The Role of Referrals in Inequality, Immobility, and Inefficiency in Labor Markets

Lukas Bolte, Nicole Immorlica and Matthew O. Jackson*

January 2020

Most Recent Version: `https://papers.ssrn.com/abstract=3512293`

Abstract

Labor markets rely on a combination of referrals and open applications in hiring. Referrals help screen candidates and so lead to better matches and increased productivity; but also give referred workers a relative advantage, and hence can lead to inequality. We examine how these different sources of hiring interact, and how that interaction determines inequality within a generation, immobility across generations, as well as the overall productivity of a society. We show that open applicants suffer from a "lemons" effect: some are applying after being rejected for jobs via their referrals. This lemons effect lowers the value of hiring through open applications and makes referred candidates look even relatively more attractive, further disadvantaging applicants without referrals. As a result, inequality in referrals, due to historic employment rates or imbalances in referral rates across groups, ties fates together across a group's generations resulting in immobility as well as inequality within a generation. Concentrating referrals among a subpopulation can also lower productivity. We show how these all interact, and identify the *different* conditions under which concentrating referrals among some group raises inequality, immobility, and/or lowers productivity. We show how characterizing these effects depends on understanding the lemons effect, and also examine extensions of the model that include the possibility of firing workers, as well education choices by workers.

JEL Classification Codes: D85, D13, L14, O12, Z13

Keywords: Inequality, Immobility, Job Contacts, Job Referrals, Social Networks, Networks, Productivity

---

*Lukas Bolte; Department of Economics, Stanford University, Stanford, CA; lbolte@stanford.edu. Nicole Immorlica; Microsoft Research, One Memorial Drive, Cambridge, MA; nicimm@microsoft. Matthew O. Jackson; Department of Economics, Stanford University, Stanford, CA; jacksonm@stanford.edu. Jackson is also an external faculty member at the Santa Fe Institute, and a fellow of CIFAR. We gratefully acknowledge support under NSF grant SES-1629446.

1 Introduction

The past few decades have seen inequality in countries like the US and UK rise to levels not seen since the Great Depression, and seen countries like China hit levels of inequality that are unprecedented. Inequality also comes together with immobility (a correlation between parent and child income), as made clear by Alan Krueger's 2012 famous demonstration of the "Great Gatsby Curve," which showed that countries with higher inequality also tend to have higher immobility.[1] There are many reasons for inequality as well as immobility, but understanding their deep roots is essential to designing policies that deal not only with their symptoms but also their causes.[2]

A key feature of labor markets is that a majority of employees, throughout the spectrum of skill levels, are hired via referrals.[3] It has been shown that this fact can result in inequality in employment and wages across groups: a higher current employment among some ethnic, gender, age or other-based group gives them an advantage in getting future jobs.[4] Note that referrals naturally tie inequality together with immobility: if one is born into a group that has poor employment, then it becomes hard to get referrals to attractive jobs. Thus, when there is high inequality across groups, referrals can make it naturally persistent.

Although referrals can lead to inequality and immobility, they also serve a productive purpose: referrals provide information about the productivity of the referred worker. Referrals thus enable firms to vet workers, and hence those hired via referrals are more productive, on average, than those hired through a less informative open application process.[5] This suggests that comparing an economy in which referrals are spread widely to one in which they are concentrated among a subset of a society should not only lead to increased inequality, but can also lower a society's productivity. Thus, one might expect inequality in employment and wages not only to be related to immobility, but also to lower productivity.

In this paper we develop a model in which firms hire both via referrals and open applications, which allows us to provide a detailed analysis of the role of referrals in driving inequality, the persistence of that inequality (immobility), and productivity. To understand the role of referrals, the key comparative static that we analyze is how things change as the

[1] See Corak (2016) for more details.

[2] See Chapter 6 in Jackson (2019) for background figures, references, and discussion concerning the inequality and its relationship to immobility mentioned in this paragraph.

[3] For background, see Granovetter (1974, 1995); Montgomery (1991); Ioannides and Datcher-Loury (2004); Topa (2011).

[4] For some of the extensive evidence on this, see Munshi (2003); Calvo-Armengol and Jackson (2004); Arrow and Borzekowski (2004); Calvo-Armengol and Jackson (2007); Laschever (2011); Beaman (2012); Patacchini and Zenou (2012); Beaman, Keleher and Magruder (2016); Lalanne and Seabright (2016); Jackson (2019).

[5] For some evidence of this difference see Fernandez, Castilla and Moore (2000); Brown, Setren and Topa (2012); Pallais and Sands (2016), and for more discussion of the many roles of referrals in different settings see Heath (forthcoming); Jackson (2019). We can also think of referrals as lowering search frictions, and we know from the search literature (e.g., see Rogerson, Shimer and Wright (2005) for a review), that decreasing frictions can raise overall productivity.

distribution of referrals across the population is varied. In particular, the natural comparative statics that we consider are the impacts of concentrating referrals among a subset of the population. Comparing an economy in which connections to currently employed workers, and hence referrals, are spread evenly across the population, to one in which those connections and referrals are concentrated among a subgroup, is a general way of seeing how gaps in past employment across groups lead to future inequality, immobility, and also affect productivity.

As we show, the conditions under which concentrating referrals among some subpopulation raises inequality or lowers productivity are different, and each requires specific features in terms of how the concentration works and how it relates to the distribution of productivities in the society. To understand this, it is important to first understand a key feature that makes our analysis subtle: a lemons effect. This refers to the phenomenon that some people who apply for jobs via open applications had referrals but were not hired via those referrals - so they were already rejected by at least one firm. The fact that the open applications include previously-rejected job seekers lowers the expected productivity from hiring via open applications. A main complication in our analysis is that the lemons effect decreases as one concentrates referrals, since there are then fewer workers who are screened by hiring firms and rejected. This effect attenuates the relationship between productivity and the concentration of referrals. It also complicates the relationship between inequality and the concentration of referrals, and causes comparative statics on inequality to differ fundamentally from the comparative statics on productivity.

It turns out that there are two different ways in which referrals can be concentrated, and each has different effects. To make this crystal clear, let us keep the overall number of referrals constant: so, for instance, each currently employed worker gets to refer one person from the next generation to replace them. The key is that some current employees might refer the same people. For instance, if white males have a higher employment rate than black males, and tend to refer other white males, then as the current employment becomes more tilted towards whites, then more whites will get referrals and fewer blacks will – with more whites getting multiple referrals. This corresponds to one way in which referrals can be concentrated, which is to have a smaller number of the next generation receive referrals. This necessarily means, since we are holding the total number of referrals constant, that people who receive at least one referral must be getting a higher number of referrals on average. In our model a worker's productive value is the same across firms and so getting multiple referrals does not help in productively matching workers to firms. Thus, it is this sort of concentration – which we call "concentration at the low end" – that is the key to understanding the impact of changing the distribution of referrals on productivity.

To understand differences in wages, one needs to understand the effects of multiple referrals. Having more than one referral can give an employee multiple offers and hence more bargaining power. Thus, determining the fraction of the population that gets higher wages as referrals are concentrated depends on the faction of the population that gets multiple referrals. Concentrating referrals at the low end is neither necessary nor sufficient for having a larger fraction of the population get multiple referrals. For instance, suppose that the population of applicants consists of people who have zero, one, two, or three referrals. By redistributing the extra referrals from those who have three, to the rest of the population it is possible to both increase the number of people who have multiple referrals at the same time as the number of people who have any referrals. Thus, it is possible to have a larger of fraction of the population getting multiple referrals, while still having more of the population get referrals, and keeping the overall total constant. The key to understanding inequality comes from the number of people getting multiple referrals, and so we call that "concentrating referrals at the high end". Thus, despite the fact concentrating referrals low end and high end often happen together, neither is sufficient for the other. Since one governs productivity and other inequality, there is a close but imperfect relationship between how productivity and inequality are influenced by referrals.

