Essays on Social Economics

By

Nicholas Hingston Tenev

A dissertation submitted in partial fulfillment of

the requirements for the degree of

Doctor of Philosophy

(Economics)

at the

UNIVERSITY OF WISCONSIN – MADISON

2018

Date of final oral examination: June 12, 2018

The dissertation is approved by the following members of the Final Oral Committee:

Steven Durlauf, Professor, Public Policy

Chao Fu, Associate Professor, Economics

Rasmus Lentz, Professor, Economics

Christopher Taber, Professor, Economics

Abstract

This dissertation comprises three papers on social economics. Each studies a different setting in which the influence of individuals' peers on their behavior shapes aggregate outcomes.

The first chapter estimates the extent to which differences in jobs found through friends can account for the aggregate wage gap between black workers and others in the US. Data from the NLSY79 are used to estimate a job search model in which individual productivity is distinguished from social capital by comparing the wages and frequency of jobs found directly with those of jobs found through friends. Jobs found through friends tend to pay more, but this premium is lower for black workers.

The second chapter, joint with Alexander Clark, develops a model in which costly voting in a large two-party election is a sequentially rational choice of strategic, self-interested players who can reward fellow voters by forming stronger ties in a network formation coordination game. The predictions match a variety of stylized facts, including explaining why an individual's voting behavior may depend on what she knows about her friends' actions. Players have imperfect information about others' voting behavior, and we find that some degree of privacy may actually be necessary for voting in equilibrium, enabling hypocritical but useful social pressure. Our framework applies to any costly prosocial behavior.

The final chapter posits that the widespread usefulness of new pricing technology crucially contributed to the 2008 financial crisis, by allowing financial services workers and regulators alike to shirk in vetting its proper use. In the model, a principal attempts to induce costly effort from a group of agents with the threat of punishment. With a convex cost of punishing agents, she may be unwilling to simultaneously punish large groups of agents, leading them

to shirk only when coordination is possible. In this setting, a new technology can actually cause an aggregate downturn specifically because it is widely useful: agents do not research it properly, knowing they will not all be punished even if their project fails. Furthermore, even agents who learn that they are using flawed technology may continue to do so.

Acknowledgements

This work has benefited enormously from discussions with so many, among them Naoki Aizawa, Enghin Atalay, Shane Auerbach, Alexander Clark, Dean Corbae, Chao Fu, Andrea Galeotti, Rasmus Lentz, George Loginov, Corina Mommaerts, Antonio Penta, Dan Quint, Marzena Rostek, Bill Sandholm, Chris Taber, Marek Weretka, and Matt Wiswall. I am grateful to the Juli Plant Grainger Institute for financial support and to the Institute for Research on Poverty for fellowship and advice. To my advisor Steven Durlauf I am grateful for grand navigation, encouragement, inspiration, solace, and exemplary professorial altruism. To my family I owe more than I can say.

This research uses data from Add Health, a program project directed by Kathleen Mullan Harris and designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris at the University of North Carolina at Chapel Hill, and funded by grant P01-HD31921 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development, with cooperative funding from 23 other federal agencies and foundations. Special acknowledgment is due Ronald R. Rindfuss and Barbara Entwisle for assistance in the original design. Information on how to obtain the Add Health data files is available on the Add Health website (http://www.cpc.unc.edu/addhealth). No direct support was received from grant P01-HD31921 for this analysis.

The NLSY79 survey is sponsored and directed by the U.S. Bureau of Labor Statistics and conducted by the Center for Human Resource Research at The Ohio State University. Interviews are conducted by the National Opinion Research Center at the University of Chicago.

List of Figures

Contents

Chapter 1

Social Capital and Racial Inequality

1.1 Introduction

Social connections play an important role in finding work.[1] But as not everyone has the same friends, some people may end up in higher-paying jobs than others, even if their qualifications are otherwise identical. This paper's contribution is to empirically assess the importance of inequality in social capital to aggregate wage inequality between groups. In particular, I focus on the oft-studied[2] persistent inequality in labor market outcomes between black workers and others in the US. I estimate that 6–10% (depending on the model) of the wage gap between these groups can be explained by differences in jobs found through friends, even after controlling for individual human capital.

The model is simple and similar to that of Burdett (1978), but with heterogeneous workers, wage growth with experience, and two types of offers. Workers receive job offers through direct search and also through friends. The chance of receiving an offer depends on its source, as well as worker characteristics including current employment status. A worker will accept an offer if it maximizes expected discounted utility from earnings. At the end of each period workers may lose their jobs; the chance of this happening depends on worker

[1] Granovetter (1973) found that roughly half of jobs are found through a social connection; more recent work on the importance of referrals includes Ioannides and Datcher Loury (2004) and Schmutte (2016).

[2] See Altonji and Blank (1999) for a review.

characteristics as well as how the job was found.

The parameters are estimated jointly via method of simulated moments on panel data from the NLSY79, which in certain years asks respondents whether their jobs were found through friends. This is the key to the identification strategy: the wages of jobs found directly provide a measure of individuals' human capital, allowing the relative wages of jobs found through friends to be used as a measure of social capital. I find that the job offers black workers find through friends pay less than those of equally productive non-black workers.

Montgomery (1991; 1992) showed how persistent inequality in wages and educational attainment can arise between two groups even if they have equal productive potential. More recently, Calvó-Armengol and Jackson (2004) explored a similar idea in an explicit network setting. The key to these papers is that if two groups are more likely to form in-group social connections (homophily), their labor market outcomes can follow different trajectories. Workers choose higher education or labor force participation because their friends are doing the same, and the payoffs to education or participation are higher if you have friends who can help you find a job (strategic complementarity). In a similar vein, Arrow and Borzekowski (2004) argue that plausible differences in network degree (number of connections) may account for black-white income disparity. I take these theoretical foundations to the data, and find that differences in jobs found through friends are indeed an important part of aggregate racial inequality in labor market outcomes.

This paper remains agnostic on exactly why friends are helpful in job search. The results are consistent with a model like that of Calvó-Armengol and Jackson (2004), in which friends are useful for merely alerting you that an opening exists, but also consistent with Simon and Warner (1992), in which referrals are valuable because they provide information to the hiring firm.[3] As my data cannot distinguish between these models, firm behavior is not modeled: arrival rates and wage distributions are reduced-form specifications flexible enough to qualitatively match a variety of possible explanations for why friends matter on the job market. The focus is instead on measuring how much the fruits of search through friends differ by race.

My results are broadly consistent with previous empirical work on the topic. Schmutte (2015) uses geographic variation to show that social interactions can explain why some workers get higher paying jobs. Holzer (1987) finds that differences in job finding rates can account for most of the racial difference in unemployment outcomes for youths, and Green et al. (1999) find that jobs found through friends pay less for black workers in certain cities. This paper makes two main contributions relative to previous empirical work. The first is scope—I use nationally representative data and consider both unemployment and wages of early- and mid-career workers. Second, I allow the wage premium of jobs found through friends to vary with human capital, to avoid misinterpreting group differences in human capital as differences in social capital (see Section 1.4.2 for further discussion).

These results are important because they have fundamentally different policy prescriptions than much of the previous literature. Prior work on racial wage inequality has often focused on differences in human capital, such as formal education or hidden investment in skill that leads to statistical discrimination (see, for example, Neal and Johnson (1996) or Coate and Loury (1993)[4]). This paper is qualitatively different in that it demonstrates inequality in labor market outcomes controlling for individual human capital. A policy such as affirmative action

[3]Brown et al. (2015) find that panel data from a large US corporation support a model such as that of in which referrals are valuable because they convey information about workers' productivity. For example, jobs found through referrals tend to pay higher initial wages–a finding corroborated by this paper.

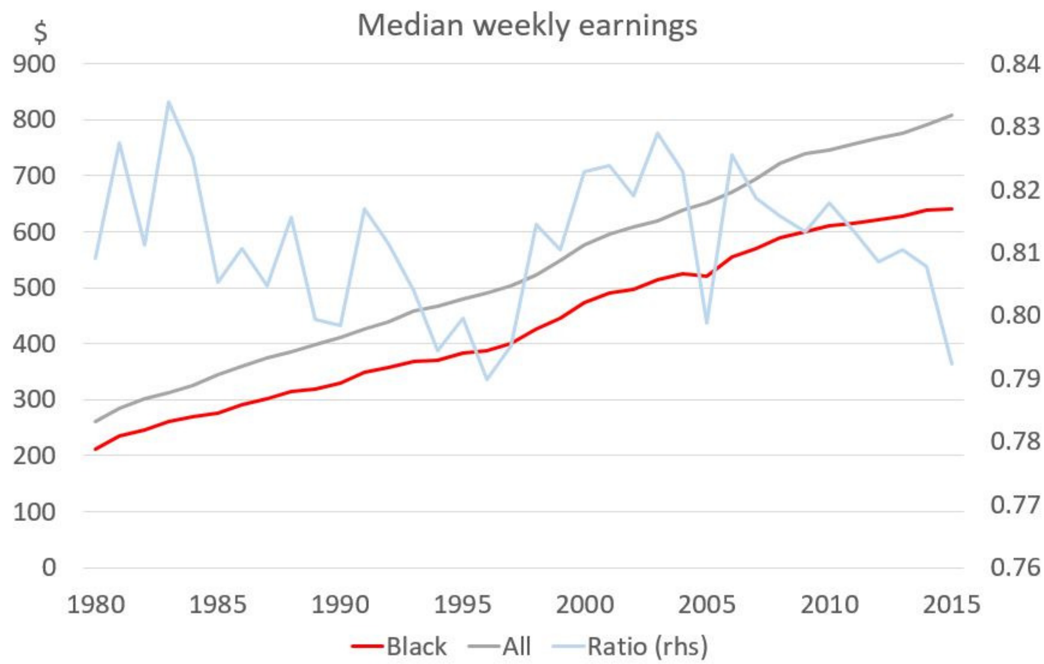
[4]Fang and Moro (2011) review other work in this vein.

in hiring may not effectively address skills disparities, but may be able to balance out social capital inequality and provide disadvantaged groups with the same rewards to education enjoyed by groups with better social connections. Furthermore, it may increase the social capital of the disadvantaged group: the more minority workers have high-status jobs, the better the job-finding prospects of their friends. This can explain empirical evidence from Miller (2017) that firms which increase their hiring of black workers to meet affirmative action standards required by government contracts tend to continue to do so even after they are no longer compelled to do so. Finally, while racial inequality is the focus of this paper, the importance of social capital may prove important for understanding inequality in other dimensions as well.

Other work in this area provides complementary explanations for persistent inequality. Becker and Tomes (1979) and Loury (1981) show that inequality can persist if lower-income parents invest less in their children's education. Durlauf (1996) shows that gaps between neighborhoods can persist where education is a local public good. This paper establishes the importance of network job-finding alongside these other mechanisms as a key driver of racial inequality. Whereas most previous work focuses on inequality driven by differences in human capital, this paper isolates differences in job-finding which owe entirely to differences in social connections.

Section 1.2 motivates the research, reviewing stylized facts about race, labor market outcomes, and social connections. Section 1.3 develops the labor market model. Section 1.4 describes the data used and performs preliminary analysis that roughly encapsulates the main result in a single fixed-effects regression, for those short on time. Section 1.5 explains the estimation and identification strategy, and Section 1.6 presents the main results.

Figure 1: Racial inequality in earnings



1.2 Stylized facts

There are two stylized facts essential to this paper. First, labor market outcomes differ by race. Figures 1 and 2 show (using BLS data) that black workers earn less and are more likely to be looking for work. Racial gaps in earnings and unemployment also exist across education levels. And across education levels, employed black workers report fewer job offers, as shown by Table 1.[5]

Second, the socioeconomic advantage of social connections is also unequally distributed by race. This is in part a mechanical consequence of racial homophily—the tendency of people to have friends of the same race. Figure 3 shows that the friends of black workers are more likely to be looking for work, especially at lower levels of educational attainment. The

[5]This table uses data from the NLSY79, described in Section 1.4.1. Wolpin (1992) finds that young black workers who did not go to college receive more job offers than average; these figures provide a broader view.

Figure 2: Racial inequality in unemployment

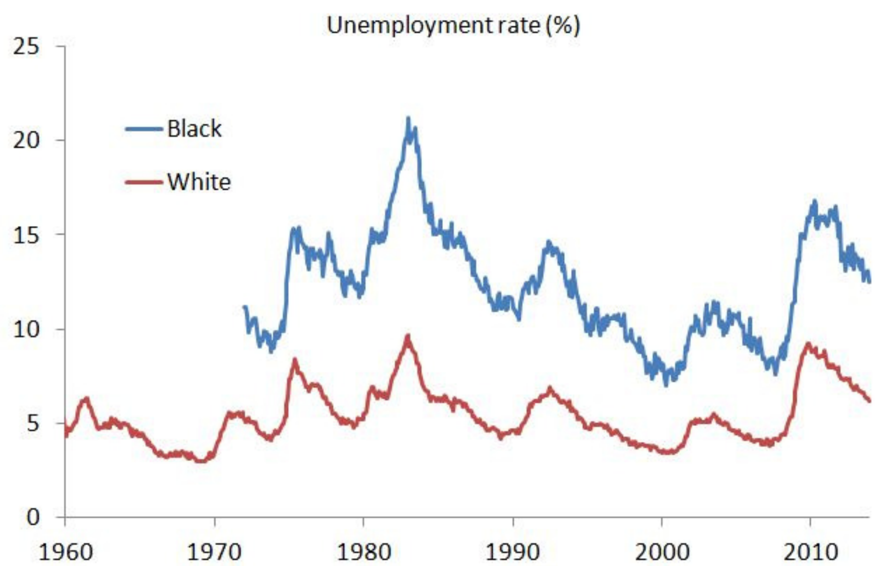


Table 1: Job offers by education and race

other offers when found job

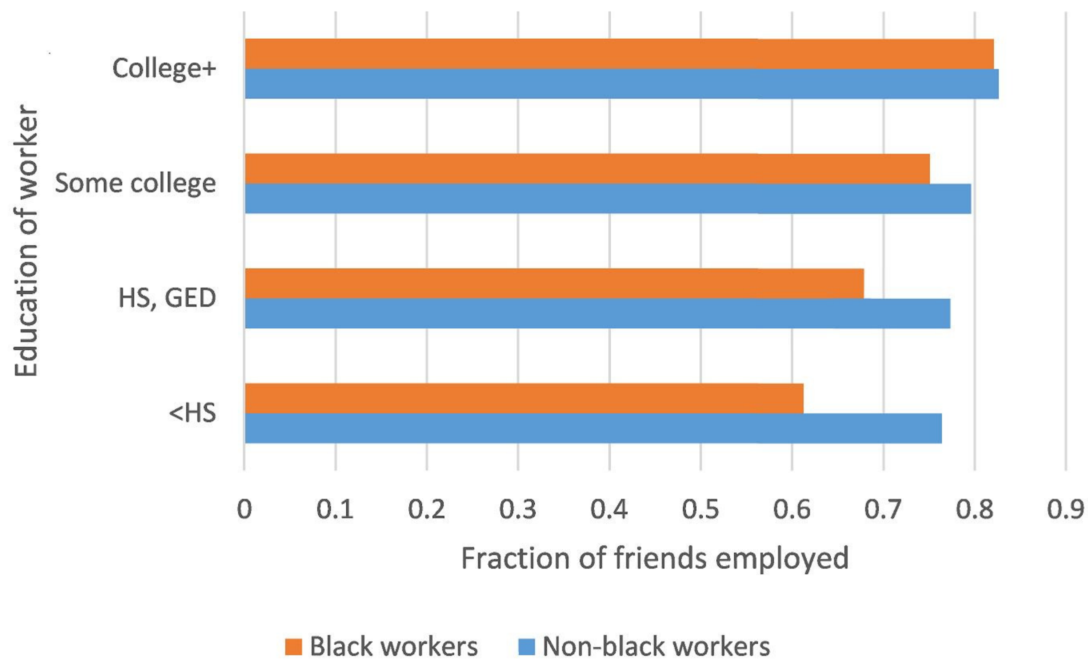
Years of school	Non-black	Black
<12	0.31	0.23
12	0.37	0.29
13-15	0.51	0.40
16+	0.74	0.70

friends of black workers at all levels of educational attainment also tend to have less formal education. For example, in Wave III of the Add Health dataset[6], 21% of the friends of black respondents with exactly 12 years of education did not finish high school, compared with 14% for the friends of non-black respondents.

Taken together, these facts suggest that non-black workers may have an advantage in finding high-paying work through friends. To fix terminology, I will use the term "social

[6]Add Health is a longitudinal study of sample of adolescents who were nationally representative of those aged 7-12 when the study began in 1994.

Figure 3: Employment rate of friends



capital" to refer to any such factors which affect jobs found through friends but not those found directly.

1.3 Model

Consider a group I of n workers who have completed their schooling and seek to maximize expected discounted utility from earnings. Time progresses in discrete periods $t = 1, 2, 3, ...$, each of which is 2 weeks long.[7] Each period, a worker may receive up to one job offer. Job offers come in two flavors: those found directly and those found through friends. The chance of receiving a direct offer is λ^d_{it} and the chance of receiving an offer through friends is λ^f_{it} (parameterized below in Equations 1.5 and 1.6). A worker who receives an offer in period t

[7]The length of each period was chosen to be as short as possible while still computationally feasible.

can either accept it or reject it and maintain the status quo—his current job (if employed) or unemployment (if not). This is the only choice made by workers in the model. At the end of the period, employed workers may lose their job with chance δ_{it}.

The log wage that worker i receives at time t from job j if the job was found directly is as follows:

$$w^d_{ij(t)} = \eta_i + h_{it} + \epsilon^d_{ij(t)}. \tag{1.1}$$

The log wage of a job found through friends is:

$$w^f_{ij(t)} = \beta_{\text{hcap}}\eta_i + \beta_{\text{exp}}h_{it} + \beta_0 + \beta_s s_{r(i)} + \epsilon^f_{ij(t)}. \tag{1.2}$$

Each worker is endowed with initial marketable human capital η_i, which can be thought of as summarizing characteristics relevant to productivity such as education and ability. While η_i will be referred to as "human capital" throughout the paper, it is best thought of as i's earning power. For example, η_i will not be i's true productive capacity if discrimination means that i is paid less than her true productive capacity. In this case η_i can be thought of how much i can expect to be paid under the current discriminatory regime. This is discussed further in Section 1.5.3 and Appendix A.3.

Initial human capital is distributed normally with parameters that may differ by race ($r_i \in \{0, 1\}$ denotes whether i reports being black), reflecting differences in early childhood circumstances,[8] educational attainment, etc.:

$$\eta_i \sim N\left(\mu_{r(i)}, \sigma^2_{r(i)}\right). \tag{1.3}$$

Worker i's wage at job j depends on acquired human capital h_{it}, which in turn grows mechanically with work experience, x_{it}:

$$h_{it} = \max\left(\alpha_1 x_{it} + \alpha_2 x^2_{it}, h_{it-1}, h_{t-2}, ...\right). \tag{1.4}$$

[8] For example, Aizer and Currie (2017) find that black children face higher exposure to lead, effecting higher rates of school suspension and juvenile detention.

This acquired human capital is parameterized as a quadratic, with the stipulation that its growth can fall to zero but cannot turn negative (this technical assumption is made to avoid the unnatural scenario of a worker preferring unemployment to avoid losing human capital). The law of motion for work experience x_{it} is simply that it increases by one every period that the worker is employed.

The shock ϵ_{ij}^{d} represents how productive worker i is at firm j, and is drawn from a normal distribution with mean normalized to zero and standard deviation σ^{d}. For jobs found through friends, the analogous match quality between worker and firm is $\beta_0 + \beta_s s_{r(i)} + \epsilon_{ij(t)}^{f}$, where $\epsilon_{ij(t)}^{f}$ is normally distributed with mean zero and standard deviation σ^{f}. This allows for the possibility that jobs found through friends provide better or worse productivity matches (β_0), and that this difference depends on race ($\beta_s s_{r(i)}$).

Each worker is endowed with social capital $s_{r(i)}$, which represents the average difference in social capital between black and non-black workers. Having no natural units, it is equal to one for black workers and zero otherwise. This is of course a simplification—the implications of heterogeneity in social capital within racial groups are discussed in Section 1.5.3 and Appendix A.3. Section 1.2 highlighted several group-level differences in social connections that we might expect to influence job search; the goal of this paper is to determine the degree to which such differences translate into aggregate differences in labor market outcomes. In the model, social capital affects both the arrival of new offers found through friends and the wages they promise, but not offers found directly.

A worker's wage can grow over the life cycle by remaining in the same job and gaining experience, or by moving to a job with a better productivity match quality ϵ_{ij}. But a worker's match quality at a given job j does not change over time.

When workers are unemployed, they receive unemployment insurance equal to 40% of the most recent wage up to a maximum of \$130/week (in 1979 dollars), roughly corresponding to typical US unemployment insurance laws. If a worker has never worked, as is the case when entering the labor market for the first time, the worker's initial human capital η_i is used in place of the most recent wage.

