\epsilon)$, a contradiction. Hence, the benefits of investment are strictly decreasing in ϵ

By applying l'Hopitals rule we now find that by choosing ϵ small enough we can make the benefits of investment arbitrarily close to $(\overline{y} - \lambda^b W^b)(f_q(1) - f_u(1))/\lambda^a f_{\pi^a}(1) \equiv b^*$. Hence the supremum is *at least* b^{a*}.

Next, we show that there is no feasible wage scheme that gives (weakly) higher benefits to invest. To do this consider an arbitrary wage scheme w^a. The benefits of investment is then simply

$$\int w^a(\theta)(f_q(\theta) - f_u(\theta))\, d\theta = \int w^a(\theta) f_{\pi^a}(\theta) \frac{(f_q(\theta) - f_u(\theta))}{f_{\pi^a}(\theta)} d\theta \quad (3.58)$$

$(f_q(\theta) - f_u(\theta))/f_{\pi^a}(\theta) < (f_q(1) - f_u(1))/f_{\pi^a}(1)$ for all $\theta < 1$ and since $w^a : [0,1] \to R_+$ there must be some set Θ of strictly positive measure such that $w^a(\theta) > 0$ for all $\theta \in \Theta$. It follows that

$$\begin{aligned} \int w^a(\theta) f_{\pi^a}(\theta) \frac{(f_q(\theta) - f_u(\theta))}{f_{\pi^a}(\theta)} d\theta &< \int w^a(\theta) f_{\pi^a}(\theta) \frac{(f_q(1) - f_u(1))}{f_{\pi^a}(1)} d\theta = \\ &\leq \frac{y - \lambda^b W^b}{\lambda^a} \frac{(f_q(1) - f_u(1))}{f_{\pi^a}(1)} = b^{a*}, \end{aligned} \quad (3.59)$$

where the second inequality follows from [BB]. ■

3.6.2. Proof of Lemma 23

Proof. We leave to the reader to verify that there is a unique solution π^{a*} to (3.9) for any fixed $y - \lambda^b W^b$. Combining Lemma 23 with continuity of G we see that it is possible to implement any fraction of investors $\pi^a < \pi^{a*}$ satisfying [IC],[IR] and [BB]. Now suppose we can implement some $\pi^a \geq \pi^{a*}$. To satisfy [IR] and [BB] it must then be that (Lemma 22)

$$\int w^a(\theta)(f_q(\theta) - f_u(\theta))\,d\theta < \frac{(y - \lambda^b W^b)(f_q(1) - f_u(1))}{\lambda^a f_{\pi^a}(1)} \tag{3.60}$$

Imposing the [IC]-constraint we find that

$$\pi^a < G\left(\int w^a(\theta)(f_q(\theta) - f_u(\theta))\,d\theta\right) < G\left(\frac{(y - \lambda^b W^b)(f_q(1) - f_u(1))}{\lambda^a f_{\pi^a}(1)}\right) = \pi^{a*}, \tag{3.61}$$

a contradiction. ■

3.6.3. Proof of Proposition 10

Proof. Using Lemma 23 we show that if there exists α such that both $G\left(\frac{\alpha y^*(f_q(1) - f_u(1))}{\lambda^a f_{\pi^{a*}}(1)}\right) > \pi^{a*}$ and $G\left(\frac{(1-\alpha)y^*(f_q(1) - f_u(1))}{\lambda^a f_{\pi^{b*}}(1)}\right) > \pi^{b*}$ holds, where $\pi^{j*} \geq G(0)$ [17], then we can find wages satisfying all constraints implementing π^* . For the other direction, suppose that the solution involving π^* can be implemented by some wage schemes. Let W^j be the total wage costs for each group and note that in order for [IC],[IR] and [BB] to hold we have from Lemma 23 that

$$G\left(\frac{(y^* - \lambda^k W^k)(f_q(1) - f_u(1))}{\lambda^j f_{\pi^{a*}}(1)}\right) > \pi^{j*},$$

Let $\alpha = (y^* - \lambda^b W^b)/y^* \in [0,1]$. Then

$$(1-\alpha)y^* = 1 - (y^* - \lambda^b W^b) = \lambda^b W^b \geq y^* - \lambda^a W^a,$$

where the last inequality comes from [BB]. The result follows. ■

[17] In order for π^{j*} to be part of an optimal solution to (3.3) it must be that $\pi^{j*} \geq G(0)$. The intuitive reason is that if $\pi^j < G(0)$, then the planner can achieve (at least) the same output by switching to $\pi'^j = G(0)$, without changing anything else in the other group.