The bulk of our analysis concerns the impact of these different forms of concentration on the lemons effect, productivity, and inequality. But we also extend the model in two directions. One is that we allow firms to fire workers. This makes hiring from open applications more attractive: if a firm ends up with low-productivity worker then it can fire that worker and rehire. Overall, this increases productivity, and also lowers the lemons effect, but also changes how picky firms are in hiring referrals. The other is that we examine the long run dynamics of the model, when generation after generation passes referrals onward. In the basic model, if one group, say blues, has a higher initial employment rate than another group, say greens, and there is homophily so that blues are more likely to refer blues and greens are more likely to refer greens, then the blues' advantage in employment and wages is passed along to the next generation. An interesting by-produce is that employed blues have higher productivity on average than hired greens (given their referral advantage so that more of them are vetted), while unhired greens are more productive than unhired blues. Over time, the advantage of blues shrinks - as some greens will be hired through open applications and able to offer referrals to the following generation. However, if people have to invest (e.g., in education) in order to become productive, then this advantage does not shrink. Greens have lower incentives to invest in education than Blues, which can perpetuate inequality. Therefore, to have persistent group-based immobility one has to augment asymmetries across groups in referrals with costs of investing in education.

As pointed out above, it has not escaped the literature that referrals, especially when combined with homophily (the tendency of people to be connected to others to whom they are similar), can lead to inequality. Also, the fact that search processes can lead to lemons effects has also been noted before (Gibbons and Katz (1991); Farber (1999); Krueger et al. (2014))

Our contribution is to develop a model that combines referrals and open applications, and to characterize their interplay. To our knowledge, there are no antecedents to our comparative statics, especially the distinction between the different ways in which referrals can be concentrated, as well as the results that we establish that establish the relationships between inequality, immobility, and productivity.

2 A Model and Preliminaries

We consider a labor market with a unit mass of risk-neutral firms, each having one current but retiring employee. Each firm wishes to hire at most one next-generation worker to replace its retiring employee. There is a mass $n \geq 1$ of risk-neutral agents seeking work at these firms; so there can be unemployment. We refer to these agents as "workers" regardless of their employment status and the currently employed agents as employees.

A worker i provides a productive value v_i to any firm that employs that worker. This value includes the worker's skill and talent, and whatever else makes the worker productive, and is the maximum amount that the firm would be willing to pay to hire the worker, if the firm had no possibilities of filling the vacancy. The distribution of v_i is denoted by F, and is independent across workers.[6] F has a finite mean and we consider the nondegenerate case in which F has weight on more than one value, and so v_i has nonzero variance. For clarity, we discuss the case in which this value is the same to any firm that employs the worker, and relegate discussion of an extension in which there can be idiosyncratic match values that differ across firms to an appendix.

Firms hire workers either through referrals or via a pool of open applications – via "referrals" or "the pool"). Referrals are generated through a network in which each firm's current (retiring) employee refers a next-generation worker. For simplicity, we assume each employee refers exactly one worker. Nonetheless, some next generation workers may receive multiple referrals as several current employees may each refer the same next generation worker to their respective firms. We denote the distribution of referrals per worker by P, with $P(k)$ being the probability that a generic worker gets exactly k referrals.[7] We allow for a variety of different referral processes, each captured via a different degree distribution. For example, if referrals are made uniformly at random across all workers then P is a Poisson distribution; i.e.,

$$P(k) = \frac{n^{-k}e^{-1/n}}{k!},$$

with $1/n$ being the average number of referrals per worker.

Workers have a minimum wage that they must be paid, w_{min} (which is presumed to be at least their outside option value). This minimum wage can either be the worker's outside option value, or else a minimum wage imposed by some government. We presume throughout that w_{min} lies below the max of the support of F, so that there is a positive mass of workers that firms find strictly worth hiring.

Each firm observes the value of its referred worker and then chooses whether to hire the worker. If a firm chooses not to hire its referred worker, it can go to an anonymous pool, consisting of all workers who either have no referrals or are not hired via any of their

[6] add References on issues of iid in continuum. Basically, these are well understood and we will not go into details here; but the appropriate approximation theorems apply.

[7] See Currarini et al. (2009) for references and a discussion on a foundation for such a matching in the continuum.

referrals, and may hire a worker picked at random from that pool.[8]

In summary, the timing of the game is as follows.

1. Each firm receives a referral to a worker.

2. Each firm chooses whether to make an offer to its referred worker and if so, at what wage.

3. Referred workers who receive at least one wage offer choose to accept one of them or to reject all of them.

4. Any accepted offer is consummated and the firm and worker are matched. The firm receives a payoff equal to the productivity of its worker minus the wage. The worker receives a payoff equal to its wage.

5. Remaining workers and firms get matched with one worker chosen uniformly at random from the pool (without replacement, so that no two firms are matched to the same worker from the pool).

6. Each firm chooses whether to make an offer to its worker from the pool and if so, at what wage.

7. Workers from the pool who receive a wage offer accept or reject it.

8. Any accepted offer is consummated and the firm and worker are matched. The firm receives a payoff equal to the productivity of its worker minus the wage. The worker receives a payoff equal to its wage.

9. Remaining firms and workers are unmatched.

A firm perfectly observes the value of a worker that is referred to it, but only knows the (equilibrium) expected value of the workers in the anonymous pool. As the mass of workers exceeds the mass of firms, some workers will be unemployed and the referral market helps in selecting more productive workers.

The fact that there is more information and advantages to hiring via referrals compared to open application process has been found for various reasons.[9] For simplicity, we look at an extreme version of this in which firms can fully assess the value of a referred worker and only the expected value of a non-referred worker, but results similar to those presented below obtain as long as there is superior information about referred workers.

[8]We assume for convenience that the firm does not search the pool and in fact only knows the expected value of workers in the pool. While dropping this assumption does not substantially change our results, we note it is justified when search in the anonymous pool is costly; e.g., because learning the value of a worker requires some trial employment period.

[9]For instance, see Rees (1966); Fernandez, Castilla and Moore (2000); Brown, Setren and Topa (2012); Pallais and Sands (2016).

2.1 Equilibrium Characterization

We examine (weak) perfect Bayesian equilibria of the game.

The basic equilibrium structure is easy to discern in this game, and can be seen from backward induction. Workers accept any wages at or above w_{min} but not below. The only possibility to have mixing at w_{min} is if that is the workers' outside option value *and* happens to be exactly the expected value of the worker to the firm.[10] In that (non-generic) case, the mixing is becomes irrelevant since there is 0 net expected value in the relationship to either side, and so it does not effect any of the results that follow. Thus, equilibrium behavior in the pool stage is such that firms hire workers from the pool at a wage of w_{min} if the expected value of workers in the pool exceeds w_{min}, do not hire from the pool if the expected value is below w_{min}, and both sides can mix arbitrarily if the expected value of workers in the pool is exactly w_{min} (and this happens to be the workers' outside option value).

Then given the continua of firms and workers, taking the strategies of others as given, firms have a well-defined value from waiting and hiring (or not) from the pool - either something positive or 0. Thus, in the first period, a firm prefers to hire a referred worker if and only if that worker's value minus w_{min} exceeds the expected value from this second period potential pool hiring. If multiple firms are competing for a referred worker whose value minus w_{min}, then they will bid.