The chance of receiving an offer directly (λ_{it}^d) and the chance of receiving an offer through friends (λ_{it}^f) are parameterized as follows. First, define the following intermediate variables:

$$\tilde{\lambda}_{it}^d = \left(\Lambda_0^d + \Lambda_{\text{race}}^d r_i + \Lambda_{\text{hcap}}^d \eta_i + \Lambda_{\text{exp}}^d h_{it} + \Lambda_u^d u_{it}\right) \tag{1.5}$$

$$\tilde{\lambda}_{it}^f = \left(\Lambda_0^f + \Lambda_{\text{race}}^d r_i + \Lambda_{\text{hcap}}^f \eta_i + \Lambda_{\text{exp}}^f h_{it} + \Lambda_s^f s_{r(i)} + \Lambda_u^f u_{it}\right). \tag{1.6}$$

Since these specifications are linear, $\tilde{\lambda}_{it}^d$ and $\tilde{\lambda}_{it}^f$ may not be bounded to the unit interval. I therefore use the logistic function $\mathcal{L}(x) = \frac{1}{1+e^{-x}}$ to parameterize the probabilities of receiving offers directly and through friends as

$$\lambda_{it}^d = \frac{\mathcal{L}\left(\tilde{\lambda}_{it}^d\right)}{\mathcal{L}\left(\tilde{\lambda}_{it}^d\right) + \mathcal{L}\left(\tilde{\lambda}_{it}^f\right)} \mathcal{L}\left(\tilde{\lambda}_{it}^d + \tilde{\lambda}_{it}^f\right) \tag{1.7}$$

$$\lambda_{it}^f = \frac{\mathcal{L}\left(\tilde{\lambda}_{it}^f\right)}{\mathcal{L}\left(\tilde{\lambda}_{it}^d\right) + \mathcal{L}\left(\tilde{\lambda}_{it}^f\right)} \mathcal{L}\left(\tilde{\lambda}_{it}^d + \tilde{\lambda}_{it}^f\right). \tag{1.8}$$

This ensures that each probability lies within the unit interval, and that their sum does as well (recall that only one offer may arrive each period, so these are the probabilities of disjoint events).

The arrival of each type of offer thus depends on i's race r_i, initial human capital η_i, accumulated human capital h_{it}, an indicator u_{it} of whether i is unemployed in period t, and a constant. The arrival rate of jobs found through friends depends on all of these and also on social capital $s_{r(i)}$. As the arrival of offers through friends (Equation 1.6) also includes the $\Lambda^d_{\text{race}} r_i$ term that determines the arrival of direct offers (Equation 1.5), the social capital coefficient Λ^f_s will capture differences in arrival of offers through friends not explained by differences in the direct offer arrival rate—this is analogous to the $\beta_s s_{r(i)}$ term in the wage equation 1.2.

Finally, at the end of each period an employed worker's job is destroyed with chance δ_{it}:

$$\delta_{it} = \mathcal{L}\left(\Delta_0 + \Delta_{\text{race}} r_i + \Delta_{\text{hcap}} \eta_i + \Delta_{\text{exp}} h_{it} + \Delta_f f_{ij(t)}\right). \tag{1.9}$$

The logistic function $\mathcal{L}$ is again used to bound this probability inside the unit interval. The chance of losing one's job depends on $f_{ij(t)}$, an indicator of whether i's job j at time t was found through friends. Jobs found through friends may be a better match for a worker's skills, or it may be more enjoyable to work with friends. More cynically, it may be harder to fire someone who is a friend or relative of the boss.

It is theoretically possible in this model that an employed worker might want to quit to unemployment. Since the chance of finding a new job changes with age, the relative value of unemployment may have increased since the worker accepted her current job offer. However, at the estimated parameter values this is never the case, and thus this possibilty is ignored: employed workers never choose to quit to unemployment.

What offers will a worker accept? Worker i's value of employment at job j with match quality ϵ_{ij} and experience x_{it} at time t can be written as follows. Note that the value also depends on race r_i, social capital s_i, and initial human capital η_i, but as they are time-invariant these arguments are suppressed and replaced with a dot $(\cdot)$.

$$V\left(\epsilon_{ij(t)}, f_{ij(t)}, a_{it}, x_{it}\cdot\right) = w_{ij(t)} + \rho \Bigg\{ \delta_{it} U\left(w_{ij(t)}, a_{it+1}, x_{it}, \cdot\right) + (1-\delta_{it}) \Bigg[V\left(\epsilon_{ij(t)}, f_{ij(t)}, a_{it+1}, x_{it}+1, \cdot\right)$$
$$\left(\left(1 - \lambda_{it}^{d} - \lambda_{it}^{f}\right) + \lambda_{it}^{d} \Phi^{d}\left(\underline{\epsilon}^{ed}\left(\epsilon_{ij(t)}, f_{ij(t)}, a_{it}, x_{it}\cdot\right)\right) + \lambda_{it}^{f} \Phi^{f}\left(\underline{\epsilon}^{ef}\left(\epsilon_{ij(t)}, f_{ij(t)}, a_{it}, x_{it}, \cdot\right)\right)\right)$$
$$+ \lambda_{it}^{d} \int_{\underline{\epsilon}^{ed}\left(\epsilon_{ij(t)}, f_{ij(t)}, a_{it}, x_{it}, \cdot\right)}^{\infty} V\left(\epsilon_{ij'}, 0, a_{it+1}, x_{it}+1, \cdot\right) d\phi^{d}\left(\epsilon_{ij'}\right)$$
$$+ \lambda_{it}^{f} \int_{\underline{\epsilon}^{ef}\left(\epsilon_{ij(t)}, f_{ij(t)}, a_{it}, x_{it}\cdot\right)}^{\infty} V\left(\epsilon_{ij'}, 1, a_{it+1}, x_{it}+1, \cdot\right) d\phi^{f}\left(\epsilon_{ij'}\right) \Bigg] \Bigg\}$$

First, the worker receives the wage flow, $w_{ij(t)}$. With chance δ_{it} (given by equation 1.9), worker i's job is lost and starts the next period (discounted by ρ) with a value of unemployment U that depends on the wage of the job lost as well as the worker's new age a_{it+1}; in this case the worker's experience remains at x_{it}. With chance $1 - \delta_{it}$, the job is not lost. The second line contains the chance of remaining in the same job—this can happen either by receiving no offers (chance $\left(1 - \lambda_{it}^{d} - \lambda_{it}^{f}\right)$) or by receiving a job offer at a firm that is below the worker's reservation match quality for offers received in employment ($\underline{\epsilon}^{ef}$ for jobs found through friends and $\underline{\epsilon}^{ed}$ for jobs found directly). The reservation match quality depends on the current match quality, and worker's age, and experience. It also depends on on how the new and old offers were found since this can affect how likely the job is to be lost. So a worker may, for example, be willing to accept a slightly lower wage for a job found through friends if it is likely to last longer. Φ^{d} and Φ^{f} are normal cumulative densities with mean zero and standard deviations σ^{d} and σ^{f}, reflecting the distribution of match quality.

The third line contains the chance and value that a direct job offer is received and the match quality $\epsilon_{ij'}$ of the new job j' is above the reservation match quality. The fourth line contains the chance and value that an offer is received through friends and it is of acceptable match quality. Here ϕ^d and ϕ^d are normal densities of mean zero and standard deviations σ^d and σ^f, again reflecting the distribution of match quality.

The value of unemployment depends on the most recent wage $w_{ij(t)}$ at time t as well as age and experience (when a worker is unemployed, define for convenience $w_{ij(t)}$ as their most recent wage).

$$U\left(w_{ij(t)}, a_{it}, x_{it}, \cdot\right) = \max\left\{w_{ij(t)} + \log\left(0.4\right), \log\left(\$130 \times 2\right)\right\} + \rho\Bigg[$$

$$\left(\left(1 - \lambda_{it}^d - \lambda_{it}^f\right) + \lambda_{it}^d \Phi^d\left(\underline{\epsilon}^{ud}\left(w_{ij(t)}, a_{it}, \cdot\right)\right) + \lambda_{it}^f \Phi^f\left(\underline{\epsilon}^{uf}\left(w_{ij(t)}, a_{it}, \cdot\right)\right)\right) U\left(w_{ij(t)}, a_{it+1}, x_{it}, \cdot\right)$$

$$+ \lambda_{it}^d \int_{\underline{\epsilon}^{ud}\left(\epsilon_{ij(t)}, f_{ij(t)}, a_{it}, \cdot\right)}^{\infty} V\left(\epsilon_{ij'}, 0, a_{it+1}, x_{it}, \cdot\right) d\phi^d\left(\epsilon_{ij'}\right)$$

$$+ \lambda_{it}^f \int_{\underline{\epsilon}^{uf}\left(w_{ij(t)}, f_{ij(t)}, a_{it}, \cdot\right)}^{\infty} V\left(\epsilon_{ij'}, 1, a_{it+1}, x_{it}, \cdot\right) d\phi^f\left(\epsilon_{ij'}\right)\Bigg]$$

The value of unemployment has four parts. First is unemployment insurance, which depends on the most recent wage and is described above. The second line contains the chance and value of remaining unemployed in the following period, which occurs if either no offer arrives or if one does but it is below the reservation match quality. Note that since the flow value of unemployment insurance depends on a worker's most recent wage but is invariant to how that job was found, the reservation match qualities in unemployment $\underline{\epsilon}^{uf}$ and $\underline{\epsilon}^{ud}$ likewise depend on the most recent wage but not its provenance. They still depend on the provenance of the new offer, however, as again this affects the chance of the job being lost. And as in Burdett (1978), a worker may prefer to remain in unemployment than take a low offer, given that the search rate is higher in unemployment. The third line contains the chance and value of an acceptable offer through direct search, and the fourth line contains the chance and value of an acceptable offer arriving through friends. In all cases the worker's experience x_{it} does not grow, as they are unemployed this period.

In this model there are three distinct ways in which friends can help you on the labor market:

1. More offers
2. Higher-paying offers
3. Jobs less likely to be lost[9]

The focus of this paper is whether or not social capital differs by race in a way that affects jobs found through friends through these three channels.

1.4 Data and preliminary analysis

1.4.1 Data

The model is estimated using the NLSY79, a longitudinal study of a sample of 12,686 people in the US who were nationally representative when first surveyed in 1979, at ages 14–22. In years 1982, 1994, 1996, 1998, and 2000, respondents who reported working were asked whether or not they found their job through asking friends or relatives. Specifically, those who reported looking for work when offered their current job are asked "Which of the methods on this card led to your being offered your job with [Name of employer]?" I identify this as a job found through friends/relatives if they marked "Contacted friends or relatives."[10]

For these jobs (up to five for each year), respondents also reported their wages, occupation, and whether they were working when they found the job. Another important feature of the data is a cumulative work experience over the entire sample (not just the years in which

[9] Loury (2006) and Brown et al. (2015) have found that jobs found through social connections last longer.

[10] Other possible search methods include checking with a state employment agency, checking with a private employment agency, contacting an employer directly, placing or answering an ad, and looking in a newspaper; respondents are instructed to mark all that apply.

job provenance is asked about). All dollar values are deflated using the CPI-U-RS published by the Bureau of Labor Statistics. Educational attainment is also reported, and the sample is restricted to include only those observations for which respondents had reached their highest level of education. The data also include sex, race, age, and employment and marital status. This paper focuses on differences in outcomes between black workers and others.

1.4.2 Preliminary analysis

Before estimating the full model, the main result can be roughly encapsulated in a single fixed-effects regression. Log weekly earnings are regressed on whether the job was found through friends, the interaction of this variable with race, and a variety of other controls including individual fixed effects. Selected coefficients are reported in Table 2.

Table 2: Wage premium of jobs found through friends

Fixed effects regression of log weekly earnings

Variable	Coefficient	Std. Err.
Job found through friends	0.194	0.015
Black*(job found through friends)	-0.080	0.035
Experience	0.139	0.011
Experience^2	-0.003	3.50E-4
Married	-5.10 E-4	0.025
Married*male	0.132	0.034
# obs.	24,433	
# individuals	8,510	

Note: Unreported coefficients include individual fixed effects, an indicator for urban location, year dummies, and a constant.

Jobs found through friends tend to pay more, but the premium is lower for black workers. This is the main result of the paper—inequality in social capital seems to be driving part of the racial wage gap, even after controlling for individual human capital.

Could this result still owe to differences in human capital, rather than social capital? Consider the following model.

$$w^d_{ij(t)} = \eta_i + \beta^d_X X_{it} + \epsilon^d_{ij(t)} \tag{1.10}$$

$$w^f_{ij(t)} = \beta_{\text{hcap}}\eta_i + \beta^f_X X_{it} + \beta_s s_i + \epsilon^f_{ij(t)}. \tag{1.11}$$

If the wages of jobs found through friends depend differently on human capital ($\beta_{\text{hcap}} \neq 1$), then a group difference in average human capital could show up in Table 2 even if $\beta_s = 0$. For example, if jobs found through friends yield higher wages only for workers with lots of human capital ($\beta_{\text{hcap}} > 1$), then a group with less human capital will have a lower average premium for jobs found through friends.

To address this concern, I first run a fixed-effects regression of the log wages of jobs found directly (Equation 1.10) to obtain estimates of individual human capital $\hat{\eta}_i$. These estimates are then used to estimate Equation 1.11. The results are given in the first column of Table 3. Note that the estimated coefficient for black workers is essentially the same as in Table 2.

Table 3: Wage premium of jobs found through friends

Regression of log weekly earnings of jobs found through friends

	(1) Naï¿œve		(2) Corrected	
Variable	Coefficient	Std. Err.	Coefficient	Std. Err.
Estimated fixed effect from Eq. 1.10	0.325	0.018	0.446	0.027
Black	-0.079	0.020	-0.071	0.020
Experience	0.114	0.010	0.122	0.010
Experience^2	-0.003	0.000	-0.003	0.000
Married	-0.093	0.026	-0.076	0.026
Married*male	0.379	0.029	0.338	0.030
# obs.			4,658	

Note: Unreported coefficients include an indicator for urban location, year dummies, and a constant.

The fixed effect estimates $\hat{\eta}_i$ are fairly noisy, as there are typically only a few wage observations per individual. This will bias lower the estimated coefficient on η, and may artificially inflate the magnitude of our coefficient of interest. Fortunately, we have a good sense how noisy $\hat{\eta}_i$ is, since we often observe multiple wages per individual at jobs found directly. Appendix A.2 details a procedure for correcting this measurement error, and the second column of Table 3 gives the corrected results: the coefficient of interest is modestly attenuated.

To determine the extent to which racial wage inequality is explained by social capital, I predict log weekly earnings for jobs found through friends using the corrected coefficient estimates from Table 3, but setting the race dummy equal to zero for all individuals. The predicted weekly earnings (converted to 2016 dollars for convenience) are given by Table 4.

Table 4: Weekly earnings predicted by linear regression (2016 $)

	Black	Non-black	Gap
Actual	363	419	56
Predicted	363	419	56
Counterfactual	368	419	50

This suggests that inequality in social capital can explain 10% of the racial gap in log weekly earnings.

Could it be the case that there is a racial difference in human capital that leads non-black workers to pursue careers in occupations in which social capital matters more? The network wage premium does vary by occupation (see Table 13 in Appendix A.1), but the estimated racial difference in network wage premium is actually slightly larger if controls for each worker's occupation are included. So the result does not seem to be driven by a difference in the importance of social capital in different occupations.

It may be the case that when a firm hires a worker through a referral, it is more certain that the worker has high productivity. If this were driving the results, then the racial difference in the network wage premium should shrink with tenure, a firms learn more about the productivity of their workers and the information gap between referrals and direct hires closes. However, the evidence does not support this: the estimated racial difference in network wage premium is slightly larger if the sample is restricted to only include jobs which workers have held for at least a year. So the main result holds even for those observations for which firms have had ample time to learn about workers' productivity.

How does this fit in with previous findings that most of the wage gap can be accounted for by premarket factors? O'Neill (1990) and Neal and Johnson (1996) find that in a regression of log wages on AFQT score and race, variation in the former is able to account for nearly all the variation in the outcome. This is interpreted as evidence that premarket factors such as human capital formation are responsible for the racial wage gap, rather than bias in hiring against black workers of equal productivity.[11] Table 14 in the Appendix shows the results of a similar exercise carried out separately for jobs found through friends and for jobs found directly. In jobs found directly, differences in AFQT can largely account (in a statistical sense) for racial inequality in earnings, as in previous work. But for jobs found through friends, the coefficient on race is large (and statistically significant), suggesting that there is racial variation in jobs found through friends that requires another explanation.

Given this result, estimating the model can fill in some important details. First, the racial discrepancy in the network wage premium could owe to better offers or simply more offers. Distinguishing these effects requires estimating the full model. Second, the exercise above is unable to address important issues of selection. In particular, changing the offers black workers find through friends could cause some workers to choose a different offer, or pull some workers out of unemployment—eventualities not addressed by the reduced-form counterfactual. Finally, the secondary effect of increased experience may also be important, and is captured in the model.

[11] To the extent that social connections are formed before entering the labor market they are a premarket factor, and variation in AFQT score may reflect both individual productivity as well as something about one's social connections, making it an imperfect way to separate an individual's characteristics from those of her social connections.

1.5 Estimation

1.5.1 Identification

A key feature of the data is that for each individual, multiple jobs are observed over time, some of which are found directly and some of which are found through friends. This allows individual human capital, which affects both types of offers, to be distinguished from the individual's social capital, which only affects the latter.

The 27 parameters are estimated jointly via method of simulated moments. Given a set of parameters, I first determine the reservation match quality in each state. To do so, I solve backwards assuming log utility, exogenous retirement at age 65, and a time discount factor ρ of 95% (per year). The continuous variables (initial human capital and match quality) as well as experience are each approximated by a 12-point grid. Finer grids were also tried, with no appreciable effect on results.

Next, I simulate each worker's wage path from the first quarter following the end of schooling through the year 2000, when the data end. When a worker's simulated human capital or the match quality of a simulated offer lie between grid points, linear interpolation is used to determine the reservation match quality. From these complete wage paths, I censor the simulated observations to include only those worker-period observations included in the NLSY79 data set.

The censored simulated data can then be compared to the actual data through the lens of certain judiciously chosen moments and regression coefficients, which are described below and summarized in Table 5. To avoid redundancy, not every coefficient from a given regression is targeted in estimation; these coefficients are marked in Table 5 with a dagger (†).

First, log wages are regressed on experience x_{it}, its square, an indicator f_{it} of whether the job was found through friends, and the interaction of this indicator with race r_i and experience and its square. This regression also includes individual dummy variables $\{d_i\}$; the estimated coefficients on these indicator variables (fixed effects) are denoted $\hat{\eta}_i$. The coefficients not including the fixed effects help identify the parameters $\{\alpha_1, \alpha_2\}$ from Equation 1.1, which determine the growth of human capital with experience and its square. They also help identify the three parameters $\{\beta_0, \beta_{\text{exp}}, \beta_s\}$ in equation 1.2, which determine the distribution of wages of jobs found through friends. The final parameter in this equation, β_{hcap}, is pinned down by the coefficient on the estimated fixed effect $\hat{\eta}_i$ in a separate regression of log wages on jobs found through friends that also includes a constant, race r_i, and experience x_{it}.

A regression of the estimated fixed effect $\hat{\eta}_i$ on a constant, race, experience, and its square identifies the means of the initial human capital distribution $\left\{\mu_{r(i)}\right\}$ from Equation 1.3 and also helps identify the dependence of human capital on experience. The standard deviations of the fixed effect for non-black and black workers identify the dispersion of initial human capital $\left\{\sigma_{r(i)}\right\}$, and the standard deviations of wages less the fixed effect $(w_{it} - \hat{\eta}_i)$ for jobs found directly and through friends identify the within-worker dispersion of offers $\{\sigma_f, \sigma_d\}$.

A regression of unemployment u_{it} on race, the estimated fixed effect $\hat{\eta}_i$, experience, and a constant targets the parameters of the job destruction rate $\{\Delta_{\text{race}}, \Delta_{\text{hcap}}, \Delta_{\text{exp}}, \Delta_0\}$ in Equation 1.9. A similar regression of log tenure at the current job on the same regressors targets the arrival rate parameters $\left\{\Lambda^d_{\text{race}}, \Lambda^d_{\text{hcap}}, \Lambda^d_{\text{exp}}, \Lambda^d_0\right\}$ and $\left\{\Lambda^f_{\text{race}}, \Lambda^f_{\text{hcap}}, \Lambda^f_{\text{exp}}, \Lambda^f_0, \Lambda^f_s\right\}$ in Equations 1.5 and 1.6 as well as the dependence of job loss on provenance Δ_f, and a regression of f_{it} (an indicator of whether the current job was found through friends) on these regressors distinguishes the arrival rate parameters for direct offers from those for offers through friends.