BIBLIOGRAPHY

[1] Akerlof, George, "The Market for Lemons: Quality Uncertainty and the market mechanism." *Quarterly Journal of Economics,* August 1970, **84**, 488-500.

[2] Akerlof, George, "The Economics of Caste and the Rat Race and Other Woeful Tales." *Quarterly Journal of Economics,* November 1976, **90**, 599-617.

[3] Aigner, Dennis J. and Glen G. Cain, "Statistical Theories of Discrimination in Labor Markets", *Industrial and Labor Relations Review.* **30,** January 1977, 175-87.

[4] Arrow, Kenneth J., "The Theory of Discrimination," in O. Ashenfelter and A. Rees, ed , *Discrimination in Labor Markets*, Princeton, NJ: Princeton University Press, 1973, 3-33.

[5] D. P. Baron: "A Theory of Collective Choice for Government Programs", *Mimeo*, Stanford University (1993).

[6] D. P. Baron and J. Ferejohn: "Bargaining in Legislatures", *American Political Science Review* **83** (1989), 1181-1206.

[7] D. P. Baron and J. Ferejohn: "The Power to Propose", in *Models of Strategic Choice in Politics*, (1989).

[8] D. P. Baron and E. Kalai, "The Simplest Equilibrium of a Majority-Rule Division Game", *Journal of Economic Theory* **61** (1992), 290-301.

[9] Becker, Gary S., *The Economics of Discrimination*, Chicago: University of Chicago Press , 1957.

[10] J. P. Benoit and V. Krishna, "Finitely repeated games", *Econometrica* **53** (1985), 890-904.

[11] V. V. Chari, L. E. Jones and R. Marimon, "The Economics of Split Voting in Representative Democracies", *Mimeo*, Northwestern University (1994)

[12] Coate, Stephen and Glenn C. Loury, "Will Affirmative Action Policies Eliminate Negative Stereotypes?," *American Economic Review* **83**, December 1993, 1220-40.

[13] Cornell, Bradford and Ivo Welch, "Culture, Information, and Screening Discrimination", *Journal of Political Economy,* **104,** (1996) 542-571.

[14] Feldman, M. and C. Gilles, "An Expository Note on Individual Risk without Aggregate Uncertainty", .*Journal of Economic Theory*, **35**, 1985, 26-32.

[15] C. Harris, "Existence and characterization of perfect equilibrium in games of perfect information", *Econometrica* **53** (1985), 613-627.

[16] Judd, Kenneth L., "The Law of Large Numbers with a Continuum of I.I.D. Random Variables", *Journal of Economic Theory* **35** (1985) 19-25.

[17] Kahn, Lawrence M., "Customer Discrimination and Affirmative Action." *Economic Inquiry*, July 1991, **26** 555-71.

[18] Lundberg, Shelly J. "The Enforcement of Equal Opportunity Laws Under Imperfect Information: Affirmative Action and Alternatives", *Quarterly Journal of Economics* CVI, February 1991, 309-26.

[19] Lundberg, Shelly J. and Richard Startz, "Private Discrimination and Social Intervention in Competitive Labor Markets", *American Economic Review* **73** (1983), 340-47.

[20] R. D. McKelvey and R. Riezman, "Seniority in Legislatures", *American Political Science Review* **86** (1992), 952-965 .

[21] Milgrom, Paul and Sharon Oster, "Job Discrimination, Market Forces and the Invisibility Hypothesis," *Quarterly Journal of Economics*, August 1987, **CII**, 453-76.

[22] Milgrom, Paul and R. Weber, "Distributional Strategies for games with incomplete information", *Mathematics of Operations Research* **50**, 619-631.

[23] Moro, Andrea and Peter Norman, "Affirmative Action in a Competitive Economy", *CARESS Working paper* **#96-08,** October 1996.

[24] Phelps, Edmund S., "The Statistical theory of Racism and Sexism," *American Economic Review* **62** (September 1972), 659-61.

[25] W. H. Riker, *The Theory of Political Coalitions*, Yale University Press, New Haven, 1962.

[26] A. Rubinstein, "Perfect Equilibrium in a Bargaining Model", *Econometrica* **50** (1982), 97-110.

[27] Schwab, Stewart, "Is Statistical Discrimination Efficient?" *American Economic Review* **76** (March 1986), 229-34.

[28] Sowell, Thomas *Preferential Policies. An International Perspective*, New York, William Morrow and Co., 1990.

[29] Spence, Michael A., *Market Signaling: Information Transfer in Hiring and Related Screening Processes*, Cambridge, MA, Harvard University Press, 1974.

[30] Welch, Finis, "Employment Quotas for Minorities," *Journal of Political Economy*, August 1976, 84, S105-39.