The key to characterizing the equilibrium is thus the first-period threshold such that firms attempt to hire a referred worker who has a value above that level, and do not hire workers below that level. That threshold corresponds to the value of waiting and possibly hiring from the pool. To characterize the threshold, the relevant function is the expected value of workers in the pool conditional on firms hiring referred workers who have values strictly above $\widetilde{v}$ and not hiring workers with values strictly below $\widetilde{v}$:

$$E_{\widetilde{v}}[v_i|i \in pool] := \frac{P(0)E[v_i] + (1-P(0))F(\widetilde{v})E[v_i|v_i < \widetilde{v}]}{P(0) + (1-P(0))F(\widetilde{v})}. \tag{1}$$

So, all equilibria are equivalent to using a threshold $\tilde{v}$ that is based on a fixed point of (1).[11] Thus, we use the term "equilibrium threshold" to refer to a fixed point of (1).

Proposition **1.** *There is a unique solution to*

$$\widetilde{v} = \max\left[w_{min}, E_{\widetilde{v}}[v_i|i \in pool]\right],$$

and that threshold value characterizes the following equilibrium behavior:

[10]Otherwise, there cannot exist an equilibrium in which workers mix since if workers mixed and were worth more than the minimum wage in expectation, then firms would deviate and offer a slightly higher wage and hire the worker for sure, but then the firm would want to offer the smallest wage higher than w_{min}, which does not exist. Thus, no such equilibrium exists, and all cases in which the expected value of the worker does not happen to be exactly the minimum wage involve workers accepting employment if indifferent.

[11]For discrete distributions F, there could be a multiplicity of thresholds that all correspond to the same decisions.

1. *a referred worker i is hired if its value $v_i > \widetilde{v}$, not hired if $v_i < \widetilde{v}$, and hired with some arbitrary probability if $v_i = \widetilde{v}$;*

2. *firms that are unsuccessful in hiring a referred worker hire from the pool if $E_{\widetilde{v}}[v_i|i \in pool] > w_{min}$, do not hire if $E_{\widetilde{v}}[v_i|i \in pool] < w_{min}$, and hire from the pool with some arbitrary probability if $E_{\widetilde{v}}[v_i|i \in pool] = w_{min}$.*

The equilibrium wages of a hired worker are $v_i - \widetilde{v} + w_{min}$ if i has more than one referral, and w_{min} otherwise.

Although we track of all cases in what follows, the reader may find it easier to concentrate on situations in which firms find it worthwhile to hire workers from the pool, since then $\widetilde{v} = E_{\widetilde{v}}[v_i|i \in pool]$ and the threshold is simply the expected productivity in the pool.

The remainder of the proof of Proposition 1 (beyond the discussion preceding the proposition) is relatively short and so we provide it here, since the intuition is also useful in what follows. The existence of an equilibrium threshold follows from the fact that the right hand side of the equation

$$\widetilde{v} = \frac{P(0)E[v_i] + (1 - P(0))F(\widetilde{v})E[v_i|v_i < \widetilde{v}]}{P(0) + (1 - P(0))F(\widetilde{v})}$$

is above $\widetilde{v}$ when $\widetilde{v}$ is small enough (as then the right hand side is $E[v_i]$) and below $\widetilde{v}$ when $\widetilde{v}$ is large enough (as then again, the right hand side is $E[v_i]$). Then given that the function is continuous in $\widetilde{v}$, there is a fixed point by the intermediate value theorem.[12] The fact that firms can mix exactly at the threshold (and also at w_{min} from the pool) comes from the fact that they are then indifferent, and it does not make a difference in the value in the pool.

The uniqueness of the fixed point is less obvious since the right hand side can be non-monotone in $\widetilde{v}$, but the argument is as follows. Suppose to the contrary that there are two distinct equilibrium thresholds, $\widetilde{v} < \widetilde{v}'$. We provide the argument for the case in which $\widetilde{v}' > w_{min}$ as otherwise it must be that $\widetilde{v} = \widetilde{v}' = w_{min}$, since no firm would ever hire a worker whose value is less than the minimum wage. The anonymous pool for threshold $\widetilde{v}'$ consists of workers without any referral, referred workers with values not exceeding $\widetilde{v}$, and referred workers with value between $\widetilde{v}$ and $\widetilde{v}'$ (and possibly some at exactly $\widetilde{v}'$ depending on the mixing). The expected productivity of workers without referrals and referred workers with values not exceeding $\widetilde{v}$ is equal to $\widetilde{v}$, since $\widetilde{v}$ is an equilibrium threshold. Averaging the value of this group with the expected value of workers with referral and values between $\widetilde{v}$ and $\widetilde{v}'$ gives the expected productivity in the pool if the threshold is $\widetilde{v}'$; but this cannot result in an expected value of $\widetilde{v}'$ as $\widetilde{v} < \widetilde{v}'$.[13]

[12] When there are atoms in the distribution, this is true when one varies the probability of hiring at the threshold from 0 to 1 at the atoms so that the right-hand side corresponds to a compact-valued upper hemicontinuous correspondence, which still admits a fixed point.

[13] Also note that no randomization of firms hiring workers exactly at the threshold is required in equilibrium as adding a mass of workers with value equal to the threshold does not affect the expected value in the pool.

2.2 Referral Inequality

The distribution of referrals across the population of workers seeking jobs, P, is central to our analysis. In particular, we investigate how the economy changes as referrals become more unequal and so we see how various aspects of the economy change as P is changed.

Let $P(2+) \equiv \sum_{k \geq 2} P(k)$ denote the fraction of workers who get two or more referrals.

The key characteristics of P that affect the matching and wages are $P(1)$ and $P(2+)$. Summed, these two track the fraction of job-seekers who have referrals and hence are vetted and have a chance of being hired at the first stage. This is the key to understanding productivity. The two separately are then key in determining the wage distribution. Once a person has two firms competing to hire them, having any additional bidders does not matter and so $P(2+)$ is sufficient (together with the cutoff) to predicting the fraction of workers who will get above minimum wage. Then $P(1)$ is needed to predict the cutoff.

Thus, for our analysis, when we change P, everything can be characterized by changes in $P(0)$, $P(1)$, and $P(2+)$. Since, these sum to 1, it is enough to track $P(0)$ and $P(2+)$.

Therefore, everything can be tracked via two ways of changing P: changing $P(0)$ and changing $P(2+)$. These can be thought of as two ways of changing referral inequality: concentration at the low end, in which more workers have no referrals (increasing $P(0)$); and concentration at the high end, in which more workers have multiple referrals (increasing $P(2+)$).

Definition **1.** *Distribution P'* concentrates referrals at the low end *compared to distribution P if $P'(0) \geq P(0)$. Concentration at the low end is* strict *if $P'(0) > P(0)$.*

Definition **2.** *Distribution P'* concentrates referrals at the high end *compared to distribution P if $P'(2+) \geq P(2+)$. Concentration at the high end is* strict *if $P'(2+) > P(2+)$.*

Neither one of these implies the other. If one considers a first order stochastic dominance change, then that will concentrate referrals at the high end but reduce concentration at the low end.

Many of the more interesting comparisons will be ones in which the overall total number of referrals is held constant (e.g., each current employee offers one referral), but their distribution is shifted. For example, if P' is a mean-preserving spread (mps) of P with both having a mean of $1/n$ (the fraction of currently employed who can offer referrals over the number looking for jobs from the next generation), then P' concentrates referrals at both the low and the high end compared to P. But note that even holding the means fixed at $1/n$, it is possible to have P' concentrate referrals at the low end but not at the high end compared to P, or vice versa.

Referral inequality can be caused by a variety of factors. One is that people can be more likely to refer their popular friends than a friend at random, as their more popular friend might be in a better position to return the favor some day. This causes referrals to be more concentrated in both senses. Referral inequality can also be induced by homophily. As we prove in Section 4, if people have groups, ethnic or geographic, and one group has an advantage in terms of its initial employment, and people tend to refer people of their own group then this will concentrate of referrals in both senses compared to a world that is group-blind. As a third example, it could be that some groups get relatively more referrals, randomly distributed within the group, per capita than others (e.g., males compared to females as seen in Lalanne and Seabright (2016)), which given that the overall mean is constant, also concentrates referrals both ends (again, see Section 4).