An indicator that i was employed when the current job was found is averaged for

Table 5: Targeted statistics

		Included independent variables									
	Dependent variable	1	r_i	x_{it}	${x_{it}}^2$	f_{it}	$f_{it} \cdot r_i$	$f_{it} \cdot x_{it}$	$f_{it} \cdot x_{it}^2$	$\{d_i\}$	$\hat{\eta}_i$
(1)	Log wage w_{it}			✓	✓	✓	✓	✓	✓	✓†	
(2)	w_{it}, job through friends	✓†	✓†	✓†							✓
(3)	Estimated fixed effect $\hat{\eta}_i$	✓	✓	✓†	✓†						
(4)	Std. dev. of $\hat{\eta}_i$	✓	✓								
(5)	Std. dev. of $(w_{it} - \hat{\eta}_i)$	✓				✓					
(6)	Unemployment u_{it}	✓	✓	✓							✓
(7)	Log tenure at current job	✓	✓	✓		✓					✓
(8)	Job found through friends f_{it}	✓	✓	✓							✓
(9)	Employed when found job	✓				✓					

†Coefficients not targeted in estimation.

jobs found directly and those through friends to distinguish on-the-job search from search in unemployment for both sources of offers, parameterized by Λ^d_{emp} and Λ^f_{emp}.

The set of targeted statistics also includes averages by race of log wages, unemployment, and the fraction of jobs found through friends. While these are featured in regressions above, the regressions include other variables endogenous to the model, and it is important that the simulation match these crucial aggregate statistics as well.

1.5.2 Procedure

Practically, estimation proceeds as follows. First, a vector of 34 targeted statistics including those indicated by Table 5 is calculated using the actual data. Then, starting values of the parameters are chosen from preliminary regressions. The model is simulated, and a vector of targeted statistics is calculated from the simulated outcomes. The difference between the the vector of true statistics and the vector of simulated statistics is weighted by W , the inverse covariance matrix of the target statistics (computed separately, from 500 draws of the original data).

$$\rho\left(v^{\text{true}}, v^{\text{sim}}\right) = \left(v^{\text{true}} - v^{\text{sim}}\right)' W \left(v^{\text{true}} - v^{\text{sim}}\right) \tag{1.12}$$

An algorithm[12] searches for the set of parameters which, when simulated, yield a vector of statistics v^{sim} that minimizes this distance $\rho\left(v^{\text{true}}, v^{\text{sim}}\right)$.

Finally, confidence intervals are constructed by bootstrapping. The set of workers is resampled with replacement to generate a synthetic sample, and the parameters are reestimated using the synthetic sample. The baseline and counterfactual simulations are run using the new parameters and synthetic sample. This is repeated 200 times to generate confidence intervals.

1.5.3 Caveats

This section concludes with some important caveats. There are reasons to suspect this paper's estimate of the role of social capital may be an underestimate. First, if there is racal bias in direct hiring, η_i will not represent i's productivity.[13] This alone does not affect the main results, but if there is more racial bias or discrimination against black workers in direct hiring than in hiring through friends, then the wages of black workers will understate their human capital. In this case, the racial difference in social capital will be larger than estimated.

Second, if there is within-race heterogeneity in social capital that is correlated with human capital, then wages of jobs found through friends may appear to depend on human

[12] Specifically, MATLAB's fminsearch is used, which employs the Nelder-Mead simplex algorithm. This was supplemented with the Artelys Knitro 10.3 solver.

[13] Bertrand and Mullainathan (2004) find that employers are less likely to respond to applications with names indicating a black applicant, though Neal and Johnson (1996) argue that most of the wage gap between black workers and others is accounted for by differences in skill as measured by the AFQT.

capital more than they do. This would lead to an underestimate of the effect of increasing social capital.[14] For these reasons, this paper's results should be seen as conservative. Appendix A.3 examines both cases in further detail.

1.6 Results

1.6.1 Parameter estimates

Table 6 presents the estimates of the parameters related to the distribution of human capital and its growth with work experience. The estimated group difference in mean initial human capital reflects inequality in educational attainment as well as other factors mentioned in Section 3.2.

Table 6: Human capital parameter estimates

		Estimate	95% CI
Mean initial human capital, non-black	$\mu_{\text{non-black}}$	5.106	[5.02, 5.21]
Mean initial human capital, black	μ_{black}	4.906	[4.82, 5.01]
Std. dev. initial human capital, non-black	$\sigma_{\text{non-black}}$	0.639	[0.606, 0.688]
Std. dev. initial human capital, black	σ_{black}	0.422	[0.369, 0.522]
Dependence of direct wages on experience	α_1	0.1923	[0.180, 0.207]
Dependence of direct wages on experience squared	α_2	-1.37 E-3	[-0.087, -0.0125]

The other parameters that determine the wage distribution of offers are given in Table 7. Recall that the dependence of offers found directly on human capital is normalized to one (Equation 1.1), so there are no analogues of β_{hcap}, β_{exp}, or β_0 for jobs found directly. The wages of jobs found through friends are higher but have slightly lower variance, and depend less on

[14] If social capital were negatively correlated with human capital, this paper would overestimate the importance of social capital. However, all available evidence contradicts this possibility: self-reported sociability is positively correlated with AFQT score and wages, and the network wage premium grows with education.

human capital, perhaps supporting the argument of Loury (2006) that jobs found through friends as a last resort can be important. The racial difference in social capital does seem to affect wage offers, even after controlling for the fact that the wages of jobs found through friends may depend on human capital differently than those of jobs found directly.

Table 7: Wage offer distribution parameter estimates

		Estimate	95% CI
Wage premium, jobs found through friends	β_0	0.459	[0.452, 0.469]
Dependence on initial human capital η_i, jobs found through friends	β_{hcap}	0.954	[0.940, 0.970]
Dep. on acquired human capital h_{it}, jobs found through friends	β_{exp}	0.977	[0.944, 1.035]
Dep. on racial difference in social capital $s_{r(i)}$, jobs found through friends	β_s	-0.039	[-0.084, -0.037]
Standard deviation of shocks, direct offers	σ_d	0.477	[0.444, 0.550]
Standard deviation of shocks, offers found through friends	σ_f	0.460	[0.357, 0.474]

Table 8 presents the offer receipt parameter estimates. At a given level of human capital, black workers receive fewer offers directly but about the same through friends as non-black workers. This is perhaps unexpected given Table 1, but consistent with Wolpin (1992).

Table 8: Offer receipt parameter estimates

		Estimate	95% CI
Constant, offers received directly in employment	Λ_0^d	-0.850	[-0.863, -0.840]
Constant, offers through friends in employment	Λ_0^f	-2.879	[-2.927, -2.832]
Difference for black workers	Λ_{race}^d	-1.19 E-2	[-1.21 E-2, -1.18 E-2]
Dependence of arrival rate on racial difference in social capital	Λ_s	1.04 E-2	[1.05 E-2, 1.08 E-2]
Dependence on initial human capital η_i, jobs found directly	Λ_{hcap}^d	3.44 E-2	[3.38 E-2, 3.49 E-2]
Dependence on initial human capital η_i, jobs found through friends	Λ_{hcap}^f	1.01 E-2	[1.00 E-2, 1.03 E-2]
Dependence on acquired human capital h_{it}, jobs found directly	Λ_{exp}^d	1.90 E-2	[1.86 E-2, 1.92 E-2]
Dependence on acquired human capital h_{it}, jobs found through friends	Λ_{exp}^f	0.55 E-2	[0.55 E-2, 0.56 E-2]
Diff. when unemployed, jobs found directly	Λ_{emp}^d	0.585	[0.577, 0.592]
Difference when unemployed, jobs found through friends	Λ_{emp}^f	0.843	[0.828, 0.857]

Table 9 gives the estimates of the job destruction parameters. Human capital lowers the rate of job destruction, and black workers are more likely to lose their jobs. Jobs found through friends are slightly less likely to be lost, as found in previous papers, though here the difference is not large.

Table 9: Job destruction parameter estimates

		Estimate	95% CI
Constant	Δ_0	-2.06	[-2.09, -2.02]
Difference for black workers	Δ_{race}	0.244	[0.240, 0.247]
Dependence on initial human capital η_i	Δ_{hcap}	-0.734	[-0.748, -0.723]
Dependence on acquired human capital h_{it}	Δ_{exp}	-3.47 E-3	[-3.53 E-3, -3.40 E-3]
Dependence on how job was found f_{it}	Δ_{f}	-6.64 E-2	[-6.75 E-2, -6.54 E-2]

1.6.2 Simulated statistics

With the parameters estimated, it is possible to simulate the model in both baseline and counterfactual scenarios. Table 10 presents the values of some key moments of interest for both simulated and actual data, as well as a counterfactual in which there is no racial difference in social capital (i.e., s_i is set to zero for all workers). Estimates of the full list of 34 targeted moments and regression coefficients from Table 5 are in Table 15 in Appendix A.1. The model does quite well at matching unemployment and the frequency with which accepted job offers are found through friends, while simulated earnings are higher than they are in the data (though the perceived difference is perhaps exaggerated by reporting these figures in dollars rather than log dollars, which is how they are treated in the model and estimation).

Table 10: Key moments

		Data	Simulated	Counterfactual	Difference
% jobs found through friends	Non-black	23.7	22.1	-	-
			[20.5, 26.4]	-	-
	Black	26.2	22.8	23.0	0.18
			[19.0, 25.5]	[19.2, 25.3]	[-0.98, 0.58]
Weekly earnings of employed (2016 $)	Non-black	442	553	-	-
			[448, 597]	-	-
	Black	381	470	474	4.44
			[365, 494]	[370, 504]	[1.04, 9.21]
Unemployment rate, %	Non-black	6.15	6.32	-	-
			[5.82, 8.05]	-	-
	Black	8.95	8.72	8.74	-0.02
			[8.25, 11.85]	[8.31, 11.92]	[-0.12, 0.26]

Note: The counterfactual sets the social capital of black workers equal to that of non-black workers; since this doesn't affect non-black workers, their outcomes are unchanged and thus marked with a dash. Bootstrapped 95% confidence intervals are given in brackets. The rightmost column gives the difference between the counterfactual simulation and the baseline simulation.

The counterfactual exercise is similar to that undertaken in Table 4. Specifically, the racial difference in social capital is set to zero to determine how much of the wage and unemployment gaps it can account for. The main finding is smaller than that of Table 4 but qualitatively similar: racial differences in jobs found through friends can account for 6% of the simulated racial gap in earnings of employed workers (95% confidence interval: [1%, 11%]). The simulated racial unemployment gap is essentially unchanged. This suggests that social connections affect labor market inequality primarily on the intensive margin.

The main takeaway from the structural estimation results is that the reduced form exercises seem to have done a pretty good job of capturing the main effect. To the extent that the model is able to discern, selection into employment and/or jobs found through friends does not seem to have much bearing on the main result, nor does the cumulative effect of experience over the life-cycle.

1.7 Conclusion

This paper is the first to estimate the contribution of social capital inequality to the aggregate wage gap between black workers and others in the US. I find that 6–10% of the gap can be explained by differences in jobs found through friends—a conservative estimate which may be biased downward by within-group correlation between social capital and human capital (discussed in Section 1.5.3). More broadly, these results demonstrate the importance of considering social capital in the study of wage inequality between groups.

The policy implications of these results may not be immediately obvious. Even if friends are important for finding work, can policy force people to make different friends? Perhaps not directly, but if social connections made early in life persist then neighborhood and school segregation may be important in determining labor market outcomes. Furthermore, this paper does have something to say about interventions that do not directly alter social networks. The main result of this paper is that a significant fraction of group inequality in labor market outcomes owes to heterogeneity other than worker human capital that prevents certain workers from finding jobs that reward their skills. Accordingly, a policy that affects the arrival rate of offers such as affirmative action in hiring may be able to counteract the racial imbalance in job-finding.

Chapter 2

Voting and Social Pressure Under Imperfect Information (with Alexander Clark)

2.1 Introduction

A growing body of empirical work suggests that social influence may have a critical role in motivating costly prosocial behaviors such as voting. Gerber et al. (2008) find that people are more likely to vote if their neighbors will learn of their participation, for example, and DellaVigna et al. (2014) estimate the value of voting 'just because others will ask' to be around \$5–\$15. Perez-Truglia and Cruces (2016) find that partisans give more to political campaigns if like-minded neighbors will learn of their contributions. These results suggest that social sanctions may be at play: civic duty may be enough to motivate some, but cannot explain why others only vote if their neighbors will find out. If social pressure is indeed critical, privacy would seem to undermine participation. We show that this is not always the case: our main result is that a certain degree of privacy and hypocrisy may actually be necessary for voter turnout or provision of a public good. These results are important because policy interventions often intentionally or unintentionally change social incentives or privacy, so understanding how information interacts with social mechanisms is critical to making sense of policy effects.

We write down a model of costly prosocial behavior enforced by social pressure in a setting with imperfect information. For ease of exposition and to make the application to an adversarial context clear, we follow the example of voting throughout the paper, despite a larger scope to the model.[1] We proceed from the simple idea that people interact in many contexts, and so it may be myopic to ignore interdependence among these interactions. In our model, players first simultaneously decide whether to abstain, vote for the left-wing candidate, or vote for the right-wing candidate. Voting actions are public knowledge for some players but non-verifiable for others, who can lie about whether they voted (throughout, we use the term 'privacy' to mean that an individual's voting action is private information). Then, each player chooses a level of cooperation with her peers. Intuitive restrictions lead us to focus on strategies in which players cooperate more with friends similar to themselves. While the model is uncomplicated, it is able to explain a variety of stylized facts from literatures in economics, political science, and psychology.

Our focus is on large elections in which no individual can hope to affect the outcome, in contrast to studies of small elections such as Borgers (2004). Nevertheless, voting can be induced by conditional network formation. Even self-interested individuals who obtain no intrinsic benefit from voting may vote in equilibrium, as they expect to be rewarded with stronger friendship with other voters of the same political affiliation. Others may abstain, because their costs of voting outweigh the benefits of stronger ties to voting peers. This social pressure may operate alongside other forces such as expressive utility, whereby voters are motivated by concerns apart from influencing the outcome of the election, but we can capture

[1]Frey and Meier (2004), for example, provide experimental evidence that individuals are more likely to contribute to a charity if they learn others are as well; such behavior can be explained by our model.

the main stylized facts about voting without expressive utility.

In our model, those who know that peers will learn of their participation will be more inclined to vote, as in the empirical literature. Despite this, we derive an unexpected non-monotonicity in the provision of information: maintaining the privacy of some individuals' participation may actually increase overall turnout (Proposition 2.5). If someone cannot be induced to vote even by social pressure, allowing them privacy at least enables hypocritical encouragement of others to vote. This can occur even if the cost of voting is identical across players (owing to asymmetries in network structure), and even if hypocrites' abstention can be inferred. Another counterintuitive finding is that increasing the level of punishment can in some cases lower turnout: if there are complementarities in the value of friendship, for example, raising punishment can lower overall cooperation to the point that social punishment loses its sting (Proposition 2.4).

Our model features a weighted and directed network that describes the capacity for benefit from cooperation among all pairs of players. This affords us specific predictions for how an individual's place in the network facilitates prosocial behavior. For any network, we characterize the information structure that maximizes voter turnout, and give conditions under which increasing privacy increases maximal equilibrium turnout. For any network and a given information structure, we find the maximal-turnout equilibrium. Our model features, as an equilibrium outcome rather than an assumption, homophily in prosocial behavior: voters are more likely to have friends who also vote, and partisans are more likely to have like-minded friends.[2] Furthermore, our model can explain not just voting but other costly prosocial behaviors as well. Appendix A.4 uses Add Health data for empirical verification of both homophily

[2]This adds to the work of Kets and Sandroni (2015), who show that homophily can arise despite the lack of a direct preference for similar friends as a way to reduce uncertainty when matched agents must play a coordination game.

and correlation with number of friends for voting as well as blood donation, community service, and organ donor registration. These results corroborate prior evidence of homophily (e.g. Knoke (1990)).

Ours is not the first paper to theorize that social pressure may be a key motivation to vote. Levine and Mattozzi (2017) consider a voter turnout model where a social norm specifies a participation rate, and the norm is enforced by audits and punishment. Two political parties choose the social norm that is optimal according to a group objective function. One party will be advantaged depending on its size and properties of the cost of enforcing a social norm. Imperfect information is treated through comparative statics on the reliability of the auditing technology, where a less reliable audit results in more punishment.

Our paper focuses on individuals, and a social norm specifies the extent to which norm adherents should punish friends observed violating the norm. Instead of considering a single auditing technology, we consider the potential privacy of a player's decision to vote or abstain. This may be thought of as allowing for the possibility that a player could fool an auditor. Further, social pressure is exerted in our model through the opportunity to withhold cooperation in a friendship game with connections on a rich network. By considering privacy and social pressure at the individual level, we are able to consider how different network positions expose an individual to varying levels of social pressure, facilitating different levels of prosocial behavior.

Harbaugh (1996) and Ali and Lin (2013) suppose that there are social benefits to appearing ethical which motivate costly prosocial behavior. We build on this approach by allowing players to choose whether or not to exert social pressure, rather than simply assuming a fixed utility penalty for social norm violations. This affords us two key benefits. First, we can explain why and when people might choose to reward social signaling (and how this might depend on network position). Our results require no committed or behavioral types, pro-social motivations are only instrumental and, using subgame-perfect Nash equilibrium as our solution concept, we address the long-standing concerns of those skeptical of the individual rationality of voting (Downs (1957)) or social punishment. Surveying the theoretical literature on voting, Feddersen (2004) notes the difficulty in informal appeals to social pressure: "Social pressure relies on followers to reward and punish each other at the direction of a leader. However, . . . it is not clear how this solves the problem, since followers will have the same incentive to shirk on exerting social pressure that they do to shirk on voting in the first place." By explicitly modeling social pressure as being exerted in a friendship game, where friendship is essentially a technology where the cooperation of two friends is complementary, we are able to show that social pressure can be rational.

Second, explicitly modeling social pressure yields an interesting prediction unique to our approach: affording more privacy can in some cases actually increase turnout as it can increase hypocritical social pressure. By contrast, in a model with fixed social pressure, increasing privacy will only decrease turnout.

One can also imagine another, similar motivation for voting: if ethical players (who will vote even absent social pressure) provide more valuable friendships than others, then even players with no ethical motivation may choose to vote as it will signal that they are ethical, attracting more friends.[3] While in our model the value of each friendship is known, players choose to vote based on similar reasoning—they expect to be rewarded in a friendship game.

Our model complements other explanations of aggregate voting behavior as well. By facilitating the alignment of individual incentives with those of the group, it can explain why self-interested agents may follow a party leader (as in Shachar and Nalebuff (1999)) or act to

[3] We are grateful to an anonymous referee for this suggestion.

maximize the utility of their group (as in Feddersen and Sandroni (2006)), even at personal cost. In our model, a news report about the closeness of a race or an announcement from a political leader can easily be used as a correlating device to determine the level of social pressure in equilibrium, explaining why aggregate turnout can vary with the closeness or importance of the election (a stylized fact unexplained by expressive theories of voting) even if sanctions occur on a personal level.

Another important strand of literature including Karlan et al. (2009) and Ambrus et al. (2014) has looked at enforcing informal lending agreements with social collateral. The social pressure in our model is similar, but is applied to enforce contribution to a public good rather than a bilateral contract. Our treatment of imperfect information, which underlies our main results and highlights the usefulness of hypocrisy, is our main relative contribution.[4] Ali and Bénabou (2016) propose a complementary reason for privacy: with uncertainty about the social value of a public good, privacy can guard against conformity in contributions, which can disguise true preferences.

We start with a simple game of perfect information in Section 2.2, which demonstrates how social pressure can induce voting and result in homophily (friendship formed among those who behave similarly). Section 2.3 discusses equilibria, including characterizing the unique turnout-maximal equilibrium (the equilibrium in which the most players vote). Section 2.4 extends the game to include imperfect information, leading to our main results on

[4]Other related work includes Levine and Modica (2014), which considers how contribution to a public good can be enforced through peer punishment: each player is audited by one other, and punishments are meted out bilaterally using the posited punishment technology. By contrast, our network setting affords us specific predictions about how an individual's place in the network facilitates prosocial behavior. Rather than incorporating noisy signals of compliance, in our model the actions of certain individuals are private. An important qualitative difference here is that we find that some level of privacy may be necessary to facilitate prosocial behavior. Our network setting and imperfect information results also distinguish our model from that of Abrams et al. (2011), though we share the idea that costly voting can be induced by social pressure.

the effects of changes in the information structure. Section 2.5 compares the model's predictions to the stylized facts of voting behavior (including two facts, homophily and correlation between degree and prosocial behavior, which we document ourselves in Appendix A.4). Section 2.6 concludes. Appendix A.5 considers aggregate welfare and shows how the utility forms assumed in leader-follower or group-utilitarian models can be interpreted as reduced forms of our framework. Appendix A.6 considers alternative strategies, which require more idiosyncratic equilibrium knowledge, but allow for the Pareto-optimal allocation. Appendix A.7 connects our paper to other results in the network literature regarding eigenvalues of the sociomatrix.

2.2 Model

There is a finite set of players I, indexed by $i \in \{1, 2, \dots, n\}$. Each player $i \in I$ has an exogenous preference parameter $\theta_i \in \{l, r\}$, indicating left- or right-wing political affiliation ("liberals" and "conservatives" respectively).

The set I and the exogenously given matrix A form a graph $G = (I, A)$, with $A \in \mathbb{R}_+^{n \times n}$ describing the network between players. An element a_{ij} of A indicates player i's capacity for friendship with player j, and A_i is a row vector indicating i's capacity for friendship with each member of I. An entry $a_{ij} = 0$ indicates that i has no capacity for friendship with j, and $a_{ij} > 0$ indicates the potential for friendship benefiting player i at some intensity, similar to Ambrus et al. (2014). We follow the convention that $a_{ii} = 0$, but do not require capacity to be symmetric. So there may exist some $i, j \in I$ such that $a_{ij} \neq a_{ji}$, but we assume that $a_{ij} > 0$ implies $a_{ji} > 0$. The matrix A and set of players I remain fixed over time.[5]

[5]While the matrix of capacities A is exogenous and fixed, the realized levels of cooperation are endogenous choices of the players—this is the sense in which ours is a model of network formation.

Definition 1. We refer to players i and j as "friends" if $a_{ij}, a_{ji} > 0$.

There are two time periods: $t \in \{1, 2\}$. In period $t = 1$, each player i chooses a voting action in a two party election: i selects $v_i \in \{-1, 0, 1\}$, where $v_i = 0$ indicates that i did not vote, $v_i = 1$ indicates i voted for the right-wing candidate, and $v_i = -1$ indicates a vote for the left-wing candidate. These decisions are elements of the $n \times 1$ vector v.

In our model, friendship between two individuals is the opportunity to cooperate for mutual benefit, and it requires mutual effort. In period $t = 2$, for any pair of friends i and j, player i chooses a cooperation level $\rho_{ij} \in [0, 1]$. Cooperation with friends is costly, but worth the cost if both players cooperate; payoffs are detailed below (Equation 2.1).

To elucidate how social pressure can induce costly prosocial behavior such as voting, we begin with the assumption that i knows A_i and, for all friends j, whether j voted ($|v_j|$) and j's political preference θ_j. Of course, full information about others' actions is not realistic. In Section 2.4, we allow for imperfect information, which allows us to study the effects of changes in information sets and obtain our main results. Formally, let $\hat{v}_i \in \{0, 1\}$ be a signal of whether or not i voted, and for now assume that no player is able to hide her true action from her friends: $\hat{v}_i = |v_i|$. [6]

A strategy of player i is then $s_i \in S_i$, consisting of a voting choice v_i in period $t = 1$ and a cooperation level $\rho_{ij}\left(v_i, \theta_i, (\hat{v}_k, \theta_k)_{k:a_{ik}>0}\right)$ for each friend j in period $t = 2$, which may be contingent on political preferences and signals of past voting behavior.

[6] Even with imperfect information, costly voting in our model requires that at least some people know the voting behavior of some of their peers. Voter participation (though not whom you voted for) is public record in the US, so it is technically verifiable. More realistically, politics is a common topic of discussion, and deceiving friends may not be easily accomplished if it comes up. The ubiquity of 'I Voted' stickers on election day suggests that voters expect some reward for advertising their participation.

Each player is endowed with an expected payoff function $U_i : \times_{i \in I} S_i \to \mathbb{R}$, where

$$U_i\left(s_i, s_{-i}\right) = -c|v_i| + f\left(\sum_{j:a_{ij}>0} \min\left\{\rho_{ij}, \rho_{ji}\right\} a_{ij} - \rho_{ij}\epsilon\right) + g_i\left(v\right). \tag{2.1}$$

Player i's expected utility has three separate components. First, if i votes ($|v_i| = 1$) she incurs a private net cost $c > 0$, which includes the opportunity cost of going to the polls as well as any other direct costs or benefits of voting. The cost of voting is common knowledge and uniform across all players, but this is without loss of generality for our analysis. The cost of voting matters only as it is compared to A_i, and so heterogeneity in costs can be captured by scaling A_i.

Second, i cares about friendship with peers. Her realized value of friendship with j depends on their cooperation choices, ρ_{ij} and ρ_{ji}, and the capacity a_{ij}. The first term, $\min\left\{\rho_{ij}, \rho_{ji}\right\} a_{ij}$, signifies the mutual nature of friendship: a pair of friends realizes their capacity for friendship only to the extent that both are willing. That is, friendship has a production function in which the efforts of two people are complementary. The second term, $\rho_{ij}\epsilon$, signifies i's cost of cooperation, which includes opportunity cost and is incurred regardless of j's level of cooperation. The intuition is simple: it is not worth spending time and effort trying to be friends with somebody who is not reciprocating. For example, player i might leave her Saturday open to socialize with a friend, j. However, if j is unwilling to socialize, no socialization occurs and i might find she has squandered an opportunity to make other plans. We assume that all friends are worth the effort: $0 < \epsilon < a_{ij}$ for all $a_{ij} > 0$. However, the total value of cooperation at stake need not be large—DellaVigna et al. (2014) estimate the value of voting 'just because others will ask' at around \$5–\$15, a small sum of influence easily wielded by one's friends. The function $f : \mathbb{R} \to \mathbb{R}$ aggregates i's realized value of friendship with each friend j such that $a_{ij} > 0$, and is strictly increasing.

Finally, i cares about the outcome of the election. Here $g_i : \{-1, 0, 1\}^n \to \mathbb{R}$, mapping a vector of votes v to i's expected utility. We normalize $g(0) = 0$, and assume that

$$g_i (v_i = 0, v_{-i}) \approx g_i (v_i = 1, v_{-i}) \approx g_i (v_i = -1, v_{-i}) . \tag{2.2}$$

In other words, no one individual has any hope of changing the outcome (as in large elections). Even when nobody else on the network is voting, this assumption will still hold if, for example, the network in question is only part of the total electorate (and others outside the network are voting). By explicitly ignoring the possibility of being pivotal, we stack the deck against the rational voter to emphasize the sufficiency of social pressure.

To motivate our focus on certain equilibria we will further assume that g_i is strictly increasing (decreasing) in v if i is conservative (liberal), such that $v' \geq v$ ($v' \leq v$)[7] implies

$$\sum_{j\,:\theta_j=\theta} (g_j(v') - g_j(v)) > c \tag{2.3}$$

for $\theta = r$ ($\theta = l$). In other words, contributing is socially efficient (within one's party) but, barring further incentives, individually irrational.

Notice that there is no intrinsic connection between the voting game in the first period and the cooperation game in the second. This is by design: social pressure in our model is a choice of the players rather than an assumption. Nonetheless, we will show that strategies that link actions in one game to actions in the other can effect a higher-welfare equilibrium and match well with the stylized facts. It is interesting to note that the existence of equilibria with nonzero turnout and social pressure is invariant to Equation 2.3; only our welfare result (Equation 2.4) will require this assumption. In other words, social pressure can be used to enforce participation in meaningless contests, or even reinforce bad behavior as discussed in Abbink et al. (2017). However, as players have an incentive not to coordinate on such an

[7] Throughout, we use the vector inequality notation $a \geq b$ to mean that $a_i \geq b_i$ for all i and strictly so for some i. The symbol $\geqq$ is used if $a_i \geq b_i$ for all i, allowing for the possibility that $a = b$.

equilibrium we think it is less likely to emerge.

While we follow the example of voting throughout for ease of exposition, our model can be applied to any prosocial behavior. In this interpretation everyone leans right, without loss of generality. Individual i contributes to the public good if $v_i = 1$, bearing personal cost c but having no discernible effect on the aggregate provision of the public good, which benefits her to the tune of $g_i(0, v_{-i})$. This setup can also handle non-partisan strategies, where cooperation is only conditioned on the decision to vote, not on political affiliation.

2.3 Equilibrium

We define Nash equilibrium as usual: a profile of strategies for all players in I such that no individual can achieve a higher expected payoff by pursuing a different strategy if the strategies of others are held fixed. While many strategies may yield equilibria, we argue for the use of just two specific strategies, based on a few simple principles which we describe and justify below. To avoid solutions based on incredible threats, we require subgame perfection throughout. Proposition 2.1 establishes that all Nash equilibria in the strategies we focus on are subgame perfect.

First, we will focus on equilibria in which voter turnout is nonzero. This choice is justified empirically: voter turnout and contributions to public goods are nonzero, and people do change their behavior when its visibility changes. Theoretically, this choice is justified because a voting equilibrium can provide higher aggregate welfare.[8] Our attention must therefore be on strategies with conditional network formation, in which players' second-period actions depend on the first-period actions of their friends. For if nobody is conditioning their

[8]Equation 2.4 provides conditions under which this is true.

behavior on your period-1 action, you will not choose a costly action like voting. An example of such conditioning is given by Sibona (2014), who finds that Facebook users are just as likely to "unfriend" work friends (with whom they presumably have good professional reason to maintain cordial relationships) for expressing "polarizing" viewpoints as they are to similarly punish high school friends.

Second, we will narrow attention to anonymous strategies. Our definition of anonymity is strong:

Definition 2. An "anonymous" strategy is one in which i's treatment of another player j can only depend on the signal $\hat{v}_j$ of j's period-1 action and the political preferences of i and j.

Voting is colloquially considered a social norm—something that everyone is supposed to do. Anonymous strategies are attractive in that they require little computation, information, or coordination for players and capture the universality of a norm. Nonetheless, they are sufficiently rich to explain a variety of stylized facts about prosocial behavior. Appendix A.6 discusses alternative strategies.

Third, we focus on a particular pair of anonymous strategies in which individuals cooperate more with friends similar to themselves in political affiliation and first-period action. Of course, it would be possible to have liberals cooperate more with conservatives, or voters with nonvoters, but this would be unnatural and contradict the data (see Section 2.5.4 and Perez-Truglia and Cruces (2016), who find evidence consistent with partisan sanctions).

Finally, how much cooperation should players withhold when called to do so? In period two, players face a choice from a continuum of cooperation levels. Given the structure of payoffs and a cooperation choice ρ_{ji}, i's best response must include $\rho_{ij} = \rho_{ji}$. We will consider strategies which specify a specific choice $\rho_{ij} \in \{1, 1-z\}$, where $z \in [0,1]$ is a publicly known parameter. z can be thought of as a punishment level; punishing a friend means denying him the opportunity to fulfill your cooperative capacity for friendship. Later, we consider how equilibria change with z.

Following these principles, we focus on two strategies. First, we define the social norm adherence strategy $s^\star(z)$. We will routinely suppress the argument z. If $s_i = s^\star$, the actions in periods 1 and 2 are as follows.

1. Vote ($v_i = 1$ if $\theta_i = r$, $v_i = -1$ if $\theta_i = l$).
2. If i voted, she chooses full cooperation ($\rho_{ij} = 1$) for all friends j unless they disagree on politics ($\theta_j \neq \theta_i$) or report abstaining ($\hat{v}_j = 0$),[9] in which case i restricts cooperation to $\rho_{ij} = 1 - z$. If i did not vote, she chooses $\rho_{ij} = 1$ if $\hat{v}_j = 0$ and chooses $\rho_{ij} = 1 - z$ for all other friends.

In other words, playing the adherence strategy $s^\star$ entails voting and cooperating fully only with voting friends of your party; with other friends, you choose a lower cooperation level $(1 - z)$.

We restrict attention to one kind of non-compliance with the adherence strategy, which we will call $s^0(z)$. The prescribed actions in periods 1 and 2 are:

1. Abstain from voting.
2. If $v_i = 0$ in period 1, i chooses $\rho_{ij} = 1$ for all friends j unless they report voting ($\hat{v}_j = 1$), in which case i chooses $\rho_{ij} = 1 - z$. In a subgame where i voted, i will take a second-period action according to the corresponding adherence strategy.

Thus playing non-compliance (s^0) entails abstaining, cooperating fully with nonvoting friends regardless of political preference, and restricting cooperation with voting friends to $(1 - z)$. We have nonvoting friends cooperate fully with each other both as a matter of conservatism and because we find it a more realistic choice. The choice is conservative because

[9] For now, of course, $\hat{v}_j = |v_j|$; in Section 2.4, this will not always be the case.

higher voter turnout in equilibrium could be obtained by having nonvoters restrict cooperation with all friends to $(1-z)$, making s^0 a less attractive strategy. A nonvoter choosing the maximum level of cooperation with another nonvoter does not necessarily imply the two are actively celebrating and rewarding each other for their abstention (though this might be a more plausible interpretation in the case of two partisan voters). Indeed, abstention can be principled or the product of apathy. Instead, we think of maximum cooperation as the most natural choice when the cooperation game is played in isolation, and so it may be as if these players simply do not let voting affect their social network formation. We will refer to those who play $\hat{v}_i = 1$ as *adherents* and those playing $\hat{v}_i = 0$ as *non-compliers*.[10]

When the z argument is suppressed, it is to be understood that both $s^\star$ and s^0 employ the same z in punishment choices. If all players are playing either $s^\star$ or s^0, then no other strategy will provide a strictly higher payoff. The strategies $s^\star$ and s^0 result in homophily, in the sense that under each of these strategies a given pair of like-minded players will cooperate more in the second period if they behaved similarly in the first period, or have similar political preferences.

Definition 3. An equilibrium in which each player plays either $s^\star$ or s^0 is called an "$(s^\star, s^0)$-equilibrium."

Definition 4. An equilibrium in which $|v_i|= 1$ for at least one player $i \in I$ is called a "voting equilibrium."

Proposition 2.1. For any graph G there exists an $(s^\star, s^0)$-equilibrium, and every $(s^\star, s^0)$-equilibrium is subgame perfect. If $c > za_{ij}$ for all $i, j \in I$ where $\theta_i = \theta_j$, then for an

[10]Note that both $s^\star$ and s^0 call on players to punish in proportion to the capacities a_{ij}. An alternative protocol of punishment could be accommodated by setting $z = 1$ and transforming A to represent the capacity of friendship that is at stake. E.g. if the social norm requires all punishments to be of a fixed severity, a_{ij} will be the minimum of this punishment value and the true capacity.

$(s^\star, s^0)$-equilibrium to be a voting equilibrium, there must exist a cycle of players who vote and share a political preference.

Proof. Existence and subgame perfection are proven in parts I and II. Part III proves the necessity of cycles.

Part I: Existence

We prove existence of a Nash equilibrium by example: the strategy profile where $s_i = s^0$ for all $i \in I$. Suppose $s_j = s^0$ for all $j \neq i$. Then player i's utility playing $s_i = s^0$ (left hand side) cannot be improved by unilateral deviation, since she is enjoying full cooperation and not incurring any voting cost. For any $\{\rho_{ij}\}_{i,j\in I}$ and v (and recalling that $\rho_{ij} \in [0,1]$),

$$f\left(\sum_{j:a_{ij}>0} a_{ij} - \epsilon\right) + g_i(v) \geq -c|v_i| + f\left(\sum_{j:a_{ij}>0} \min\{\rho_{ij}, \rho_{ji}\} a_{ij} - \rho_{ij}\epsilon\right) + g_i(v).$$

Thus $s_i = s^0$ is a best response, and we have a Nash equilibrium.

Part II. Every $(s^\star, s^0)$-equilibrium is subgame perfect

Consider an outcome in which each player is playing either s^0 or $s^\star$. Consider a possibly off-path history for which R (or symmetrically, L) is the set of players who have political preference $\theta_j = r$ ($\theta_j = l$) and voted ($\hat{v}_j \neq 0$) at $t = 1$.

If $i \in R$ ($i \in L$), she will fully cooperate ($\rho_{ij} = 1$) with her friends $j \in R$ (L), and they will cooperate with her ($\rho_{ji} = 1$) as well, according to $s^\star$ and s^0. i will mutually punish her friends $j \in R^C$ (L^C), playing $\rho_{ij} = \rho_{ji} = 1 - z$—all best responses in the subgame, by Lemma 2.3. If $i \in (R \cup L)^C$, she will mutually cooperate with all friends $j \in (R \cup L)^C$, $\rho_{ij} = \rho_{ji} = 1$, and punish those in $j \in (R \cup L)$ at level z so that $\rho_{ij} = \rho_{ji} = 1 - z$—again, all best responses in the subgame, by Lemma 2.3. Thus, we have a subgame perfect equilibrium.

Part III.

Consider a voting equilibrium where some conservatives, without loss of generality, vote. From these voting conservatives, consider voters $1, \ldots, m$ such that $a_{i\,i+1} > 0$ for $i = 1, \ldots, m-1$. Such a collection of voters must exist, because $c > za_{ij}$ implies that each voter must be friends with at least two other voters to be induced to vote. For no cycle to exist among these players, these must be distinct players and we must have $a_{mi} = 0$ for all $i < m-1$. However, for $v_m = 1$ in equilibrium, m must be connected to another distinct player $m+1$ where $v_{m+1} = 1$ in equilibrium. This is true for any m, but this is impossible for a finite set of players. Therefore, a cycle of voters must exist: for some collection of voters $1, \ldots, m$, we must be able to find $m' < m-1$ and $m'' < m'-1$ so that $(m'-1, m', m'', m''+1, \ldots, m'-1)$ is a cycle. This establishes that some voters must be on a cycle.

□

We see that strategies $s^\star$ and s^0 will deliver an equilibrium free of incredible threats. Proposition 2.1 does not guarantee the existence of a voting equilibrium, but we see cycles to be a necessary network property for voting if costs are such that a player must be pressured by at least two friends before they might prefer to vote. If only one co-affiliated norm-adhering friend is required for each player, full turnout is trivially an equilibrium for any graph such that the subgraphs of only left- or right-wing players are both connected. Later, Proposition 2.3 will provide a procedure for finding the highest voter turnout achievable in equilibrium. Cycles are both computationally easy to detect on a graph and indicative of the health of a community. Communities are stronger when ties are resilient enough that removing one individual does not render the graph disconnected, and that an action provides feedback ultimately felt at the action's origin—both properties of cycles.

For any pair of friends $i, j \in$ and a fixed ρ_{ji}, there is a best response for i that includes $\rho_{ij} = \rho_{ji}$. There is no subgame perfect equilibrium in which $\rho_{ij} \neq \rho_{ji}$ after some subhistory for any friends $i, j \in I$.

Proof. Assume i and j are friends. Player i's realized net benefit of friendship from j is $\min\{\rho_{ij}, \rho_{ji}\}a_{ij} - \rho_{ij}\epsilon$. If $\rho_{ij} < \rho_{ji}$ after some subhistory, then i's payoff from friendship with j in the second period (left hand side below) is lower than if she matched ρ_{ji} (right hand side), since $a_{ij} > \epsilon$:

$$\begin{aligned}\min\{\rho_{ij}, \rho_{ji}\}a_{ij} - \rho_{ij}\epsilon &= \rho_{ij}a_{ij} - \rho_{ij}\epsilon \\ &< \rho_{ji}a_{ij} - \rho_{ji}\epsilon.\end{aligned}$$

If $\rho_{ij} > \rho_{ji}$, then again i's net benefit (left hand side) is lower than if she played ρ_{ji} (right hand side):

$$\begin{aligned}\min\{\rho_{ij}, \rho_{ji}\}a_{ij} - \rho_{ij}\epsilon &= \rho_{ji}a_{ij} - \rho_{ij}\epsilon \\ &< \rho_{ji}a_{ij} - \rho_{ji}\epsilon.\end{aligned}$$

Recall that U_i is strictly increasing in this benefit of friendship (Equation 2.1). Thus no strategy can include $\rho_{ij} \neq \rho_{ji}$ after any history in subgame perfect equilibrium. □

As a consequence, we see that reciprocal punishment is necessary to the existence of voter turnout. This helps emphasize the difference between punishment in our model and the kind of punishment, or discipline technology, at work in Levine and Mattozzi (2017) and Levine and Modica (2016).

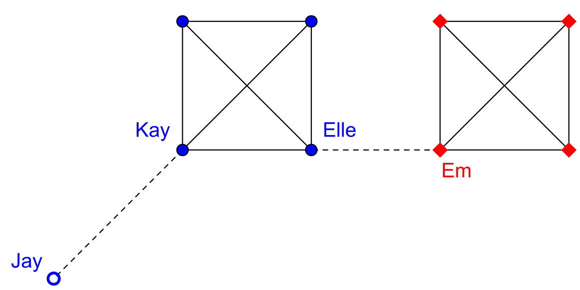
Proposition 2.2. There is no subgame perfect voting equilibrium in which, $\forall i$, i cooperates fully with all friends in every history in which she did not vote.

Proof. By way of contradiction, suppose a SPE exists in which i votes but cooperates fully with all friends in every history in which she did not vote.