2.3 The lemons effect

Given that hiring on the referral market follows a simple cutoff rule: workers with referrals are hired if their productivity is above some cutoff, and rejected workers go to the pool, this lowers the value of workers in the pool. The impact of this selection on the pool productivity is what we term the lemons effect. So, it is the referred workers who have low productivity and are then rejected and thrown back into the pool who "spoil" the anonymous pool. This lemons effect plays a central role in our analysis, sometimes amplifying and sometimes mitigating, the various effects that we study.

Formally, we outline the lemons effect, and its comparative statics, as follows. The proof of the following proposition appears in the appendix.

Proposition **2.** *In equilibrium there is a (strict) lemons effect:*

$$E_{\widetilde{v}}[v_i | i \in pool] < E[v_i]$$

where $\widetilde{v}$ is the equilibrium threshold.

Furthermore, consider two distributions over the number of referrals a worker has, P and P', and let the unique equilibria be given by $\widetilde{v}$ and $\widetilde{v}'$ respectively. Then, if P' concentrates referrals at the low end with respect to P (i.e., $P(0) \leq P'(0)$), then $E_{\widetilde{v}}[v_i | i \in pool] \leq E_{\widetilde{v}'}[v_i | i \in pool]$, and strict concentration leads to a strict decrease in the lemons effect and increase in the expected value of workers in the pool.

The lemons effect is an important characteristic of labor markets with referrals: the unemployed pool of workers includes unlucky people who had no contacts, together with people who have been rejected. This inclusion of previously rejected workers makes hiring from the pool less attractive, and provides additional incentives for firms to hire referred workers. This feedback effect, means that the hiring threshold for referrals, $\widetilde{v}$, is less than the unconditional expected productivity value in the population. So, some workers who are hired via referrals have lower than average productivity, but benefit from the lemons effect. This leads to an increased inequality between the employment prospects of workers who have referrals (who have two opportunities to be hired, via referrals at an advantage and also from the pool) and those who do not (who can only be hired from the pool and are put together with rejected workers).

This relation is exacerbated by the lemons effect as a result of which firms knowingly hire below-average referred workers because they are better than the expected value in the pool, even though the pool also contains some above-average workers.

Furthermore, when comparing two economies, the one in which fewer workers have referrals – i.e., referrals are concentrated at the low end – has *less* of a lemons effect since fewer people are rejected and pushed into the pool (even accounting for the raised threshold, as that raised threshold is itself a reflection of the extent of the lemons effect[14]).

2.4 Constrained efficiency

In equilibrium, firms hire referred workers with productivity that is below its unconditional expectation as long as it is above the hiring threshold. As a result, one may expect that the equilibrium hiring is not efficient and that instead a different rule could increase productivity. However, as the proposition below shows, a social planner who takes the distribution of referrals as given would choose the same threshold for hiring on the referral market as in the unique equilibrium.

PROPOSITION **3.** *Total production is maximized when the hiring threshold on the referral market is given by the equilibrium threshold $\tilde{v}$.*

The lemons effect described in Section 2.3 goes hand in hand with total production in the economy: the average productivity of the workers in the pool equals the average productivity of the unemployed so that, as average productivity of all workers is held fixed, the average productivity of the employed increases with the size of the lemons effect.

Choosing a hiring threshold on the referral market different to the equilibrium threshold affects the average productivity in the pool as described in equation (1). To maximize total production, the threshold is chosen to minimize the average productivity in the pool: When the threshold is below the average productivity in the pool, increasing it (and thus sending workers with that value to the pool) further decreases the average; while, when the threshold is above the average, decreasing it decreases the average. To minimize the average productivity in the pool, one would thus set the hiring threshold equal to the average productivity in the pool. But this is exactly the condition that uniquely characterizes the equilibrium. Hence, the equilibrium threshold maximizes total production.

3 Impact of Referral Inequality

We now examine the relationship between the inequality in a society, in terms of the distribution of referrals, and the productivity and income inequality in the society. To keep the analysis uncluttered (as it is already rich), we consider a case throughout this section in which w_{min} is the value of workers' outside options: their productivity at home or their wage in some other form of employment if they are not hired on this market. This assumption allows us to assess the value of unhired workers and consider them in assessing the overall welfare/productivity/income of the society.

[14]If the lemons effect had actually gone up, then the threshold would have had to fall.

3.1 Concentrating Referrals Decreases Productivity

We first examine the impact on overall production as we concentrate referrals.

Given an equilibrium threshold $\widetilde{v}$ and a distribution of referrals P, the overall production is:

$$\begin{aligned} Y_{\widetilde{v}}(P) &= n(1-P(0))(1-F(\widetilde{v}))E[v_i|v_i \geq \widetilde{v}] \\ &+ (1-n(1-P(0))(1-F(\widetilde{v})))\widetilde{v} \\ &+ (n-1)\, w_{min}. \end{aligned} \tag{2}$$

The first term in (2) describes the expected productivity of a worker who is hired through the referral market multiplied by the mass of workers hired through this channel; the second term describes the worker's productivity when she is hired through the anonymous pool (or turns out to be equal to her outside option if no hires are made through the pool), again multiplied by the mass of workers hired through this channel; and the third term describes the outside option value of the unhired workers.

To understand how productivity changes with changes in the referral distribution it is useful to note that a single referral is productivity enhancing: If a worker has high productivity (above the equilibrium threshold), then that worker is hired. If the referred worker has low productivity, then the firm has the option of taking a match from the pool. When some worker gets two or more referrals instead of one, then that does not improve the matching of that worker beyond the first referral: if they are low value then the referrals are all wasted, and if they are of high value then any referral past the first one is wasted. This then suggests that any change that increases $P(0)$ (decreasing the number of people getting a referral) is detrimental to productivity: the total value of production in the society goes down as fewer high value workers are hired overall in the economy.

There are some subtleties since the threshold is raised as we increase from $P(0)$ to $P'(0)$ and the lemons effect is lowered. The proof shows that these are more than overcome. The basic intuition is that having a bigger $P'(0)$ means that fewer workers get vetted overall, and more ultimately end up being hired without any vetting (lemons or otherwise). Ultimately, jobs that are not filled by a high value worker end up being filled by someone of lower value or not filled at all (depending on the equilibrium, presuming that we are not changing from one type of equilibrium which hires from the pool to the other which does not), and replacing those with high value is good. The full proof takes care of all the possible cases, including in which we change forms of equilibria.

This is captured in the following proposition.

Proposition 4. *Consider two distributions P and P'. If P' concentrates referrals at the low end with respect to P (i.e., $P'(0) \geq P(0)$), then the total production in the economy associated with P' is smaller than that associated with P: $Y_{\widetilde{v}'}(P') \leq Y_{\widetilde{v}}(P)$, and strict concentration leads to a strict decrease in productivity.*

Proposition 4 states that concentrating referrals among a smaller part of the population, leads to decreased productivity.

This result relies on the assumption that the outside option of workers equals the minimum wage. For instance, if w_{min} is above outside option value to unemployed workers, then overall production could be higher under P' under some circumstances. That would require that there was no hiring from the pool under P, and that the difference between P' and P be large enough to ease the lemons sufficiently to make hiring under the pool attractive under P', and that the gain in value from hiring those workers (the difference between the minimum wage and their outside values) is sufficiently large to offset the loss from reduced vetting.

In addition to the changes in productivity, it is possible that a concentration in referrals can affect the size of an industry. If a change from P to P' causes a sufficient change in the lemons effect to move from not hiring from the pool to hiring from the pool, then that could increase the size of the industry in terms of overall employment.

3.2 Concentrating Referrals has an Ambiguous Impact on Wage Inequality

Next, we explore how concentrating referrals impacts wage inequality, which in this model is equivalent to income inequality (more comments on this below).