Consider a history where i has not voted. Since i is cooperating fully in this history, by Lemma 2.3, any player $j \neq i$ must also choose $\rho_{ji} = 1$ if i and j are friends, to achieve Nash equilibrium in the subgame. Not voting will thus achieve i's maximal possible utility, and at least $c > 0$ more utility than any history involving voting. Therefore, voting was not a best response for i—a contradiction. $\square$

Example. Before formally discussing voting equilibria, we present the following example to elucidate the model. Figure 4 illustrates the turnout-maximal equilibrium for the example network pictured, assuming voting cost $c = \frac{3}{2}$, full punishment ($z = 1$), and that all nonzero capacities for friendship (indicated by lines) are equal to one.[11] Jay is not voting, since the cost ($\frac{3}{2}$) is greater than the potential benefit of cooperation with Kay (1). Kay gives up friendship with Jay and the cost of voting ($2\frac{1}{2}$ total) in order to vote, but receives the full cooperation of three fellow liberals in return (3). The other liberals are also voting, since the benefits of friendship with fellow voters outweigh the costs. Similarly, the conservatives are all voting, but Elle and Em do not realize their full capacity for friendship since they are politically opposed.

Figure 4: A two-party example



Circular nodes are liberals, diamond nodes are conservatives. A dashed line between two nodes i and j indicates bilateral capacity for friendship ($a_{ij} = a_{ji} = 1$) that is not fully realized ($\rho_{ij} = \rho_{ji} = 1 - z$); a solid line indicates that it has been fulfilled ($\rho_{ij} = \rho_{ji} = 1$). Filled-in nodes indicate players who are voting, and hollow nodes indicate abstention.

Following the discussion above, the remainder of this section focuses on equilibria

[11]This example also has three other equilibria: one with zero turnout, one in which only the conservatives vote, and one in which only the liberals save Jay vote.

involving only the adherence and non-compliance strategies $s^\star$ and s^0. There is a unique turnout-maximal equilibrium in these strategies. It is easy to find (Proposition 2.3), and its properties are described by Proposition 2.4 (these results are delayed until the more general imperfect information model is laid out in Section 2.4). The turnout-maximal equilibrium is of most interest when the social benefits of voting outweigh the private costs, providing greater aggregate utility than other equilibria that might be easily achieved, such as the one in which nobody votes and everybody cooperates—Equation 2.4 will characterize parameter values for this to be true for linear f.

The welfare-optimal equilibrium may not be achievable using strategies $s^\star$ and s^0, as it may require non-anonymous strategies. Appendix A.6 discusses non-anonymous strategies, and characterizes the welfare-optimal equilibrium. Several important contributions to the study of games on networks use properties of the eigenvalues computed from the adjacency matrix A. In our setting, however, the neither the eigenvalues nor the eigenvectors of a graph will determine the set of equilibria, as discussed in Appendix A.7.

2.4 Imperfect information

This section extends the baseline model to include imperfect information about voting behavior, and includes our main results on the effects of changes in information sets. Some players' voting actions can be hidden by lying, allowing them to avoid voting without incurring punishment. Making a player's voting action public makes her more inclined to vote (Lemma 2.4), as found in the empirical literature (see Gerber et al. (2008)). But the overall effect on turnout is ambiguous, as learning that one's peers are not voting makes one less inclined to vote. So giving non-voters privacy can increase overall turnout (Proposition 2.5), as it enables them to (hypocritically) pressure others into voting.

We begin by relaxing the assumption that all i's friends observe whether or not she voted. Instead, let $Q \subseteq I$ be the set of individuals who can choose to lie and report $\hat{v}_i \neq |v_i|$. Essentially, these players either incur no mental cost of lying or are unafraid of being caught, perhaps for example because they vote in a location distant from their friends. As before, each individual first takes a voting action. Next, if $i \notin Q$, i is compelled to play $\hat{v}_i = |v_i|$ as before. But if $i \in Q$, i can choose $\hat{v}_i \in \{0, 1\}$ regardless of her actual voting action. In other words, if $i \notin Q$ her friends will know whether she voted; if $i \in Q$ they will only know what she tells them (though of course they needn't believe her).[12] This choice of $\hat{v}_i$ is, for those who have it, taken in period 1. Finally, individuals choose levels of cooperation in period 2, as before. Q is public knowledge before any actions are taken.

The strategies s^0 and $s^\star$ remain as defined in Section 2.3, with the amendment that each involves truthfully signalling $\hat{v}_i = |v_i|$ even if $i \in Q$. In addition, we introduce a new strategy $s^H(z)$ called *hypocrisy*:

1. Abstain from voting.
2. Keep your first-period action hidden by reporting voting for your party ($\hat{v}_i = 1$).
3. If you reported voting, cooperate fully ($\rho_{ij} = 1$) with every friend $j : a_{ij} > 0$ of your party who reported voting. With all other friends, play $\rho_{ij} = 1 - z$. If you reported abstention, cooperate fully with friends who also reported abstention and restrict cooperation with others to $\rho_{ij} = 1 - z$.

Note that this strategy includes a combination of actions not available to those $i \in Q^C$: not voting, but reporting having voted. Thus, it is a strategy only available to those with privacy,

[12] A slightly richer model might also permit those whose actions are observed to lie, but this would complicate strategies without providing any new insights.

allowing voting behavior to be kept private in order to avoid punishment from voters. Those playing s^H will punish anyone observed abstaining, which is hypocritical since they themselves are not voting. In this section both those playing $s^\star$ and s^H are called *adherents*. Note that even if you're able to hide your voting behavior, you may choose to reveal your abstention by playing s^0 if, for example, most of your friends are nonvoters.

Because $c > 0$, it may be inferred that those in Q are hypocrites and not actually voting. However, punishing these players would only lower turnout—they cannot be induced to vote via social pressure since their voting behavior is not verifiable, but allowing them to hypocritically pressure others may increase overall turnout.

Lemma 2.4 states that if the only strategies played by others are s^0, s^H, and $s^\star$, then a player will have a best response among these strategies. A subgame-perfect Nash equilibrium exists in which all players play a strategy in $\left\{s^0, s^H, s^\star\right\}$. Moreover for any strategy profile in which $s_j \in \left\{s^0, s^H, s^\star\right\}$ for all $j \neq i$, an element of $\left\{s^0, s^H, s^\star\right\}$ is a best response for player i. For a player with privacy ($i \in Q$), $s^\star$ is never a best response.

Proof. Existence of an SPE is again easily shown by example: all players play s^0. If i deviates by voting, she will incur cost c and enjoy weakly less cooperation; if i deviates by reducing cooperation, this will only hurt her.

To prove the second part of the lemma, without loss of generality suppose throughout that $\theta_i = r$.

If i plays $\hat{v}_i = 1$ in the first period, then all conservative friends j playing $\hat{v}_j = 1$ will choose $\rho_{ji} = 1$. By Lemma 2.3, i's best response must include $\rho_{ij} = 1$. All liberal friends and those playing $\hat{v}_j = 0$ will choose $\rho_{ji} = 1 - z$. By Lemma 2.3, $\rho_{ij} = 1 - z$ is a best response in period 2. For $i \in Q$, s^H prescribes these period 1 and 2 actions. For $i \notin Q$, $s^\star$ prescribes these period 1 and 2 actions. Accordingly, the maximum utility achievable among all strategies that include $\hat{v}_i = 1$ can be attained by playing a strategy in the set $\{s^H, s^\star\}$.

Now suppose i chooses $\hat{v}_i = 0$. Then all friends j playing $\hat{v}_j = 1$ will choose $\rho_{ji} = 1 - z$. By Lemma 2.3, $\rho_{ij} = 1 - z$ is a best response in period 2. All friends j playing $\hat{v}_j = 0$ will choose $\rho_{ji} = 1$. By Lemma 2.3, i's best response must include $\rho_{ij} = 1$. s^0 prescribes these period 2 actions. Accordingly, s^0 gives the maximum utility possible among all strategies that include $\hat{v}_i = 0$. These cases are exhaustive, so there is always a best response in $\{s^0, s^H, s^\star\}$.

Finally, if $i \in Q$ then $s^\star$ is never a best response, since to other players it is observationally equivalent to s^H while incurring the voting cost $c > 0$. □

Lemma 2.4 implies that removing a player's privacy can only make her more inclined to vote—voting is never a best response for players with privacy, but may be for others. In particular, if s is an equilibrium under $Q \ni i$ in which i prefers $s^\star$ to s^0, then there exists an equilibrium with i removed from Q in which i plays $s^\star$ and all others' strategies remain as before. This is because the other players are indifferent between i playing s^H when $i \in Q$ and i playing $s^\star$ when $i \notin Q$.

Lemma 2.4 states that the adherence of others makes adherence as attractive as possible—a form of complementarity. This auxiliary result leads to Proposition 2.3, which provides existence of the turnout-maximal equilibrium given the new set of strategies, and gives a simple way to find it. Among strategy profiles in which $s_j \in \{s^0, s^\star, s^H\}$ for all $j \in I \setminus \{i\}$, the one in which those in Q play s^H and all others play $s^\star$ maximizes $U_i(s^\star, s_{-i}) - U_i(s^0, s_{-i})$ for $i \in I$ and maximizes $U_i(s^H, s_{-i}) - U_i(s^0, s_{-i})$ if $i \in Q$.

Proof. Throughout, define $\tilde{f}(x) = f((1-\epsilon)x)$—the utility benefit net of costs ϵ to x units of friendship realized when cooperation is reciprocated, as it always is in equilibrium.

Without loss, let $\theta_i = r$. If s_{-i} features only adherents, i will be punished by all friends for playing s^0: for each of i's friends j, j will choose $\rho_{ji} = 1 - z$ in response to observing $\hat{v}_i = 0$. When i plays s^0, she will reciprocate with $\rho_{ij} = 1 - z$. Therefore, the sum of realized friendship capacities for i is $\sum_{j:a_{ij}>0}(1-z)a_{ij} = (1-z)A_i e$, where e is an $n \times 1$ vector of ones.

When i chooses $s^\star$ and s_{-i} features only adherents, then each of i's friends j with $\theta_j = l$ will choose $\rho_{ji} = 1 - z$. All other friends j are adherents of the same party, and so will choose $\rho_{ji} = 1$ in response to observing $\hat{v}_i = 1$. Therefore, the sum of realized friendship capacities for i from playing $s^\star$ is $\sum_{j:a_{ij}>0} a_{ij} - z \sum_{j:a_{ij}>0,\theta_j=l} a_{ij} = A_i e - z A_i e_L$, where e_L is a binary vector indicating liberal players.

Recall, that given $s_j \in \{s^\star, s^H, s^0\}$, the minimum ρ_{ji} is $1 - z$. This strategy profile $s_{-i} \in \{s^\star, s^H\}^{n-1}$ thus provides the lowest possible payoff to i for playing s^0 among all profiles $s_{-i} \in \{s^0, s^\star, s^H\}^{n-1}$, by ensuring the minimum sum of realized friendship capacities. Furthermore, this strategy profile also maximizes the payoff to $s^\star$ for i, since the only friends withholding cooperation are those politically opposed friends (with $\theta_j \neq \theta_i$), who will never cooperate with i if choosing a strategy $s_j \in \{s^\star, s^H\}$. Then, we have maximized the difference as

$$U_i(s^\star, s_{-i}) - U_i(s^0, s_{-i}) = -c + \tilde{f}\left(A_i e - z e_L\right) - \tilde{f}\left((1-z)A_i e\right).$$

Now suppose $i \in Q$. The profile s_{-i} maximizes the payoff to s^H for i by the same logic as above—all friends are adherents, and the only cooperation forgone is with politically opposed friends who will never cooperate with i if she plays s^H. This profile thus also maximizes

$$U_i(s^H, s_{-i}) - U_i(s^0, s_{-i}) = \tilde{f}\left(A_i e - z e_L\right) - \tilde{f}\left((1-z)A_i e\right).$$

□

Proposition 2.3. The following procedure delivers the turnout-maximal subgame-perfect Nash equilibrium in which $s_i \in \{s^0, s^\star, s^H\}$ for all $i \in I$.

1. Start with those $i \notin Q$ playing $s^\star$ and $s_i = s^H$ if $i \in Q$.
2. Any player not playing a best response given the current strategy profile updates their strategy to s^0.
3. Repeat from step 2 until a fixed point is reached.

Proof. As argued in Lemma 2.4, those $i \in Q$ will never vote in equilibrium, choosing instead $s_i \in \{s^H, s^0\}$. Also, $i \in Q^C$ will have a best response in the set $\{s^0, s^\star\}$. Therefore, maximizing the number of voters is equivalent to maximizing the number of players in Q^C who will vote.

Let X^k be those $i \in I$ who switch to s^0 in the k^{th} iteration of the algorithm. If X^1 is empty, the algorithm reaches a fixed point immediately, and we have maximal turnout because no player in Q will ever vote, and the algorithm results in an equilibrium with all other players voting. Now we consider if X^1 is nonempty. We proceed by induction, showing that for any player in X^k for some k who is not voting according to the algorithm, there cannot exist an equilibrium in these strategies where they do vote. That is, the algorithm does not remove any players from the set of voters who might have voted in an equilibrium.

Basis step: Let $i \in X^1$.

Case 1: $i \notin Q$.

Then $U_i(s^\star, s_{-i}) < U_i(s^0, s_{-i})$ for all permissible profiles $s_{-i} \in \{s^\star, s^H, s^0\}^{n-1}$ because, by Lemma 2.4, step 1 maximizes $U_i(s^\star, s_{-i}) - U_i(s^0, s_{-i})$. Therefore, no strategy profile exists in which i would prefer $s^\star$ in equilibrium.

Case 2: $i \in Q$.

Similarly, $U_i(s^H, s_{-i}) < U_i(s^0, s_{-i})$ for all permissible profiles s_{-i} because, again by Lemma 2.4, step 1 maximizes $U_i(s^H, s_{-i}) - U_i(s^0, s_{-i})$. Therefore, no strategy profile exists in which i would prefer s^H in equilibrium.

Inductive step: Assume no player $j \in X^k$ will adhere to the social norm (play $s^\star$ or s^H). If $i \in X^{k+1}$, then i would only adhere if at least one player in X^k adhered to the social norm, i having preferred to deviate to s^0 only after we removed players in X^k from the set of adherents. However, since by assumption there is no equilibrium in which any $j \in X^k$ adheres, i cannot adhere either.

Thus, we have shown that if i is removed from the set of adherents according to the algorithm, there cannot exist an equilibrium in strategies $s^\star$, s^H, and s^0 where they do adhere. Therefore, the algorithm delivers the turnout-maximal equilibrium. □

The procedure described in Proposition 2.3 terminates in at most n steps, since someone is removed in every iteration and $\{s_i = s^0\}_{i \in I}$ is a fixed point. Similar results can be obtained if those playing s^H cooperate with everyone in the stage game, and those playing s^0 cooperate with those whose votes are hidden. Such an equilibrium would have greater aggregate cooperation in the stage game, but lower voting turnout, as the social cost of abstention would be diminished.

Why focus attention on the turnout-maximal equilibrium? First, the turnout-maximal equilibrium tells us whether a given network and privacy allocation permit a voting equilibrium. As we will show below in Figures 5 and 8, many networks require some privacy for there to be any turnout in equilibrium. Second, we might expect this equilibrium to emerge in the real world if it is preferred by players—when the aggregated welfare benefit to a political party outweighs the private costs of voting and forgone cooperation with friends (recall that friends of different parties cooperate only if neither reports voting). Equation 2.4 characterizes the parameter values for which the turnout-maximal equilibrium found by the algorithm (with associated voting vector $v^\star$) provides more aggregate utility to the type-θ party than the equilibrium in which nobody votes and everybody cooperates fully, assuming $f(x) = x$ to illustrate this fundamental tension cleanly.

$$\sum_{i\,:\theta_i=\theta} \Bigg(g_i\left(v^\star\right) \\ - c|v_i^\star| - z \left(\sum_{j:\theta_i=\theta_j} \mathbf{1}\left(\hat{v}_i^\star \neq \hat{v}_j^\star\right)\left(a_{ij}-\epsilon\right) + \sum_{j:\theta_i\neq\theta_j} \mathbf{1}\left(\hat{v}_i^\star + \hat{v}_j^\star \neq 0\right)\left(a_{ij}-\epsilon\right) \right) \Bigg) > 0. \tag{2.4}$$

While voting is socially efficient for the group (Equation 2.3), this equation emphasizes that actually motivating voting through social pressure comes at a cost of foregone cooperation, which will only be enough to justify the group benefits under certain conditions.

Proposition 2.4 describes some interesting properties of the turnout-maximal equilibrium.

Proposition 2.4. Assume f is weakly concave and twice differentiable. The maximal $\left(s^\star, s^H, s^0\right)$-equilibrium turnout level is nondecreasing in the punishment level z, and the sets of norm adherents in the maximal $\left(s^\star, s^H, s^0\right)$-equilibria across different values of z are ordered by the subset ordering.

Proof. If f is weakly concave and twice differentiable, then so is $\tilde{f}$ (defined in the proof to Lemma 2.4). Let $X^k(z)$ be the set of players who have updated their best response to s^0 by the k^{th} iteration of the algorithm from Proposition 2.3. Consider an arbitrary player $i \in X^k(z)^C$, who is either playing $s^\star$ or s^H at iteration k. Without loss of generality, let $\theta_i = r$. Let e be the $n \times 1$ vector of all ones and $\hat{v}_R^k$ ($\hat{v}_L^k$) be the binary vector identifying those conservatives (liberals) playing $s^\star$ or s^H at the current iteration of the algorithm.

It must be true for this player that

$$-c\mathbf{1}_{\{i \notin Q\}} + \tilde{f}\left((1-z)\, A_i e + z A_i \hat{v}_R^k\right) \geq \tilde{f}\left((1-z)\, A_i e + z A_i \left(e - \hat{v}_R^k - \hat{v}_L^k\right)\right).$$

We will show that $i \notin X^k(z)$ implies $i \notin X^k(z')$ for any $z' > z$. $i \notin X^k(z)$ implies

$$\tilde{f}\left((1-z)\, A_i e + z A_i \hat{v}_R^k\right) - \tilde{f}\left((1-z)\, A_i e + z A_i \left(e - \hat{v}_R^k - \hat{v}_L^k\right)\right) \geq c\mathbf{1}_{\{i \notin Q\}}.$$

We will show that the difference on the left hand side is increasing in z, establishing that if $i \in X^k(z)^C$ then $i \in X^k(z')^C$ for any $z' > z$. The derivative is

$$\tilde{f}'\left((1-z)\, A_i e + z A_i \hat{v}_R^k\right)\left[-A_i e + A_i \hat{v}_R^k\right] - \tilde{f}'\left((1-z)\, A_i e + z A_i \left(e - \hat{v}_R^k - \hat{v}_L^k\right)\right)\left[-A_i e + A_i \left(e - \hat{v}_R^k - \hat{v}_L^k\right)\right].$$

The previous inequality implies $A_i \hat{v}_R^k \geq A_i \left(e - \hat{v}_R^k - \hat{v}_L^k\right)$. By concavity, $\tilde{f}'\left((1-z)\, A_i e + z A_i \hat{v}_R^k\right) \leq \tilde{f}'\left((1-z)\, A_i e + z A_i \left(e - \hat{v}_R^k - \hat{v}_L^k\right)\right)$. As $\tilde{f}$ is strictly increasing, both terms are strictly positive. We also know $\left[-A_i e + A_i \hat{v}_R^k\right] \geq \left[-A_i e + A_i \left(e - \hat{v}_R^k - \hat{v}_L^k\right)\right]$ given that adherence was preferred. Both of these terms are weakly negative, as e is an upper bound for either binary vector of voters. Therefore,

$$\tilde{f}'\left((1-z)\, A_i e + z A_i \hat{v}_R^k\right)[-A_i e + A_i \hat{v}_R^k] - \tilde{f}'\left((1-z)\, A_i e + z A_i \left(e - \hat{v}_R^k - \hat{v}_L^k\right)\right)[-A_i e + A_i \left(e - \hat{v}_R^k - \hat{v}_L^k\right)] \geq 0.$$

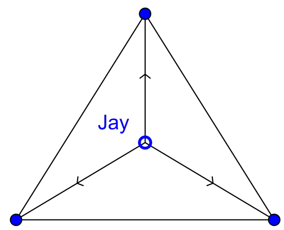
Thus, since the difference is increasing in z, $i \in X^k(z)^C$ implies $i \in X^k(z')^C$ for any $z' > z$. Thus, for every step of the algorithm (and in particular, at the terminal step) $X^k(z')^C \supseteq X^k(z)^C$. Accordingly, for a higher punishment level, there are more players choosing $s^\star$ or s^H in the turnout-maximal equilibrium. Those who play s^H will never vote, so we know this corresponds to higher voter turnout. The set of norm adherents is ordered by the subset ordering. □

It is known in the network literature that results like Propositions 2.3 and 2.4 hold for games of strategic complementarity.[13] While Proposition 2.4 shows that increasing punishment (z) increases maximal equilibrium turnout when the friendship aggregation function f is concave, the reverse can occur when f has a convex region. When there are complementarities in the value of friendship, for example, turnout may be non-monotonic in punishment, decreasing for high levels of punishment. Strikingly, militancy in social pressure can actually undermine the cause. The intuition is that when punishment is high, the level of realized friendship available to a player in equilibrium may be low no matter the strategy chosen. With f convex, some player may be less swayed to vote by the threat of punishment at low levels of realized friendship than if punishment were lower and the additional cooperation bought by voting more appreciable.

With the machinery of the full model with imperfect information laid out, we now derive our main results; Section 2.5 discusses them in light of the empirical evidence. As demonstrated by Lemma 2.4, public disclosure of an individual's voting participation makes that individual more inclined to vote. This is supported empirically by Gerber et al. (2008), who go on to interpret this as evidence that public disclosure of norm-related behavior increases compliance with norms. However, this need not be the case. As revealing the behavior of a nonvoter may discourage others from voting, increasing an individual's likelihood to vote does not necessarily imply higher overall turnout. In a similar vein, Panagopoulos (2010) uses an experiment to examine the mobilizing effects of pride and shame in voting behavior. Without accounting for the indirect effects of participants on others, the results of such experiments may be biased.

[13]Jackson and Zenou (2014) discuss this in their section 3.3.2. Our particular set of anonymous strategies effectively induce strategic complementarity, permitting these results to hold even though the game we consider is one of neither strategic complements nor substitutes.

Figure 5: Privacy can be critical



Arrows on edges indicate lopsided capacity for friendship, while unmarked edges indicate balanced capacity. Filled nodes indicate equilibrium voters; hollow, abstainers.

Eliminating privacy may eliminate hypocrisy and its potential uses. This is demonstrated in the following example. Consider 4 mutual friends of the same political preference and voting costs $c = \frac{3}{2}$ depicted in Figure 5. There is one central player, Jay, who is like a celebrity in terms of influence. Jay is an influence on others, but is himself barely influenced by his friends. Letting Jay be player 1, we can represent this with the asymmetric matrix of capacities

$$A = \begin{bmatrix} 0 & \frac{1}{1000} & \frac{1}{1000} & \frac{1}{1000} \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{bmatrix}.$$

Jay will never vote, having almost nothing to gain in the friendship game. As a result, for a high enough punishment level z there can be no voting in equilibrium. However, were Jay's vote private and the others' public, he might hypocritically affirm the importance of voting, thus creating an equilibrium where all other players are pressured into voting. In the privacy regime which permits the highest turnout, Jay alone has the option of privacy. This demonstrates that turnout need not be monotonically related to privacy.

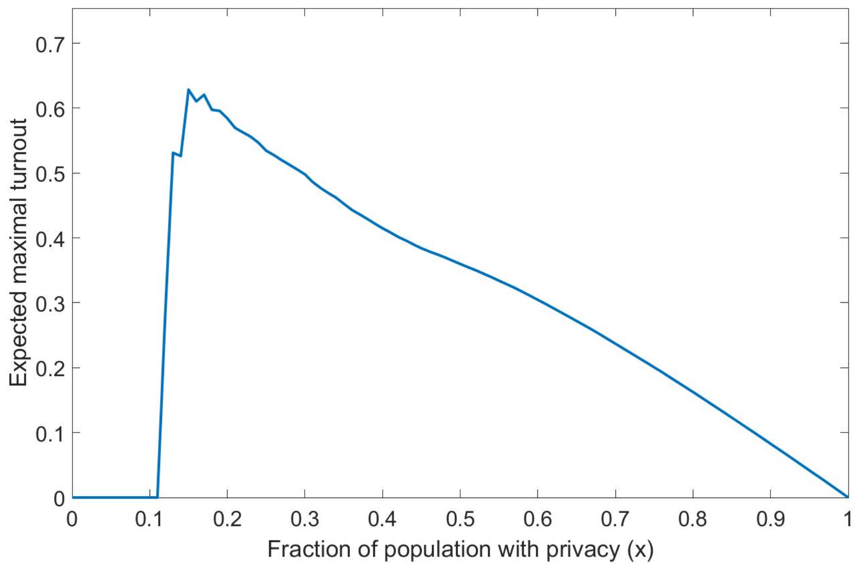
The potential for increased privacy and hypocrisy among non-voters to increase voter turnout extends beyond small contrived networks like that in Figure 5, as the following exercise shows. Consider a scenario in which a social planner randomly allocates privacy to a fraction x of the population without knowledge of individuals' network positions. How would expected maximal turnout[14] vary with x for a given network? The answer of course varies from network to network, but to get a better sense of the forces at play we perform the following exercise. First, we generate a random network of 100 nodes where each pair is connected with chance 7% (Figure 7). The cost of voting c is set to 3.5, all nonzero capacities equal 1, the punishment level is $z = 1$, and everyone belongs to the same political party. Next, we fix an x and calculate the expected maximal turnout at this privacy level by allocating privacy randomly 10,000 times and finding the turnout-maximal equilibrium each time. Finally, we repeat this step for each x between 0 and 1 in increments of 0.01.

Figure 6 shows that for this network, expected turnout is clearly non-monotonic in the level of privacy. Giving one more person privacy can lower turnout (as they will no longer vote) or increase it (as it enables them to pressure others). For this network, these forces tend to roughly balance when 15% of people have privacy, providing the highest expected maximal turnout of 62.8%.

Figure 7 shows this example network in the unique maximal turnout equilibrium for one particular allocation of privacy to 15 people. The voters (dark circles) cooperate with each other as well as those who hide their abstention (light circles). Of those publicly abstaining (diamonds), some simply forgo cooperation (lower right) as it would not outweigh the cost of voting. However, others (top left) are cooperating amongst themselves. In fact, one player (light diamond) has the opportunity for privacy but chooses to reveal her abstention to gain the

[14]Here the expectation is over random allocations of privacy to x people, and "maximal" refers to the equilibrium with the highest turnout.

Figure 6: Expected maximal turnout over privacy levels



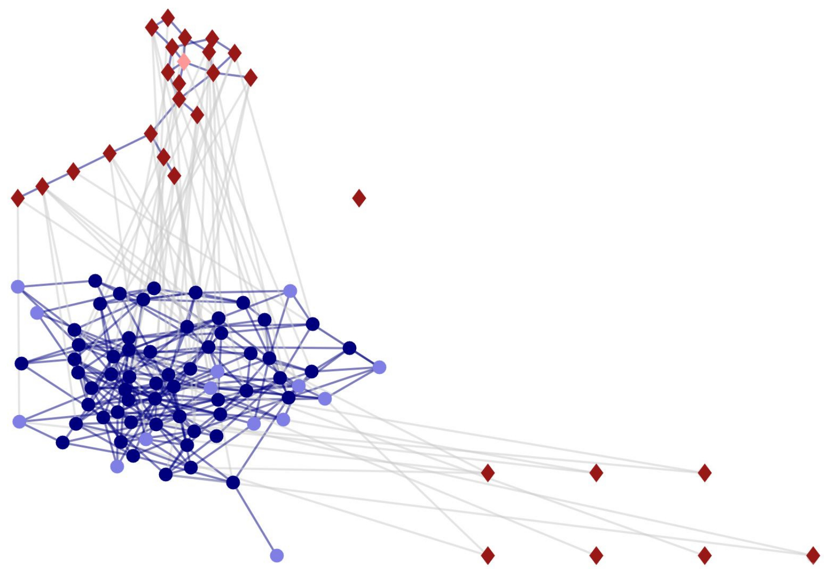
cooperation of her non-compliant friends. This emphasizes that non-compliance with a social norm can itself unify people within a subculture.[15] Since costs are uniform, all heterogeneity in actions owes entirely to the network structure.

Of course this is just an example, but the results are qualitatively robust. Other networks generated in the same manner usually share the same basic characteristics. Figure 8 plots summary statistics for another hundred[16] such networks. Each position on the horizontal axis represents a different randomly generated network (ordered by expected maximal turnout with optimal privacy), and the vertical positions of the markers above it represent the results for that network. For each network we compute, as above, the optimal privacy level

[15]In what Chandrasekhar et al. (2015) call 'DeGroot action models,' players learn about some state of the world by observing the actions of their neighbors. In such models, a clan (group that has more in-group connections than connections to the rest of the network), can fail to converge to the correct belief (see also Mossel and Tamuz (2014) and Golub and Jackson (2010)). While there is no learning in our model, this result is similar in flavor to our groups of non-compliers in that both models motivate players to imitate peers yet do not necessarily result in unanimous action, because the network structure affords some individuals more influence than others.

[16]Increasing the number of networks results in qualitatively similar results; the number 100 was picked to make the figure readable.

Figure 7: A large simulated network

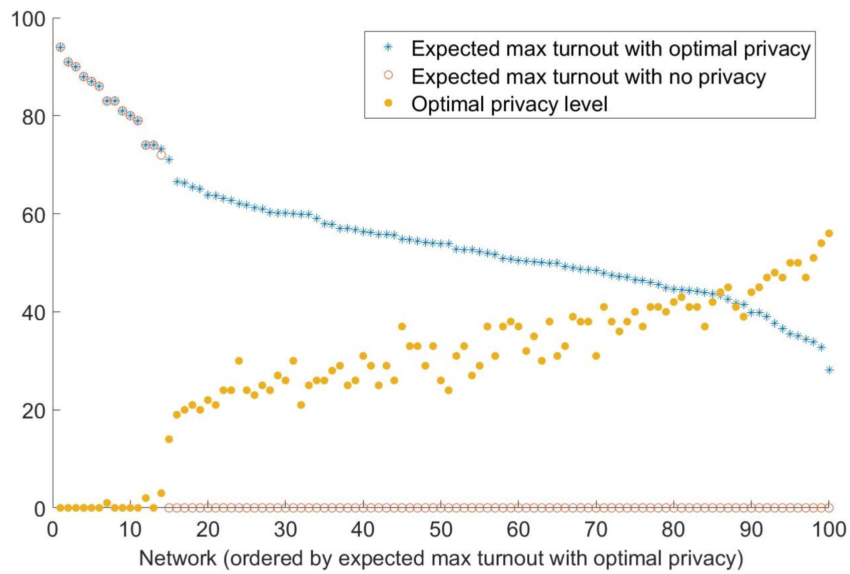


Round nodes are adherents and diamonds are non-compliers; lighter shades indicate privacy. Solid lines indicate full cooperation while faint lines indicate unfulfilled potential for cooperation.

(the fraction of the players that results in the highest expected maximal turnout, if privacy is randomly allocated), expected maximal turnout at the optimal privacy level, and expected maximal turnout with no privacy. On the left side of the figure are the networks for which zero privacy is optimal. For most networks, however, turnout is maximized by some nonzero level of privacy. Expected maximal turnout with optimal privacy is negatively correlated with the optimal privacy level, as nobody afforded privacy will vote. The most interesting part of this figure is the expected maximal turnout with no privacy (unfilled circles). This is zero for most of the networks, indicating that some privacy is often necessary for positive equilibrium turnout.

We obtained similarly non-monotonic turnout/privacy plots using different types of

Figure 8: Turnout for 100 simulated networks



networks,[17] costs,[18] and numbers of nodes. The main necessary condition for this non-monotonicity is that some people who cannot be induced to vote influence others who can. This is formalized in Proposition 2.5, which says exactly when increasing privacy will increase maximal turnout. The condition is simple and plausible: if there is a non-voter who could be swayed to vote with more pressure from his friends, then turnout can be increased by expanding privacy.

Proposition 2.5. The turnout in the turnout-maximal $\{s^\star, s^H, s^0\}$-equilibrium can be strictly increased with increased privacy $Q' \supset Q$ if and only if, in the maximal voting equilibrium under Q, there exists some non-voter $i \in Q^C$ for whom $U_i(s^\star, s_{-i}) \geq U_i(s^0, s_{-i})$ for some $s_{-i} \in \{s^\star, s^H, s^0\}^{n-1}$.

[17]Increasing the chance of pairwise attachment without changing costs increases the chance that a network's optimal privacy level will be zero, since it raises the fraction of people who can be pressured into voting. In addition to varying the chance of pairwise connection, we also considered preferential attachment, meeting through friends, and core-periphery networks.

[18]Increasing the cost of voting increases the number of people who cannot be pressured into voting, and makes it more likely that the optimal level of privacy is nonzero.

Proof. First we assume that in the maximal voting equilibrium under Q, there exists some non-voter in Q^C for whom $U_i(s^\star, s_{-i}) \geq U_i(s^0, s_{-i})$ for some $s_{-i} \in \{s^\star, s^H, s^0\}^{n-1}$ and show that turnout can be increased with increased privacy. Let S be the set of players who do not comply with the social norm. We can partition the set $S \cap Q^C$ (players in Q^C who choose s^0) into the two sets B_1 and B_2:

$$B_1 = \left\{i \in S \cap Q^C : U_i(s^\star, s_{-i}) \geq U_i(s^0, s_{-i}) \text{ for some } s_{-i} \in \{s^\star, s^H, s^0\}^{n-1}\right\},$$

$$B_2 = \left\{i \in S \cap Q^C : U_i(s^\star, s_{-i}) < U_i(s^0, s_{-i}) \text{ for all } s_{-i} \in \{s^\star, s^H, s^0\}^{n-1}\right\}.$$

By assumption B_1 is nonempty. Let i belong to B_1, and suppose without loss of generality that $\theta_i = r$. If $U_i(s^\star, s_{-i}) \geq U_i(s^0, s_{-i})$ for some $s_{-i} \in \{s^\star, s^H, s^0\}^{n-1}$, then $U_i(s^\star, s_{-i}) \geq U_i(s^0, s_{-i})$ if $s_{-i} \in \{s^\star, s^H\}^{n-1}$, since by Lemma 2.4, this s_{-i} maximizes $U_i(s^\star, s_{-i}) - U_i(s^0, s_{-i})$.

B_2 must also be nonempty—otherwise, all players in B_1 would have played $s^\star$ in the turnout-maximal equilibrium, having no noncompliant friends to tempt them to do otherwise (see the proof of Proposition 2.3).

Now we consider an alternative privacy regime $Q' = Q \cup B_2$, and we show the new maximal voting equilibrium under Q' will feature more players choosing $s^\star$.

If we apply the algorithm in Proposition 2.3, we see that no players will want to switch to s^0 in the first step of the algorithm, where $s_{-i} \in \{s^\star, s^H\}^{n-1}$. For every $i \in B_1$, $U_i(s^\star, s_i) \geq U_i(s^0, s_{-i})$ by the previous argument. This increases the voter turnout level by $|B_1|$.

Players in B_2 now play s^H. By Lemma 2.4, we need only check if a player $i \in B_2$ would prefer s^H to s^0. This earns them the highest possible payoff in the first period of the game, by avoiding the cost of voting, and in the second period they best respond by cooperating with all friends of the same political party and

$$U_i(s^H, s_{-i}) - U_i(s^0, s_{-i}) = \tilde{f}\left(A_i e - zA_i e_L\right) - \tilde{f}\left(\left(1 - z\right) A_i e\right).$$

Clearly, $A_i e - zA_i e_L \geq (1 - z) A_i e$, and with f increasing, the above difference is nonnegative. This demonstrates that players in B_2 best respond by choosing s^H.

Any player who was not in B_1 or B_2 found adherence to be a best response in the original maximal equilibrium. So adherence must still be a best response for these players, since all other players are now adherent, and Lemma 2.4 shows that full adherence maximizes the relative payoff to adherence. Altogether, we increased the number of voters by $|B_1| > 0$.

Now, we prove the opposite direction—that the possibility of strictly increasing the turnout in the turnout-maximal equilibrium by expanding privacy will imply that B_1 is nonempty. We prove the contrapositive. Suppose that B_1 is empty. For maximal turnout to increase by expanding Q, a nonvoter under the original privacy regime must switch to $s^\star$. But players in Q will never play $s^\star$ by Lemma 2.4, and players in B_2 will never play $s^\star$ by definition. So if B_1 is empty, maximal turnout cannot increase under Q'.

□

In general, the privacy set Q which allows for the highest turnout among maximal equilibria is characterized by Proposition 2.6. The intuition is simple: anyone who cannot be induced to vote by social pressure should be allowed privacy, so that they can themselves costlessly support voter turnout by pressuring others.[19] Note that Proposition 2.6 requires full knowledge of the network structure to formulate the optimal allocation of privacy.

Proposition 2.6. The privacy regime that creates the highest turnout in a turnout-maximal

[19]Similarly, maximal turnout can be achieved by non-anonymous strategies in which anyone whose costs prohibit them from being induced to vote is nonetheless afforded full cooperation by voters.

$\left\{s^{\star}, s^{H}, s^{0}\right\}$ equilibrium can be computed as follows.

1. Let $Q = \{\}$ and set $s_i = s^{\star}$ for all $i \in I$.
2. Find all $i \in I$ for whom $s^{\star}$ is not a best response, and put $i \in Q$. For all such i, their strategy is now s^{H}. Leave all else unchanged.

Proof. First we show this is an equilibrium. By construction, those who play $s^{\star}$ and vote according to this algorithm are best-responding. Updating other players' strategies to s^{H} does not affect the voters' utility since hypocrites and adherents play identically in the second period. Thus, those playing $s^{\star}$ are still best responding. Similarly, we see that hypocrites are best responding. They are best responding in the cooperation game per Lemma 2.3, and avoid the cost of voting. No hypocrite would prefer to deviate to a strategy s^{0} as they would be punished by voters and hypocrites, gaining cooperation in no friendships (since there are no other non-compliers). By Lemma 2.4 this is the only deviation to check, and so this must be an equilibrium.

Now suppose, by way of contradiction, there were a privacy set Q' permitting an equilibrium with higher turnout. This would require some player $i \in I$ who plays hypocrisy according to the algorithm to instead vote in equilibrium. Without loss, let $\theta_i = r$. At best, this player obtains utility

$$U_i(s^{\star}(z), s_{-i}) = -c_i + g_i(0, v_{-i}) + \tilde{f}\left(A_i\left(e - ze_L\right)\right),$$

equal to the utility obtained by playing $s^{\star}$ when everybody else is as well. However, $s^{\star}$ was not a best response for this player in this situation. Accordingly, there can be no equilibrium in which i votes. Thus, the algorithm generates the privacy set Q having the turnout-maximal equilibrium given the strategy constraints. □

2.5 Model predictions and stylized facts

This section compares the predictions of our model to the findings of the empirical literature on prosocial behavior and privacy. To our knowledge ours is the only model that jointly rationalizes this wide range of stylized facts about prosocial behavior and privacy, including explaining why individuals would choose to enact social punishments.

2.5.1 Social sanctions

Our basic premise that social sanctions are enacted to enforce norms is well grounded in the empirical literature. In the experimental context, for example, Fehr and Fischbacher (2004) find evidence of cooperative norms enforced by peer punishment. In the context of elections, Knack (1992) presents survey evidence that people both enact and expect social sanctions for failure to vote.

2.5.2 Publicizing voting behavior can induce voting

Gerber et al. (2008) show that threatening to disclose people's voting behavior to neighbors or household members can induce them to vote. In our model, this empirical phenomenon is easily explained. Making someone's voting decision public may induce them to vote, since it raises the possibility of punishment by peers (Lemma 2.4). Note that this stylized fact cannot be explained using a simple peer effects model in which players derive utility from taking actions similar to those of their peers. Player i already knows her own voting behavior, and the experiment does not change her knowledge about her peers' behavior, so it should not change her utility or her choices.

Conversely, making voting behavior more private can lower turnout. Funk (2010), using the introduction of the Swiss mail ballot option as a natural experiment, shows the importance of social incentives in voting. Despite making voting easier, the effect on aggregate turnout was negligible and voter participation was more negatively affected in smaller communities. This is consistent with our model: lowering the cost of voting may fail to increase turnout if the privacy of voting behavior is simultaneously enhanced. Furthermore, it highlights the relative importance of our social pressure mechanism.

2.5.3 Injunctive vs. descriptive norms

The non-monotonicity in the effect of information demonstrated in Section 2.4 can explain evidence from the psychology literature of the difference in effects between appeals to injunctive norms (what ought to be done) vs. descriptive norms (what is commonly done), which is important for policy design. Telling people that others are flouting a social norm is often found to be ineffective in effecting compliance—see e.g. Cialdini et al. (1990) and Cialdini et al. (2006). If we interpret such a message as increasing individuals' beliefs of how many people are unlikely to punish them, it is easy to see in the context of our model how they might encourage non-compliance. Making individuals' actions public only makes a social norm more popular if those individuals are adherents.

2.5.4 Homophily

In our model, a pair's capacity for friendship is more likely to be fulfilled if both people exhibit the same voting behavior (Remark 2.3). Political homophily (in the sense of friends sharing attitudes and behaviors) has been documented (e.g. Knoke (1990)), and in Appendix A.4 we corroborate previous work, empirically verifying homophily in both voting and party affiliation using data from the Add Health survey on a young (aged 18-26), nationally representative cohort. We also find homophily in other prosocial behaviors: giving blood, organ donor registration, and community service. This suggests that our model of enforcing social norms via conditional network formation may apply not just to voting but to other prosocial behaviors as well. Preferences surely play a role in explaining homophily, but our model provides another complementary reason for this empirical regularity.