The first thing to note is the key factor in determining the wage distribution in our model is the number of people who get more than one referral: those are the people who have some competition for their services and earn above the minimum wage. Thus, rather than how many people get referrals, what is key here is what fraction of people get multiple referrals. So, it is our concentration at the high end that plays the central role in understanding inequality. Concentration at the low end still determines the expected value from hiring from the pool, and thus impacts wages that people with multiple referrals obtain, so it is also involved but with a different (marginal) effect.

To measure inequality, we use the Gini coefficient of wages. As comparing distributions is generally a incomplete exercise, using such a coefficient allows one to make comparisons of partially ordered distributions. However, even using this standard one-dimensional measure, and looking at simple settings, we will see that inequality can move in different ways from the same comparative statics, depending on the specifics of the environment.

One might expect that concentrating referrals at the high end would be sufficient to increase inequality; but things are not so direct. First, even if fewer workers are high wage levels, increasing the size of that group can actually increase or decrease inequality, depending on relative wages and the relative size of the group to begin with, as the Gini coefficient makes relative comparisons. In addition, the lemons effect can be alleviated. This tends to decrease the wages of the workers who have multiple referrals, lowering the high wage, which can then decrease inequality. Thus, there can be different sorts of countervailing effects.

To get a characterization of when inequality is raised by concentrating referrals at the high end, for this subsection only, we consider a situation in which productivity takes on two values $v_H > v_L$, with fraction f_H of the population being of the higher productivity.

Combined with the assumption that w_{min} is the value of workers' outside options, this implies that there are just two levels of wages. All of the countervailing effects described above already exist with just two types, and this makes it possible to see the intuition.

Letting $\pi_H = P(2+)f_H$ be the fraction of the total set of workers (employed and unemployed) who earn the high wage, and $W_L = \frac{w_{min}}{w_H}$ be the low wage relative to the high wage, one can write the Gini coefficient of an economy as follows:

$$Gini = \frac{\pi_H (1-\pi_H)(1-W_L)}{\pi_H + (1-\pi_H)W_L}. \tag{3}$$

Let us consider what happens when the fraction of people earning the high wage is increased (π_H is increased), which is a consequence of concentrating referrals at the high end. Straightforward calculations show that:

$$\frac{\partial Gini}{\partial \pi_H} = \frac{(1-W_L)(W_L(1-\pi_H)^2 - \pi_H^2) - \frac{\partial W_L}{\partial \pi_H}\pi_H(1-\pi_H)}{(\pi_H + (1-\pi_H)W_L)^2}.$$

The change in the Gini consists of two parts: the effect of changing the relative fraction of people earning the high wage compared to the low wage, and then the effect of changing the wage via the lemons effect.

The first part (ignoring $\frac{\partial W_L}{\partial \pi_H}$) is positive for low π_H and high W_L, but then becomes negative as π_H increases and W_L decreases. However, then the overall expression is lowered by the factor of $\frac{\partial W_L}{\partial \pi_H}$, which accounts for the lemons effect. The change in the lemons effect depends on the level of concentration at the low end and not at the high end.

To fully sign $\frac{\partial W_L}{\partial \pi_H}$, we also need to know what happens to concentration at the low end. In particular, the change in W_L comes from the change in $1/w_H$, which is governed by the concentration at the low end. Knowing how $P(2+)$ changes, tells us about π_H, but we need to know how $P(0)$ changes, to determine the change in w_H. If concentration occurs at both ends, then, an increase in π_H also increases $P(0)$ which, using Proposition 2, decreases the lemons effect and thus the high wage so that $\frac{\partial W_L}{\partial \pi_H} > 0$. This then counteracts things and can reverse

These results are summarized in the following proposition, which should be clear from the above discussion, so we omit a formal proof.

PROPOSITION **5.** *Ignoring changes in the lemons effect (e.g., holding $P(0)$ constant), the Gini increases as we concentrate referrals at the high end if and only if $\frac{w_{min}}{w_H}(1-P(2+)f_H)^2 - (P(2+)f_H)^2 > 0$, which holds for low values of $P(2+)f_H$ and high values of $\frac{w_{min}}{w_H}$. The derivative of the Gini is negative with respect to* concentration at the low end *(decreases in $P(0)$).*

Proposition 5 implies that if we consider a first order stochastic dominance shift in P, so that we concentrate referrals at the high and lower concentration at the low end, presuming $\frac{w_{min}}{w_H}(1-P(2+)f_H)^2 - (P(2+)f_H)^2 > 0$, then the Gini will increase. If instead, (again presuming that $\frac{w_{min}}{w_H}(1-P(2+)f_H)^2 - (P(2+)f_H)^2 > 0$) we consider a mean preserving spread in P, then we can end up with more people earning the higher wage which pushes inequality up, but they earn a lower higher wage due to an improved value of the pool which pushes inequality down. Either force can dominate depending on the particular parameters in question.

Firms' Profits Our analysis above is focused on wages. Firms in this economy are earning profits since they can hire workers at w_{min} and earn a higher expected value. Accounting for who gets those profits as income can affect the inequality calculations.

Firms earn the same profit from workers they compete (those with high values and $P(2+)$, as pulling from the pool). This means that expected profits take a simple form (here presuming that the expected value from the pool exceeds the minimum wage): Profits are:

$$\Pi = nP(1) f_H v_H + (1 - P(1) f_H)\widetilde{v} - w_{min}. \tag{4}$$

So, firms' profits depend entirely on how many people get just one referral, as well as what the cutoff is and hence the lemons effect. Thus, firms' profits can be fully characterized if we know $P(0)$ and $P(2+)$ (and hence $P(1)$).

To provide some simple intuition, let us consider a distribution for which workers receive either $0, 1$ or k referrals. Together with the assumption that the total number of referrals is constant equaling 1, this implies that P is completely characterized by one parameter: $P(k)$. In that case, simple but tedious calculations show that $\partial\Pi/\partial P(k) = -\frac{f_L f_H (v_H - v_L)(f_L + (n-1)k)}{(P(0) + (1-P(0)) f_L)^2} < 0$. This is not obvious, since once again the lemons effect has a counteracting force, but in this situation the derivative can be unambiguously signed.[15]

This means that if profits are distributed uniformly across all workers, then concentrating referrals (an increase in $P(k)$) decreases profits, and so decreases income uniformly across all workers which increases the Gini and thus inequality.[16] Thus, compared to Proposition 5, this is another pressure increasing inequality. If profits instead go to some special class of citizens who are owners of the firms, then it depends on who they are, and so then the further effects are ambiguous.

4 Inequality Across Groups and Immobility

In our analysis to this point, we have examined changes in the distribution of referrals, but without indicating what might lead a society to have more or less concentration in referrals. Also, our discussion was focused on productivity and inequality, but we have not yet discussed immobility. We now expand to offer insight on both of these issues by explicitly modeling

[15]The weaker assumption that concentration at the high end implies concentration at the low end is not sufficient to sign the effect on profit.

[16]An easy way to see this is to note that the Gini decreases in W_L (see equation (3)) and so lowering the level of all incomes decreases W_L, thereby increasing the Gini.

types of agents (e.g., ethnicity, gender, etc.). Coupled with homophily (tendencies to refer own type) or some other asymmetry (all biased towards referring some type), this offers sources of concentration in referrals. Moreover, it enables us to see how less employment among a particular group in one generation translates into lower employment (and lower wages) in the next generation: hence tracking immobility.

For simplicity we consider two groups, but the results are easy to extend. We refer to one group as *blue* and the other as *green*, with respective masses $n_b > 0$ and $n_g > 0$ of workers per generation, such that $n_b + n_g = n$. Let e_b and e_g be the current masses of employed blue and green workers respectively. Then there is a current employment bias towards blues if $\frac{e_b}{e_g} > \frac{n_b}{n_g}$.

To model homophily in referrals, we introduce group-dependent referral-bias parameters $h_b \in [0, 1]$ and $h_g \in [0, 1]$. A fraction h_b of blue workers refer blue workers; the remaining $(1 - h_b)$ fraction of blue workers refer green workers; with h_g defined analogously. If there was no current employment bias, then the blue workers looking for jobs would receive a total of

$$r_b = \frac{h_b n_b + (1 - h_g) n_g}{n}$$

referrals and the green workers would receive

$$r_g = \frac{h_g n_g + (1 - h_b) n_b}{n}.$$

If $\frac{r_b}{r_g} = \frac{n_b}{n_g}$, then there is a balance in referrals, namely each population receives a number of referrals proportional to their representation in the population, conditional on balanced employment. A bias in employment rates and/or referral (homophily) rates can advantage one group of the other.

To make the model concrete, suppose additionally that referrals to a population are made uniformly at random across workers from that population such that the degree distribution within a population is Poisson. We relax this condition and study general degree distributions in the appendix.

Equilibrium is determined by the overall distribution of referrals, as discussed in Section 2, as types do not impact firm payoffs. Varying employment rates and homophily impacts the distribution of referrals and so can impact the productivity and wage inequality of all workers. For example, with full homophily ($h_b = h_g = 1$), increased historic employment inequality (increasing $\frac{e_b}{e_g}$) results in a concentration of referrals at both the low and high ends, which harms productivity and can increase the Gini coefficient.

Of particular interest to us, however, is how historic employment rates and homophily impact type-specific outcomes. These factors impact the relative rates of referrals across blues and greens, and thereby employment and wage inequality. If historic employment rates are balanced ($\frac{e_b}{e_g} = \frac{n_b}{n_g}$) and referrals are balanced ($\frac{r_b}{r_g} = \frac{n_b}{n_g}$), then outcomes are equal for the two groups, regardless of the degree of homophily. On the other hand, should either of these conditions strictly favor blues ($\frac{e_b}{e_g} > \frac{n_b}{n_g}$ or $\frac{r_b}{r_g} > \frac{n_b}{n_g}$), then blues have more referrals and more individuals with multiple referrals than greens (per capita), resulting in better wages and higher employment rates. It also leads to a worse lemons effect for blues since relatively more of them are screened, and rejected. As the expected productivity of employed and unemployed workers needs to average out to the same value across groups, this also implies that employed blue workers have higher productivity than employed green workers.[17] Given that referral-hired workers tend to outperform other workers (again, see the discussion in the Introduction), this is consistent with observation. Furthermore, the difference in average productivity of employed workers across types, could perpetuate a biased perception of their respective abilities if observers do not understand the selection process. In particular, estimating productivity by looking at employed populations (the way that productivity is standardly measured) systematically overestimates blues' productivity and underestimates greens' productivity.

This discussion is summarized in the following proposition.

Proposition **6.** *If there is (weak) employment bias ($\frac{e_b}{e_g} \geq \frac{n_b}{n_g}$) and (weak) referral imbalance ($\frac{r_b}{r_g} \geq \frac{n_b}{n_g}$) in favor of blues, then*

- *the wage distribution of blue workers first order stochastically dominates the wage distribution of green workers,*
- *and the employment rate of blue workers is (weakly) higher than the employment rate of green workers.*

Both of these advantages to blues are strict if either of the above conditions hold with strict inequality; and, then the average productivity of employed blue workers is greater than that of employed green workers (and, correspondingly, the average productivity of unemployed green workers is greater than that of blues).

We can extend our analysis one step further: thus far we have discussed a model with just one hiring period. If instead we repeat this game ad infinitum, using workers hired in period t to refer workers in period $t+1$, and firms are myopic in each period, then we derive the long-run employment rates. In this setting, if referrals are balanced, then the fact that hiring from the pool is unbiased causes employment rates to eventually equalize. On the other hand, if referrals are unbalanced, there is perpetual over-representation of blues among employed workers, and long-run immobility.

Proposition **7.** *Suppose that both initial employment rates are positive, or that there is an interior homophily rate for a group having all the initial employment. There is long-run bias in employment rates in the infinite-period game* if and only *if there is strict referral imbalance ($\frac{r_b}{r_g} \neq \frac{n_b}{n_g}$).*

[17]On the flip side, the productivity of unemployed greens dominates that of unemployed blues, suggesting firms should prefer to hire minorities from the pool if they can use types to make hiring decisions.

Proposition 7 shows that if referrals are balanced initial employment rates are irrelevant in the long run. So, in considering long-run differences across groups, all that matters is imbalance in referral rates and not short-run employment rates (provided they are not degenerate).

This ignores other incentives in the matching process that may induce long-run inequality in employment rates such as poverty traps.[18] For instance, if workers must make a costly investment, e.g., in education, to realize a productive value, then we may observe perpetual inequality. In particular, since Proposition 6 implies that the expected wage of blue workers of participating in the labor market is greater than that of green workers, if there is some cost to educating one's self before knowing whether one will have a referral, green workers may drop out while blue workers do not.

5 Firing Workers

We now consider how the analysis is enriched when some firms can fire workers and hire a different worker through open application.

In particular, let us suppose that after some time has elapsed, firms have learned the value of any worker they have hired from the pool (they already know the value of a referred worker if they kept him or her), and can then choose whether to fire that worker and hire a new one from the pool for the remaining time. Let λ be the fraction of time that is left in the period for which they would get the replacement worker, and $1 - \lambda$ is the fraction of the period that has to elapse before they can fire a worker. In equilibrium, as we prove below, they would never fire a referred worker that they kept initially, so the decision will only be relevant for workers they hired from the initial pool.

So the timing is as follows:

- Firms get referrals from their old employees.
- Firms can choose whether to (try to) hire that worker.
- Firms that chose not to hire or lost in competition to hire a worker can choose whether to hire from the initial pool.
- After $1 - \lambda$ of the period has elapsed, firms can choose to fire their current worker.
- Firms that have fired a worker can choose to hire from the pool again (which now has all currently unemployed workers, including ones just fired).

We refer to these pools of anonymous workers as pool 1 and pool 2, respectively. We assume that the distribution of values is high enough so that firms prefer to hire from these

[18] For example, see Calvo-Armengol and Jackson (2004) for a discussion of one in a job-referral setting.

pools to having no worker at all, in order to simplify the exposition; but the full details are worked out in the appendix.

A firm's production is given by $1-\lambda$ of the value of its worker before any firing, and then λ of its worker that it has after any firing and rehiring decisions are made.

There is another closely related variant of the model for which the proposition below also holds, which is one in which some fraction λ of firms immediately learn the value of their worker (from the pool, and they already know the value from the referral) and can then immediately fire that worker and draw again from the pool. In terms of expected values, this variant of the model looks exactly the same from a firm's perspective, as they just have λ weight on an expectation of firing and rehiring, as opposed to a fraction of time. This second variant does have some differences in terms of values in the pools, as fewer firms have an opportunity to fire workers; but the basic structure outlined in Proposition 8 is exactly the same.

As before, an equilibrium is characterized by threshold strategies of which there are now two: Referred workers are hired if their value is above $\tilde{v}_1$ and otherwise firms hire from the first pool. Then at the second decision point firms fire workers, when given the opportunity, if the worker's value is below $\tilde{v}_2$. Again, firms can mix with some probability when indifferent. The equilibrium thresholds, as before, are unique, so that we can perform comparative statics with respect to unique equilibrium quantities.

Note that our base model is nested in this formulation with $\lambda = 0$. Let $\tilde{v}$ denote the hiring threshold for that version, essentially without firing.