2.5.5 Degree/voting correlation

Appendix A.4 also demonstrates that empirically, people with higher degrees (more friends) are more likely to vote. Those who reported voting in the most recent presidential election tend to report more friends than non-voters, robust to a variety of demographic controls (though not, of course, an exhaustive set). Furthermore, we document similar correlations between degree and other prosocial behaviors: blood donation, organ donor registration, and community service. This correlation is not a direct implication of our model for all networks, as homophily is. However, our mechanism gives an economic reason that prosocial behavior may depend on degree: those who lack the necessary influences may not be subject to sufficient social pressure to outweigh the cost of voting.[20]

2.5.6 Hypocrisy

Another interesting phenomenon arises in the imperfect information setting: hypocrisy. Individuals whose votes are hidden cannot be coerced to vote with social pressure. However, they

[20]Note, however, that this is only the case if the sum $\sum_{j \in I} a_{ij}$ is increasing in degree, contrary to the assumption sometimes made in the empirical literature for identification purposes that $\sum_{j \in I} a_{ij} = 1$ for all individuals.

can still contribute to overall turnout by punishing observed non-voters. Lying about voting is well established empirically (see e.g. Bernstein et al. (2001)), and our result is also consistent with the stylized fact that even those who do not vote themselves may publicly support voting in principle: a 2006 Pew poll[21] finds that even among those who admit that they rarely vote, 60% "completely agree" that it is one's "duty as a citizen to always vote."

2.5.7 Turnout/closeness correlation

Empirical work has found that voter turnout is higher for closer races (see, e.g., Blais (2000)). An important strand of the theoretical literature on voting, in particular Shachar and Nalebuff (1999) and Feddersen and Sandroni (2006), has sought to resolve the tension between this stylized fact and the seeming irrationality of individual voting. While our model does not address this directly, it does provide a mechanism that can rationalize key assumptions made in previous work. Appendix A.5 sketches out how a news report or an announcement from a political leader can be used as a correlating device to determine the level of punishment and thus the level of turnout, allowing players to coordinate on an equilibrium that provides higher group utility. Supposing that higher turnout leads to higher group utility in close elections (as is true under mild conditions assumed in the literature), all that is required for our model to match this stylized fact is that people know the stakes of the election, and expect punishment to depend on closeness.

[21] The Pew Research Center For The People & The Press (2006)

2.6 Discussion

This paper contributes to the literature on social enforcement by considering imperfect information about peer behavior, which allows us to explain a variety of stylized facts about social influence and voting. We demonstrate that privacy and the hypocrisy it enables can be essential elements of public good provision. In theorizing a mechanism by which individuals might be induced to bear the cost of prosocial behavior, we align individual incentives with those of the group, facilitating dynamic coordination as the group benefits of coordination change over time. Our mechanism thus enriches the microfoundations for leader-follower or group-utilitarian models of voting, demonstrating that their empirically verifiable predictions of aggregate dynamics (such as correlation between voter turnout and election closeness) can emerge even with a rational, self-interested electorate.

While our model ties together a wide range of stylized facts from the empirical literature, several key implications remain untested. Does cooperation in unrelated activities vary over election cycles, for example, as individuals enact social sanctions? Do those with the ability to hide their actions pressure others hypocritically? And if they do, does removing an individual's privacy prevent him from exerting such pressure? The ideal experiment might directly observe social sanctions as well as changes in voting behavior in response to randomly assigned privacy (and if possible, randomly assigned opportunities for cooperation). If social sanctions are essential to public good provision, then understanding if and when they occur may be essential to aggregate welfare accounting. In addition to the potentially adverse effect of reducing privacy, our model suggests that increasing punishment may in some cases lower aggregate turnout—evidence of this would also be interesting.

Finally, while our model includes only self-interested players, we do not necessarily view our results as an alternative to models in which civic concerns enter individuals' utility functions. Rather, our model can potentially explain how, over time, self-interested players might develop such civic concerns as a heuristic. Modeling this evolution explicitly in a bounded rationality setting may be another fruitful line of future research.

Chapter 3

Coordinated Shirking

3.1 Introduction

Given its outsize importance, it is no surprise that the financial crisis of 2008 quickly engendered a literature trying to explain its causes and mechanisms. Some examine root causes, like Taylor (2009), who argues that government policies allowed easy credit to create a bubble. Others focus on transmission: Acharya and Richardson (2009), for example, discuss how regulatory arbitrage prompted banks to maintain exposure to risky asset-based securities, allowing a shock in the housing market to propagate to the rest of the economy, while Hall (2010) discusses spillover from financial sector crises.

Yet there remains an important outstanding question: why didn't analysts at financial institutions and rating agencies better predict the failure of the securitized products which precipitated the crisis? Was there simply an unforeseeable shock given the state of financial models at the time, or was there more that could have been done to evaluate their riskiness? And if better research could have been done, why wasn't it?

This paper models an agency friction which may have led to the misuse of new financial technology and thus contributed critically to the financial crisis. Specifically, workers tasked with evaluating new financial technology may shirk their duty when a technology is sufficiently widespread in use, since their employers may not credibly be able to threaten to fire all of them.

In the model, a lone principal seeks to induce effort from a group of agents. New technology arrives exogenously each period; using it usually increases agents' production but on rare occasions causes production to fail. Those agents with access to the new technology can exert costly research effort to obtain a noisy signal of its promise. After obtaining this signal (or shirking and seeing no signal), these agents decide whether or not to use the new technology. The principal is unable to observe the agents' effort or the signal, but she can observe their production. She is able to punish agents who fail to produce, but at a cost which is convex in the number of agents punished. The main idea is that the principal may be unable to credibly punish large groups of agents who shirk simultaneously: a "too many to fail" problem. This is similar to the work of Acharya and Yorulmazer (2008, 2007) and Farhi and Tirole (2012), in which the promise of government bailouts in an aggregate downturn can prompt banks to correlate the returns of their investments. However, the focus here is on the friction introduced by the interaction of technological progress with the convexity in the principal's punishment function.

The main application is a worker-firm setting, and particularly the case of the financial services industry leading up to the financial crisis of 2008. Leading up to the crisis, new financial pricing models facilitated a credit boom, as the ability to package risky loans into seemingly safe securities expanded the market for such loans. Even when doubts arose as to the soundness of these models, their use continued right up until the recession. This paper's model can explain how the introduction of an especially useful technology can lead to such a downturn: when such a technology comes along, workers may be less likely to properly vet it, since if it fails the firm cannot credibly commit to firing everyone who used it. Furthermore, even workers who find out that the technology is flawed may continue to use it, since they do not fear punishment for doing so. This behavior has a flavor of herding; Devenow and Welch (1996) summarize papers on herding in financial markets.

Others (e.g. Acharya and Richardson (2009) and Rajan (2008)) have noted that analysts' compensation structures gave little incentive to care about the long-term performance of their work (a theme repeated here), but do not explain why reputation concerns or the threat of firing were not more effective restraints. This paper specifically explains why the threat of firing workers for poor performance may have been inadequate in these cases. Furthermore, it explicitly links the downturn to the advent of a new financial technology, helping to explain the timing of the crisis.

It also illuminates several possible solutions to the threat of coordinated shirking. First, expectations are important: if workers for some reason expect that nobody else is shirking, they too will be diligent. Second, shifting worker compensation from variable (bonuses) to fixed (salary) may exacerbate shirking, as it uncouples workers' incentives from their output. Finally, the problem of coordinated shirking can be resolved if there is a commonly known complete strict ordering of workers (such as seniority) that the firm can use as a basis for punishment. The most senior person of any group of prospective shirkers will be motivated to exert effort, unraveling the shirking behavior of the group.

Section 3.2 introduces the model. Section 3.3 derives conditions for equilibria, calculates production loss compared to the first-best case, and gives an example to aid intuition. Finally, Section 3.4 applies the model to the financial crisis and social crime such as speeding, and discusses policy implications for both. The Appendix considers another application of paper's model: herding in crime, such as speeding or looting, where the presence of other criminals encourages criminal activity by reducing the chance of punishment. This case has been considered by e.g. DiPasquale and Glaeser (1998); Becker (1968) is the foundational analysis of the economics of crime. This paper's relative contribution is to highlight the significance of agent anonymity: if there exists a natural ordering of agents, the principal can punish according to that ordering, causing the coordinated shirking to unravel.

3.2 Model

Time is discrete and infinite: $t = 0, 1, 2, ...$ There exists a continuum of risk-neutral agents of measure one,[1] indexed by $i \in [0, 1]$, and one risk-neutral principal. The agent maximizes expected wages and the principal maximizes expected output less wages; they share a common time discount rate $\delta \in (0, 1)$.

3.2.1 Individual production

An individual agent i's production function is

$$f_{it} = (1 - \tau_{it}\mu_{it}(1 - \chi_t)) a_{it}.$$

The arguments are individual i's technology level a_{it}, technology choice $\tau_{it} \in \{0, 1\}$, and an indicator of whether or not the new technology is sound $\chi_t \in \{0, 1\}$. Agent i can produce his technology level a_{it} unless he has access to a new technology ($\mu_{it} = 1$), chooses to use it ($\tau_{it} = 1$), and it turns out to be unsound ($\chi_t = 0$). The principal owns the output of production.

[1]Using a continuum of agents rather than a finite set of discrete agents eases the analysis, but is not qualitatively important.

3.2.2 Technology

Each period, new technology arrives exogenously. It affects the production of a fraction μ_t of the unit mass of workers, where μ_t is a random variable with support on the unit interval.

Specifically, assume that each period μ_t is drawn from a distribution Q, independent of history:

$$\mathrm{P}\left[\mu_t < \mu\right] = Q\left(\mu\right). \tag{3.1}$$

For agent i let μ_{it} be a binary variable indicating whether or not period t's new technology is applicable to agent i. The chance that the technology is relevant is μ_t:

$$\mathrm{P}\left[\mu_{it} = 1\right] = \mu_t$$

$$\mathrm{P}\left[\mu_{it} = 0\right] = 1 - \mu_t.$$

The agents for whom the new technology is relevant are those for whom $\mu_{it} = 1$ and these chances are independent across agents, so

$$\int_0^1 \mu_{it} di = \mu_t.$$

The individual's technology level is a_{it}, which indicates potential production. Technology is given by

$$a_{it} = 1 + \tau_{it}\mu_{it}g.$$

Here τ_{it} is an indicator variable which is 1 if agent i chooses to use the new technology and 0 otherwise. The new technology expands productivity by $g > 0$ percent, but is only available to agent i if $\mu_{it} = 1$, which occurs with probability μ_t.

The new technology is sound with probability π, indicated by the binary random variable χ_t:

$$\mathrm{P}\left(\chi_t = 1\right) = \pi$$

$$\mathrm{P}\left(\chi_t = 0\right) = 1 - \pi.$$

The random variables $\{\chi_t\}$ are independent across time.

3.2.3 Effort/shirking

When given the opportunity to use a new technology ($\mu_{it} = 1$), agents cannot directly observe if it is sound (χ_t). However, they have the opportunity to undertake costly effort (E) to obtain a noisy signal of χ_t. Specifically, by paying a cost c , the agent learns the value of η_t, where $\epsilon \in \left[0, \frac{1}{2}\right]$ is the chance that the signal is wrong:

$$\mathrm{P}\left(\eta_t = \chi_t\right) = 1 - \epsilon$$

$$\mathrm{P}\left(\eta_t = 1 - \chi_t\right) = \epsilon.$$

The random variables $\{\eta_t\}$ are independent across time. Forgoing effort will, as is usual in the literature, be called 'shirking' (S).

3.2.4 The principal

The principal pays each agent wa_{it} before production is realized. The principal then receives the fruits of production: $\{f_{it}\}_{i\in[0,1]}$, such that aggregate production is given by

$$f_t = \int_0^1 f_{it}\mathrm{d}i.$$

The principal cannot observe each agent's effort, but can observe output, and decides whom to punish. The cost of punishing a fraction λ of the agents is $r\left(\lambda\right)$. This cost is assumed to be weakly increasing in the number of agents punished and convex. Also assume that there are negligible fixed costs to punishing agents[2] , so $r\left(0\right) = 0$. Punished agents exit the game and receive the outside option b; they are replaced by agents identical to them.

Convexity of the punishment is a crucial assumption in the model. In the worker/firm context, punishment is akin to firing and replacing a worker. There may be a small number

[2] If there are fixed costs, the principal may prefer not to punish small numbers of failures.

of people who are nearby and qualified for the job whom the firm can use to replace a small number of workers at low cost. Replacing large numbers of workers, though, may oblige the firm to pay relocation and/or retraining costs for some of the new hires.

The principal does not have to punish all agents who fail to produce, but rather can punish failure at a rate γ, which may depend on the number of potential failures μ, such that an agent whose production fails is punished with probability $\gamma(\mu)$.

3.2.5 Timing

Each period, the following occur:

1. A technology state μ_t is realized
2. Agents with access to the new technology ($\mu_{it} = 1$) choose whether to exert research effort (E) or to shirk (S)
3. Agents who chose effort pay a cost c and receive a signal η_t of the new technology's promise, χ_t
4. Agents with access to the new technology choose whether or not to use it (τ_{it})
5. Agents receive wa_{it}
6. Production f_{it} is realized for each agent, given technology a_{it} and the realization of χ_t
7. The principal decides what fractions of agents to punish (at a cost $r(\lambda)$ of replacing a fraction λ of the agents)
8. Punished agents receive b and exit the game; the rest continue to the next period

In the worker-firm application, this is equivalent to workers negotiating a salary equal to a fraction w of their prospective productivity, which they are paid unless they are fired. The case of compensation based on performance is discussed in 3.4.1.

3.2.6 Strategies

Each agent with access to a new technology must choose effort (E) or shirking (S), which may depend on how widely useful the technology is (μ_t). For each realization of signal η_t as well as a null signal (if the agent shirks), the agent must choose whether to use the new technology. So an agent's strategy is a map from $[0,1]$ (the support of μ_t) to $\{\text{E}, \text{S}\} \times \{0,1\}^3$.

Given that a fraction $\bar{f}$ of agents produced successfully, the principal's strategy consists of a rate $\gamma\left(\bar{f}\right) : [0,1] \to [0,1]$ which determines the fraction of agents whose production failed to punish. If all agents pursue the same strategy, this is equivalent to writing $\gamma(\mu)$, since either μ_t agents will fail or none will.

3.3 Results

3.3.1 First-best

Fix a new technology μ_t and focus on the problem of an agent who has access to it. By ignoring the new technology, the agent is guaranteed production of 1.

The expected production if the agent does not exert research effort (shirks) but uses the new technology is

$$\text{E}\left(f_{it} | e_{it} = 0, \tau_{it} = 1\right) = \pi\left(1 + g\right).$$

If the agent does exert research effort and receives a favorable signal ($\eta_{it} = 1$), expected production using the new technology is

$$\text{E}\left(f_{it} | e_{it} = 1, \eta_t = 0, \tau_{it} = 1, \right) = \left(1 - \epsilon\right)\left(1 + g\right).$$

This is greater than the production of one guaranteed by choosing $\tau_{it} = 0$ if

$$g > \frac{\epsilon}{1 - \epsilon}.$$

If the agent exerts research effort and receives an unfavorable signal ($\eta_{it} = 0$), expected production using the new technology is

$$\mathrm{E}\left(f_{it}|e_{it} = 1, \eta_t = 1, \tau_{it} = 1, \right) = \epsilon \left(1 + g\right).$$

Trusting the signal and forgoing the new technology will be preferable if

$$g < \frac{1-\epsilon}{\epsilon}.$$

Thus it will be efficient for agents who exert effort to act according to the signal received if

$$g \in \left(\frac{\epsilon}{1-\epsilon}, \frac{1-\epsilon}{\epsilon}\right). \tag{3.2}$$

Expected production when exerting effort is then

$$\mathrm{E}\left(f_{it}|e_{it} = 1\right) = \pi\epsilon + \left(1-\pi\right)\left(1-\epsilon\right) + \pi\left(1-\epsilon\right)\left(1+g\right).$$

This means that exerting effort is efficient if the cost (c) and chance the signal is wrong (ϵ) are small enough relative to the chance of the technology being good (π) and the amount it improves production, (g):

$$c < \pi\left(1-\epsilon\right)g - \left(1-\pi\right)\epsilon. \tag{3.3}$$

For the remainder of the paper, assume that 3.2 and 3.3 hold. This is the basic requirement to ensure an interesting problem. If all agents were exerting effort, any failures in production would owe to chance. Since punishing agents is costly, in the first-best case all agents would exert effort (since it is efficient) and the principal would never punish. Given technology μ_t, expected aggregate production in the first-best case is

$$\int_0^1 f_{it}\mathrm{d}i = \left(1-\mu_t\right) + \mu_t\left[\left(1+\pi g\right)\left(1-\epsilon\right) + \pi\epsilon\right]. \tag{3.4}$$

Total welfare is simply output less agent effort costs, since wages are a transfer. Total welfare in the first-best case is

$$\left(1-\mu_t\right) + \mu_t\left[\left(1+\pi g\right)\left(1-\epsilon\right) + \pi\epsilon - c\right].$$

3.3.2 Equilibria

As in other principal-agent settings, it is a Nash equilibrium here for agents to always shirk and for the principal to never punish. However, the more interesting question is under what conditions effort can be induced.

The principal's problem

The principal's strategy consists of a policy $\gamma(\mu) : [0,1] \to [0,1]$: given the fraction of projects that failed, she has to decide what fraction of the failures to punish (γ). Assume that punishing at rate $\gamma \geq \underline{\gamma}$ will induce the agent to exert effort (conditions for this to hold will be derived in 3.3.2). Since punishment is costly, the principal will in all cases choose either $\gamma = 0$ or $\gamma = \underline{\gamma}$; anything between these two would be costly without inducing effort, while anything greater than $\underline{\gamma}$ would be effective but unnecessarily costly. The question then becomes when the principal should punish and when she should not.

Given a policy $\gamma\left(\bar{f}\right)$, let Γ be the subset of the unit interval for which the principal punishes agents. That is,

$$\Gamma := \{\mu : \gamma(\mu) = \underline{\gamma}\}.$$

For $\mu_t \in \Gamma$, agents know that they will be punished for failure with probability $\underline{\gamma}$, and are thus induced to exert effort in these states.

If the principal never punishes ($\gamma(\mu) = 0 \forall \mu$) , the agents will always shirk and the principal's expected value is simply the expected discounted value of production less wages:

$$\underline{\mathrm{U}} \equiv \frac{1}{1-\delta}\left\{(1-\bar{\mu}) + \bar{\mu}\pi(1+g) - w(1+\bar{\mu}g)\right\}.$$

The principal's value of employing policy $\gamma(\cdot)$ given a technology state μ_t for which the principal does punish is

$$U^P(\gamma(\cdot),\mu_t\in\Gamma)=(1-\mu_t)+\mu_t[\pi(1-\epsilon)(1+g)+\pi\epsilon+(1-\pi)(1-\epsilon)]$$
$$-(1-\pi)\epsilon r(\mu_t\gamma(\mu_t))-w(1+\mu_t g)+\delta\mathrm{E}U(\gamma(\cdot),\mu_{t+1}).$$

The principal's value of employing policy $\gamma(\cdot)$ given a technology state μ_t for which the principal does not punish is

$$U^N(\gamma(\cdot),\mu_t\notin\Gamma)=(1-\mu_t)+\mu_t\pi(1+g)-w(1+\mu_t g)+\delta\mathrm{E}U(\gamma(\cdot),\mu_{t+1}).$$

The difference is

$$\mu_t[-\pi\epsilon g+(1-\pi)(1-\epsilon)]-(1-\pi)\epsilon r(\gamma(\mu_t)\mu_t).$$

For a policy $\gamma(\cdot)$ to be optimal, this must be greater than zero. Rearranging and using the fact that in equilibrium the principal will punish at rate $\underline{\gamma}$, we have

$$\frac{(1-\pi)(1-\epsilon)-\pi\epsilon g}{(1-\pi)\epsilon}\mu>r(\underline{\gamma}\mu). \tag{3.5}$$

This equation defines when the principal will prefer to punish if she is able to commit to a punishment policy $\gamma(\cdot)$. Figure 11 plots both sides of this equation for the sample parameters and functional forms given in Section 3.3.2.

Definition 5. Let $\Gamma^\star$ be the set of all $\mu\in[0,1]$ that satisfy Equation 3.5 - the set of μ for which the principal can credibly threaten punishment under commitment.