For $\lambda \in (0, 1]$, not hiring a referred worker and instead going to Pool 1, now features an *option value*. The firm randomly draws a worker but may get a second draw to replace him or her if the worker's value is too low. As a result, firms have a higher threshold on the referral market compared to the base model. This lessens the Lemons Effect in the first pool and so the first pool value is actually higher than $\tilde{v}$. The second pool, however, has a worse lemons effect since it then includes all fired workers as well as rejected referral workers. The fact that the second threshold is less than $\tilde{v}$ takes some proof, but is true. Essentially, the extra opportunity to fire improves the overall productivity, and so the workers left in the second period pool are worse than in the case where there is just one chance to hire.

We summarize these findings in the following proposition.

Proposition **8.** *When firms are given the option to fire their worker, with a parameter* $\lambda \in (0, 1]$, *then*

$$\tilde{v}_2 < \tilde{v} < \tilde{v}_1;$$

and

$$\tilde{v}_1 = (1-\lambda)E_{\tilde{v}_1}[v_i | i \in \textit{pool } 1] + \lambda E_{\tilde{v}_1}[\max\{v_i, \tilde{v}_2\} | i \in \textit{pool } 1]; \tag{5}$$

$$\tilde{v}_2 = E_{\tilde{v}_1, \tilde{v}_2}[v_i | i \in \textit{pool } 2], \tag{6}$$

where $E_{\tilde{v}_1}[v_i | i \in \textit{pool } 1]$ *is defined as in equation* (1) *and* $E_{\tilde{v}_1,\tilde{v}_2}[v_i | i \in \textit{pool } 2]$ *as the expected* *value in pool 2 if the hiring threshold on the referral market is given by $\tilde{v}_1$ and the firing threshold by $\tilde{v}_2$.*

Equations (5) and (6) give the consistency conditions for proposed equilibrium thresholds ensuring that firms hire their referral worker if the value is above $\tilde{v}_1$ and fire a worker from the pool if their value is below $\tilde{v}_2$.

6 Discussion and Policy Thoughts

We have shown how referrals can lead to inequality, inefficient production, and persistent differences across groups. We have also shown that in many cases there is a relationship between inequality and inefficiency: concentrating referrals decreases productivity and increases inequality. However, we have also showed that the detailed relationships between these are more subtle that one might superficially imagine, due to the lemons effect, in which hiring from a pool of workers to which a firm has no referral can be more or less attractive depending on the extent to which that pool contains workers that were rejected by other firms. A sufficiently large change in the lemons effect can reverse one or both of the relationships between the concentration of referrals and productivity and inequality.

The role of the lemons effect in inequality, inefficiency, and immobility can be counteracted by policies that encourage hiring from the pool, e.g., by overcoming its information disadvantage via subsidizing internships and hiring. In addition, various forms of affirmative action that make hiring from the pool more attractive help overcome employment imbalances, which via homophily and network effects, further reduce future imbalances. Thus, even without subsidies it can be improving to regulate more intense search from the pool (e.g., requiring interviewing of qualified minorities for open positions as in many sports leagues for coaching positions). Also, even though efficiency is not reached in the setting with firing, the ability to fire does improve overall productivity and reduce inequality; which provides insight into one advantage of flexible labor markets. Finally, providing more information about non-referred workers, via various sorts of certification, can help attenuate informational disadvantages.

References

Arrow, Kenneth J. and Ron Borzekowski, "Limited network connections and the distribution of wages," *FEDS Working Paper*, 2004.

Beaman, Lori A., "Social Networks and the Dynamics of Labour Market Outcomes: Evidence from Refugees Resettled in the U.S.," *Review of Economic Studies*, 2012, *79 (1)*, 128–161.

__ , **Niall Keleher, and Jeremy Magruder**, "Do job networks disadvantage women? Evidence from a recruitment experiment in Malawi," *Unpublished. http://faculty. wcas. northwestern. edu/˜ lab823/BKM_recruitment_Oct2013. pdf*, 2016.

Brown, Meta, Elizabeth Setren, and Giorgio Topa, "Do informal referrals lead to better matches? Evidence from a firm's employee referral system," *Evidence from a Firm's Employee Referral System (August 1, 2012). FRB of New York Staff Report*, 2012, (568).

Calvo-Armengol, Antoni and Matthew O. Jackson, "The effects of social networks on employment and inequality," *The American Economic Review*, 2004, *94* (3), 426–454.

__ **and** __ , "Networks in labor markets: Wage and employment dynamics and inequality," *Journal of Economic Theory*, 2007, *132* (1), 27–46.

Corak, Miles, "Inequality from Generation to Generation: The United States in Comparison," *IZA DP No. 9929*, 2016.

Currarini, Sergio, Matthew O. Jackson, and Paolo Pin, "An economic model of friendship: Homophily, minorities, and segregation," *Econometrica*, 2009, *77* (4), 1003–1045.

Farber, Henry S, "Mobility and stability: The dynamics of job change in labor markets," *Handbook of labor economics*, 1999, *3*, 2439–2483.

Fernandez, Roberto M, Emilio J. Castilla, and Paul Moore, "Social capital at work: Networks and employment at a phone center," *American journal of sociology*, 2000, pp. 1288–1356.

Gibbons, Robert and Lawrence F Katz, "Layoffs and lemons," *Journal of labor Economics*, 1991, *9* (4), 351–380.

Granovetter, Mark, *Getting a Job: A Study of Contracts and Careers*, University of Chicago Press, 1974.

Granovetter, Mark S., *Getting a job: A study of contacts and careers (Second Edition)*, University of Chicago Press, 1995.

Heath, Rachel, "Why do firms hire using referrals? Evidence from Bangladeshi Garment Factories," *Journal of Political Economy*, forthcoming.

Ioannides, Y.M. and L. Datcher-Loury, "Job information networks, neighborhood effects and inequality," *Journal of Economic Literature*, 2004, *424*, 1056 – 1093.

Jackson, Matthew O., *The Human Network: How Your Social Position Determines Your Power, Beliefs, and Behaviors*, Pantheon Books: New York, 2019.

Krueger, Alan B, Judd Cramer, and David Cho, "Are the long-term unemployed on the margins of the labor market?," *Brookings papers on economic activity*, 2014, *2014* (1), 229–299.

Lalanne, Marie and Paul Seabright, "The old boy network: The impact of professional networks on remuneration in top executive jobs," *SAFE Working Paper No 123*, 2016.

Laschever, Ron A., "The Doughboys Network: Social Interactions and Labor Market Outcomes of World War I Veterans," *SSRN discussion paper 1205543*, 2011.

Montgomery, James D., "Social networks and labor-market outcomes: Toward an economic analysis," *The American Economic Review*, 1991, *81* (5), 1408–1418.

Munshi, Kaivan, "Networks in the Modern Economy: Mexican Migrants in the US Labor Market," *Quarterly Journal of Economics*, 2003, *118* (2), 549–599.

Pallais, Amanda and Emily Glassberg Sands, "Why the Referential Treatment: Evidence from Field Experiments on Referrals," *Journal of Political Economy*, 2016, *124:6*, 1793–1828.

Patacchini, Eleonora and Yves Zenou, "Ethnic networks and employment outcomes," *Regional Science and Urban Economics*, 2012, *42* (6), 938–949.

Rees, Albert, "Information networks in labor markets," *The American Economic Review*, 1966, *56* (1/2), 559–566.

Rogerson, Richard, Robert Shimer, and Randall Wright, "Search-Theoretic Models of the Labor Market: A Survey," *Journal of Economic Literature*, December 2005, *43* (4), 959–988.

Topa, Giorgio, "Chapter 22 - Labor Markets and Referrals," in Jess Benhabib, Alberto Bisin, and Matthew O. Jackson, eds., *Jess Benhabib, Alberto Bisin, and Matthew O. Jackson, eds.*, Vol. 1 of *Handbook of Social Economics*, North-Holland, 2011, pp. 1193 – 1221.

A Idiosyncratic values

In our base model, we assume that the value of a worker is constant across firms. However, the value of a worker may depend on an idiosyncratic fit of that worker at a particular firm. In this section, we explore this assumption further.

Consider our model as introduced in Section 2; however, now with the value of a worker idiosyncratic across firms. In particular, let worker i's value at firm j be given by v_{ij} which is distributed according to some distribution F (finite mean and non-degenerate) and independent across i, j.

First, note that the value of a worker in the pool is redrawn according to F so that all workers in the pool have an expected value equal to the unconditional expectation of F, that is

$$E^{\tilde{v}}[v_i | i \in \text{pool}] = E[v_i] \coloneqq \bar{v}$$

for any hiring threshold $\tilde{v}$.

Thus, there is no lemons effect and furthermore the equilibrium threshold is given by the unconditional expected value (for simplicity, throughout this section, assume that $\bar{v} \geq w_{min}$ so that firms hire from the pool).

However, inequality in referrals, here understood as a mean-preserving spread relationship between two distributions, negatively affects productivity. For $k = 1, 2, \ldots$, let $v_i(k)$ denote the random variable of the highest value a worker with exactly k referrals has at one of the firms shes referred to. For notational simplicity, define $E[v_i(0)|v_i(0) \geq \bar{v}] \coloneqq 0$. For a given distribution of referrals P, total production when the outside option of workers is given by w_{min} is given by

$$\begin{aligned} Y^I(P) = &\sum_{k=0}^{\infty} P(k)(1 - F(\bar{v})^k) E[v_i(k)|v_i(k) \geq \bar{v}] \\ &+ \left(1 - \sum_{k=0}^{\infty} P(k)(1 - F(\bar{v})^k)\right) \bar{v} \\ &+ (n-1) w_{min}. \end{aligned}$$

Let P' be a mean-preserving spread of P.

Proposition 9. *Total production in the economy is larger given P than given P', i.e.*

$$Y^I(P) \geq Y^I(P').$$

Proof. It is enough to show that

$$(1 - F(\bar{v})^k) E[v_i(k)|v_i(k) \geq \bar{v}] \tag{7}$$

is concave in k.

Note that

$$F^k(v) - F^{k-1}(v) = -F^{k-1}(v)(1 - F(v)).$$

Taking the difference of (7) for consecutive values of k, we have

$$\begin{aligned}
&(1-F(v)^{k+1})E[v_i(k+1)|v_i(k+1)\geq \bar{v}]-(1-F(v)^k)E[v_i(k)|v_i(k)\geq \bar{v}]\\
=&\int_{\bar{v}}^{\infty} xd(F^{k+1}-F^k)\\
=&-\int_{\bar{v}}^{\infty} xd(F^k(1-F))\\
=&-\left[xF^k(x)(1-F(x))\right]_{x=\bar{v}}^{\infty}+\int_{\bar{v}}^{\infty} F^k(x)(1-F(x))dx\\
=&\bar{v}F^k(\bar{v})(1-F(\bar{v}))+\int_{\bar{v}}^{\infty} F^k(x)(1-F(x))dx
\end{aligned}$$

which is decreasing in k and the result follows. □

Next and similarly to before, we introduce horizontally differentiated groups, blue workers and green workers, to discuss about inequality across groups.

Suppose that blues have more referrals than greens. This may be due for example to historic employment bias and or referral imbalance. Then, the labor market outcomes of blues dominate those of greens as the following proposition makes precise.

PROPOSITION **10.** *If the referral degree distribution of blue workers first-order stochastically dominates the one of green workers, then wage distribution of blue workers also first-order stochastically dominates the one of green workers.*

Proof. The proposition follows from the fact that the wage distribution of a worker with exactly $k+1$ referrals first-order stochastically dominates the wage distribution of a worker with exactly k referrals. □

Similarly to before, Proposition 10 implies that the employment rate of blue workers is higher than the employment rate of green workers following such referral distributions

COROLLARY **1.** *If the referral degree distribution of blue workers first-order stochastically dominates the one of green workers, then the employment rate of blue workers in the next period is larger than the employment rate of green workers.*

Proof. The employment rate is simply equal to the fraction of workers with non-zero wage which is larger for blue workers than for green workers by Proposition 10. □

B Proofs

Proof of Proposition 2. Clearly $E_{\tilde{v}}[v_i|i \in \text{pool}] \leq E[v_i]$. $E_{\tilde{v}}[v_i|i \in \text{pool}] = E[v_i]$ only if $P(0)=1$ or F is degenerate, both of which are ruled out by assumption. Thus $E_{\tilde{v}}[v_i|i \in \text{pool}] < E[v_i]$.

For any $\tilde{v}$, as $E_{\tilde{v}}[v_i|i \in \text{pool}] < E[v_i]$ the right-hand side of (1) is strictly increasing in $P(0)$ and the result follows.

□

Proof of Proposition 4. There are three cases to consider: given both P and P', firms either hire from the pool or do not hire from the pool, or they hire under the pool under P' but not under P.

Suppose first that firms hire from the pool in the economies associated with P and P'. As

$$F(\widetilde{v})E[v_i|v_i < \widetilde{v}] = E[v_i] - (1 - F(\widetilde{v}))E[v_i|v_i \geq \widetilde{v}],$$

some simple algebra shows that we can rewrite (1) as

$$(P(0) + (1 - P(0))F(\widetilde{v}))E_{\widetilde{v}}[v_i|i \in \text{pool}] = E[v_i] - (1 - P(0))(1 - F(\widetilde{v}))E[v_i|v_i \geq \widetilde{v}].$$

We substitute the above expression for part of the first term in (2) and after some further algebra derive a simple expression for total output given by

$$nE[v_i] - (n - 1)E_{\widetilde{v}}[v_i|i \in \text{pool}] + (n - 1)w_{min}.$$

By Proposition 2, $E_{\widetilde{v}}[v_i|i \in \text{pool}] \leq E_{\widetilde{v}'}[v_i|i \in \text{pool}]$ implying $Y_{\widetilde{v}'}(P') \leq Y_{\widetilde{v}}(P)$.

Suppose now that firms do not hire from the pool under either economy. Total production is now given by

$$n(1 - P(0))(1 - F(w_{min}))E[v_i|v_i \geq w_{min}] + (n - n(1 - P(0))(1 - F(w_{min}))w_{min},$$

which is clearly decreasing in $P(0)$; again, implying $Y_{\widetilde{v}'}(P') \leq Y_{\widetilde{v}}(P)$.

Lastly, consider the case in which workers from the pool are hired in the economy associated with P' but not with P. Note that if $E_{\tilde{v}}[v_i|i \in \text{pool}] = w_{min}$, then both expressions for production, one in which firms hire from the pool and one in which not, evaluate to $nE[v_i]$ so that productivity changes continuously in whether workers are hired from the pool. Thus, as we already derived the comparative statics for when hiring from the pool does not change, we thus conclude that also in this case $Y_{\widetilde{v}'}(P') \leq Y_{\widetilde{v}}(P)$. □

Proof of Proposition 8. Existence of equilibrium thresholds is proven analogously as in Proposition 1. Similarly, equations (5) and (6) are simply optimality conditions that need to be satisfied for firms to adhere to the equilibrium thresholds thus showing equilibrium.

Lastly, to order the thresholds, starting with $\lambda = 0$, it is easy to see that the equilibrium threshold on the referral market as well as the firing threshold are given by $\tilde{v}$. Holding $\tilde{v}_1, \tilde{v}_2$ constant, it also holds that the right-hand side in (5) increases in λ. For a given $\tilde{v}_1$, let $\tilde{v}_2$ be determined in (6). Then, increasing λ increases the right-hand side of (5) so that the value of the equilibrium threshold on the referral market also has to increase. □