However, given that wages are realized before production, if the principal is unable to commit to a policy $\gamma(\cdot)$ then she must prefer punishing the agents to deviating to autarky for all $\mu_t\in\Gamma$:

$$-r(\underline{\gamma}\mu_t)+\delta\mathrm{E}U(\gamma(\cdot),\mu_{t+1})\geq\delta\underline{\mathrm{U}}. \tag{3.6}$$

The principal's expected utility is

$$\mathrm{E}U(\gamma(\cdot),\mu_t)=\frac{1}{1-\delta}\int_{\mu\in\Gamma}\{\mu[-\pi\epsilon g+(1-\pi)(1-\epsilon)]-(1-\pi)\epsilon r(\mu\underline{\gamma})\}\,\mathrm{d}\mu$$
$$+\frac{1}{1-\delta}\int_0^1\{(1-\mu)+\mu\pi(1+g)-w(1+\mu g)\}\,\mathrm{d}\mu.$$

Definition 6. Let $\Gamma^{\star\star}$ be the set of all $\mu \in [0,1]$ that satisfy Equation 3.6 - the set of μ for which the principal can credibly threaten punishment when commitment is not possible.

Proposition 3.1. Regardless of whether she can commit, the principal's optimal policy $\gamma(\mu)$ will be a threshold rule:

$$\gamma^{\star}(\mu) = \begin{cases} \underline{\gamma} & \mu \in [0, \tilde{\mu}) \\ \gamma \in \{0, \underline{\gamma}\} & \mu = \tilde{\mu} \\ 0 & \mu \in (\tilde{\mu}, 1] \end{cases} . \tag{3.7}$$

(The optimal policy at exactly the threshold $\tilde{\mu}$ will depend on the distribution $Q(\mu)$ and the parameters of the game, and is explained in the proof.)

Proof. First consider the case with commitment. Since $r(\cdot)$ is convex, $r(\underline{\gamma}\mu)$ is convex in μ given any $\underline{\gamma} \in [0,1]$. Recalling that $r(0) = 0$, this implies that given $\underline{\gamma}$, if $\mu \in [0,1]$ satisfies Equation 3.5, then so does any μ' satisfying $0 < \mu' < \mu$. So letting $\tilde{\mu}$ be the supremum of the set of values in $[0,1]$ for which Equation 3.5 holds, the principal can and will optimally punish for all μ up to the threshold $\tilde{\mu}$. As discussed before, the principal will in all cases choose either $\gamma = 0$ or $\gamma = \underline{\gamma}$. So under commitment the principal's optimal policy is the threshold rule given by Equation 3.7, with $\gamma^{\star}(\tilde{\mu}) = \underline{\gamma}$ if and only if $\tilde{\mu}$ satisfies Equation 3.5.

Next, consider the case where the principal cannot commit to a policy $\gamma(\cdot)$. Consider a rule $\gamma(\mu) : [0,1] \to \{0, \underline{\gamma}\}$, and $\mu_0, \mu_1 \in [0,1]$ such that $\mu_0 < \mu_1$. Since $r(\cdot)$ is increasing, if μ_1 satisfies Equation 3.6 then so does μ_0. So if $\gamma(\mu_1) = \underline{\gamma}$, $\gamma(\mu_0) = \underline{\gamma}$. Letting $\tilde{\mu}$ be the supremum in $[0,1]$ such that Equation 3.6 holds, the principal will again follow a threshold rule given by Equation 3.7, with $\gamma^{\star}(\tilde{\mu}) = \underline{\gamma}$ if and only if $\tilde{\mu}$ satisfies Equation 3.6. □

If $r(\cdot)$ is differentiable at zero, a sufficient condition for $\Gamma^{\star}$ to be non-empty is

$$\frac{(1-\pi)(1-\epsilon) - \pi\epsilon g}{(1-\pi)\epsilon r'(0)} > \underline{\gamma}.$$

Otherwise, Equation 3.5 would hold with equality only at $\mu = 0$, after which the costs of punishment would outweight the benefits. While the case without commitment is perhaps more realistic, this paper will focus on the case with commitment, to emphasize that the driving friction here is not necessarily a commitment problem but rather the convexity of the principal's punishment function.

The agent's problem

Assume the chance of punishment is $\gamma(\mu)$ for agents who fail to produce, which may depend on the technology state μ, as this governs the maximum number of agents whose production can fail in the period. The agent will exert effort if the expected value of doing so exceeds the expected value of shirking. The value function therefore depends on the usefulness of this period's new technology, μ_t. The agent is guaranteed wa_{it} and pays c if the technology is useful to him (chance μ_t), and continues in the game ($\mathrm{E}V(\mu_{t+1})$) if the new technology is good (chance π) and the agent receives the correct signal (chance $1-\epsilon$). If either the technology is good but the signal wrong (chance $\pi\epsilon$) or the technology is bad and the signal correct (chance $(1-\pi)(1-\epsilon)$), the agent forgoes using the new technology and continues on in the game $\mathrm{E}V(\mu_{t+1})$. Only if the agent receives a false positive signal (chance $(1-\pi)\epsilon$) is punishment triggered, and even then there is only a $\gamma(\mu)$ chance of receiving it, with the agent continuing on ($\mathrm{E}V(\mu_{t+1})$) with complementary probability. If the technology is not useful to the agent (chance $1-\mu_t$), the agent continues on in the game with $\mathrm{E}V(\mu_{t+1})$.

$$\mathrm{E}V(\mu_t) = w + \int_{\mu\in\Gamma} \mu \left\{\pi\left[(1-\epsilon)(wg + \delta \mathrm{E}V(\mu_{t+1})) + \delta\epsilon \mathrm{E}V(\mu_{t+1})\right]\right.$$

$$\left. + \delta(1-\pi)\left[(1-\epsilon)\mathrm{E}V(\mu_{t+1}) + \epsilon\left(\gamma(\mu_t)b + (1-\gamma(\mu_t))\mathrm{E}V(\mu_{t+1})\right)\right]\right\} \mathrm{d}Q(\mu)$$

$$+ \int_{\mu\notin\Gamma} \mu\left\{\delta \mathrm{E}V(\mu_{t+1}) + \pi wg\right\} \mathrm{d}Q(\mu) + \delta\int_0^1 (1-\mu)\mathrm{E}V(\mu_{t+1})\,\mathrm{d}Q(\mu).$$

Simplifying and gathering terms, this becomes

$$\mathrm{E}V(\mu_t) = w + \int_{\mu\in\Gamma} \mu\left[\pi(1-\epsilon)wg + \delta(1-\pi)\epsilon\gamma(\mu)b\right]\mathrm{d}Q(\mu) + \int_{\mu\notin\Gamma}\mu\pi wg\mathrm{d}Q(\mu)$$

$$+\delta\left\{\int_{\mu\in\Gamma}\mu\left[\pi+(1-\pi)(1-\epsilon\gamma(\mu))\right]\mathrm{d}Q(\mu)+\int_{\mu\notin\Gamma}\mu\mathrm{d}Q(\mu)+\int(1-\mu)\mathrm{d}Q(\mu)\right\}\mathrm{E}V(\mu_{t+1}).$$

Using the fact that $\mathrm{E}V(\mu_{t+1}) = \mathrm{E}V(\mu_t)$ and solving for $\mathrm{E}V(\mu_t)$ yields

$$\mathrm{E}V(\mu_t) = \frac{w + \int_{\mu\in\Gamma}\mu\left[\pi(1-\epsilon)wg + \delta(1-\pi)\epsilon\gamma(\mu)b\right]\mathrm{d}Q(\mu) + \int_{\mu\notin\Gamma}\mu\pi wg\mathrm{d}Q(\mu)}{1-\delta\left\{\int_{\mu\in\Gamma}\mu\left[\pi+(1-\pi)(1-\epsilon\gamma(\mu))\right]\mathrm{d}Q(\mu)+\int_{\mu\notin\Gamma}\mu\mathrm{d}Q(\mu)+\int(1-\mu)\mathrm{d}Q(\mu)\right\}}. \tag{3.8}$$

For an agent for whom $\mu_{it} = 1$ when $\mu_t \in \Gamma$, with $\gamma(\mu_t) = \gamma$, the expected value of effort is

$$w - c + \pi(1-\epsilon)(wg + \delta\mathrm{E}V(\mu_{t+1})) + \delta(\pi\epsilon + (1-\pi)(1-\epsilon))\mathrm{E}V(\mu_{t+1}) + \delta(1-\pi)\epsilon(\gamma b + (1-\gamma)\mathrm{E}V(\mu_{t+1})).$$

The expected value of shirking is

$$w + \pi(wg + \delta\mathrm{E}V(\mu_{t+1})) + \delta(1-\pi)(1-\gamma)\mathrm{E}V(\mu_{t+1}) + \delta(1-\pi)\gamma b.$$

The difference is

$$-c - \pi\epsilon wg + \delta(1-\pi)(1-\epsilon)\gamma\left[\mathrm{E}V(\mu_{t+1}) - b\right].$$

So the agent will exert effort if the increased chance of continuation in the game compared with the outside option outweighs cost of effort and loss owing to false negative signals:

$$\delta(1-\pi)(1-\epsilon)\gamma\left[\mathrm{E}V(\mu_{t+1}) - b\right] > c + \pi\epsilon wg. \tag{3.9}$$

Equation 3.8 can be used to yield a condition for effort containing only parameters. So given a policy of the principal $\gamma(\mu)$, Equation 3.9 determines when the agents will exert effort. Figure 10 plots both sides of Equation 3.9 for the sample functional forms and parameters given in Section 3.3.2.

Definition 7. Let $\underline{\gamma}$ be the lowest possible punishment rate which induces effort (Equation 3.9 holds). With smooth expected values this is

$$\underline{\gamma} \equiv \frac{c + \pi \epsilon g \mathrm{E}V\left(\mu_{t+1}\right)}{\delta\left(1-\pi\right)\left(1-\epsilon\right)\left[\mathrm{E}V\left(\mu_{t+1}\right)-b\right]}. \tag{3.10}$$

It is obvious from Equation 3.9 that if costs and the outside option are positive, the value of the signal alone is not enough to induce effort in the absence of punishment ($\gamma = 0$). That is, only the threat of punishment can induce effort. To further ensure this is an interesting problem, assume that the level of noise ϵ is small enough that if failure is always punished ($\gamma = 1$), effort is induced. Equivalently, assume that

$$\underline{\gamma} \in [0, 1]. \tag{3.11}$$

In other words, assume that the noisiness of the signal is not so high as to prevent the principal from being able to induce effort. An equilibrium with effort is then simply characterized by $\{\underline{\gamma}, \tilde{\mu}\}$, where $\tilde{\mu}$ characterizes the principal's policy rule $\gamma(\mu)$ of the form given by Equation 3.7, and $\underline{\gamma}$ is characterized by Equation 3.9. More specific characterization requires functional forms, as in the example (Section 3.3.2).

Welfare loss

The equilibrium with effort only differs from the first-best case when $\mu_t > \tilde{\mu}$. In such a state, expected aggregate output is

$$\int_0^1 f_{it} \mathrm{d}i = \left(1-\mu_t\right) + \mu_t \pi \left(1+g\right).$$

The expected welfare loss for $\mu_t > \tilde{\mu}$ (compared to the first best, Equation 3.4) is thus

$$\mu_t \left[\left(1-\epsilon\right)\left(1-\pi\right) - \pi g \epsilon + c\right].$$

If Equations 3.2 and 3.3 hold this welfare loss will be positive.

Example

To aid understanding, this section considers a case with simple functional forms and some sample parameters. Let the principal's cost of punishing μ agents be

$$r(\mu) = R\mu^2.$$

Then from Equation 3.5, the principal's condition for punishment (assuming she can commit to a policy) becomes

$$\frac{(1-\pi)(1-\epsilon) - \pi\epsilon g}{(1-\pi)\epsilon}\mu > R\underline{\gamma}^2\mu^2.$$

The highest measure of failures the principal is willing to punish is then

$$\tilde{\mu} = \frac{(1-\pi)(1-\epsilon) - \pi\epsilon g}{(1-\pi)\epsilon R\underline{\gamma}^2}. \tag{3.12}$$

Intuitively, as the rate of punishment rises, the largest group the principal is willing to punish shrinks.

Assume a uniform distribution for μ on $[0,1]$, such that

$$Q(\mu) = \begin{cases} 0 & \mu < 0 \\ \mu & \mu \in [0,1] \\ 1 & \mu > 1 \end{cases} \cdot$$

Given that the principal employs a threshold strategy (equation 3.7), the agent's expected continuation value $\mathrm{E}V(\mu_t)$ (Equation 3.8) becomes

$$\mathrm{E}V(\mu_t) = \frac{w + \frac{\tilde{\mu}^2}{2}\left[\pi(1-\epsilon)wg + \delta(1-\pi)\epsilon\underline{\gamma}b\right] + \frac{1-\tilde{\mu}^2}{2}\pi wg}{1 - \delta\left\{\frac{\tilde{\mu}^2}{2}\left[\pi + (1-\pi)(1-\epsilon\underline{\gamma})\right] + \frac{1-\tilde{\mu}^2}{2} + \frac{1}{2}\right\}}. \tag{3.13}$$

No commitment

Since the principal will use a threshold strategy in equilibrium, her expected utility in equilibrium will be

$$\mathrm{E}U\left(\gamma^{\star}\left(\cdot\right),\mu_{t}\right)=\tilde{\mu}\left\{\left(1-\frac{\tilde{\mu}}{2}\right)+\frac{\tilde{\mu}}{2}\left[\pi\left(1-\epsilon\right)\left(1+g\right)+\pi\epsilon+\left(1-\pi\right)\left(1-\epsilon\right)\right]\right.$$
$$\left.-\left(1-\pi\right)\epsilon R\frac{\tilde{\mu}}{2}\underline{\gamma}^{2}-w\left(1+\mu_{t}g\right)\right\}$$
$$+\left(1-Q\left(\tilde{\mu}\right)\right)\left\{\left(1-\mu_{t}\right)+\mu_{t}\pi\left(1+g\right)-w\left(1+\bar{\mu}g\right)\right\}+\delta\mathrm{E}U\left(\gamma\left(\cdot\right),\mu_{t+1}\right).$$

This can be simplified to

$$\mathrm{E}U\left(\gamma^{\star}\left(\cdot\right),\mu_{t}\right)=\int_{\mu\in\Gamma}\left\{\mu\left[-\pi\epsilon g+\left(1-\pi\right)\left(1-\epsilon\right)\right]-\left(1-\pi\right)\epsilon r\left(\mu\underline{\gamma}\right)\right\}\mathrm{d}\mu$$
$$+\int_{0}^{1}\left\{\left(1-\mu\right)+\mu\pi\left(1+g\right)-w\left(1+\mu g\right)\right\}\mathrm{d}\mu+\delta\mathrm{E}U\left(\gamma\left(\cdot\right),\mu_{t+1}\right).$$

Sample parameters and results

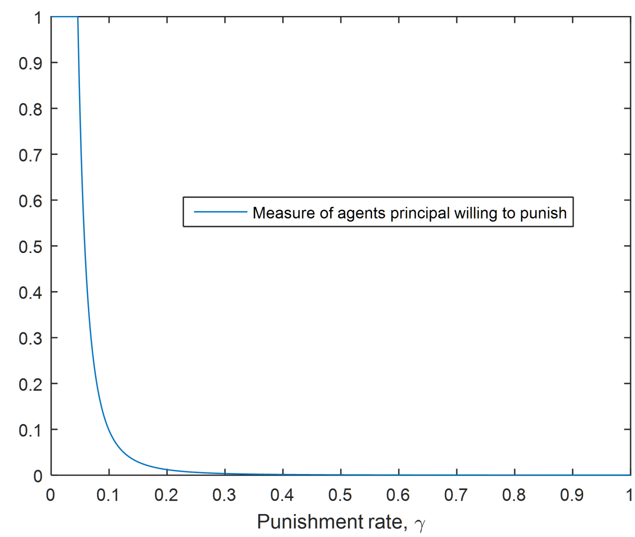
Consider the sample parameters in Table 11, for which Equations 3.2 and 3.3 both hold.

Table 11: Example parameter values

Parameter	Value	Description
δ	0.95	Common discount rate
π	0.85	Chance that new technology is sound
ϵ	0.01	Chance that signal is wrong
b	0	Agent's outside option
w	0.20	Agent's share of potential productivity
g	0.25	Growth rate
c	0.03	Agent's cost of research effort
R	1E6	Principal's punishment cost parameter

Given these parameters, Equation 3.12 gives the maximum measure of agents that the principal is willing to punish in equilibrium; Figure 9 plots $\tilde{\mu}$.

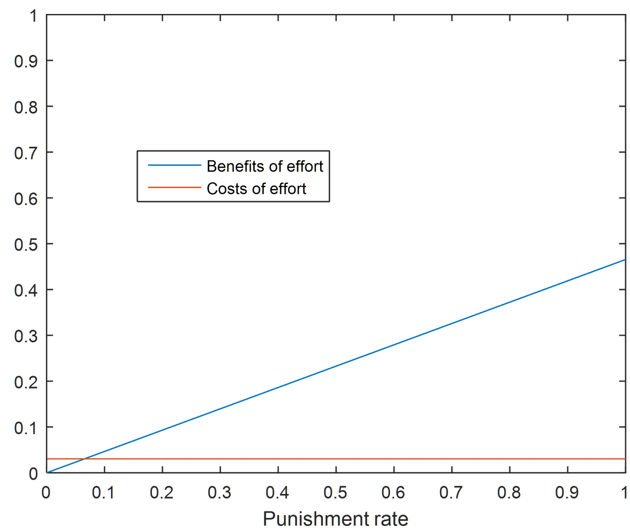
Figure 9: The principal's willingness to punish as a function of punishment rate



As expected, the willingness to punish falls with the punishment rate. Given this willingness to punish, the agent's costs of effort (right hand side of Equation 3.9) and benefits (left hand side) are given by Figure 10 as a function of punishment rate γ. The agent is only willing to exert effort if threatened with a punishment rate of $\gamma \geq 5\%$ — this is where benefits outstrip costs.[3]

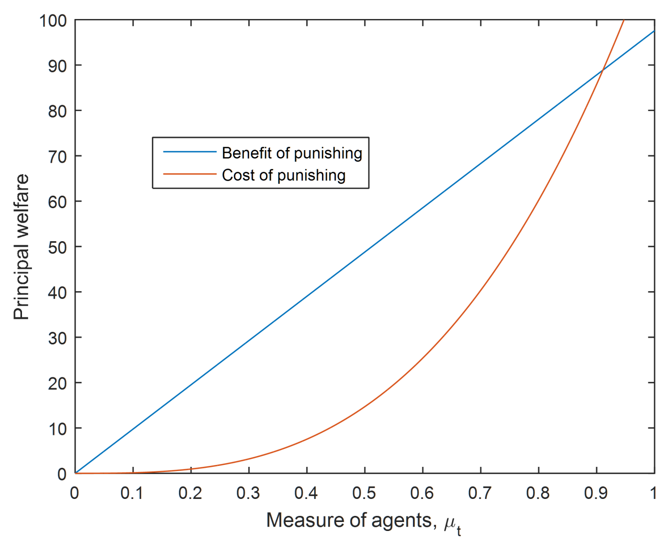
[3]For the case without commitment, the lower bar is roughly double (10.8%).

Figure 10: The agent's tradeoff as a function of punishment rate



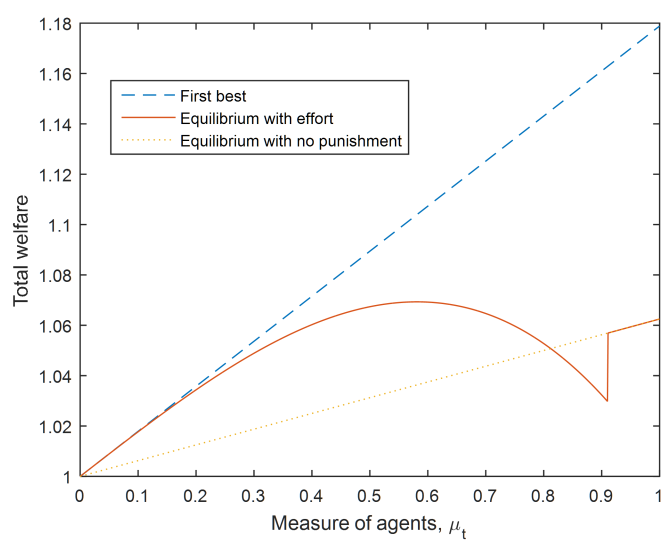
Finally, given $\gamma = 5\%$, Figure 11 plots the principal's costs (right hand side of 3.5) against her benefits (left hand side). She is willing to punish a group of agents as large as 0.91, but no larger. This demonstrates the main mechanism of the paper: the principal may be unwilling to punish large groups of shirkers.

Figure 11: The principal's tradeoff in equilibrium



In this equilibrium, workers shirk whenever $\mu > \tilde{\mu} = 0.91$, since they know they will not be punished. Figure 12 shows expected one-period welfare in this equilibrium compared to the first-best case with full effort and the autarkic equilibrium with no punishment. Total expected welfare (integrating over μ) is greater in the equilibrium with effort, but falls well short of the first best. Furthermore, in equilibrium both production and welfare exhibit non-monotonicity in the applicability of technology μ, as technology that is useful to everyone facilitates coordinated shirking.

Figure 12: Total welfare



3.4 Applications and discussion

3.4.1 Firms, workers, and aggregate downturns

Applied to a worker-firm setting, this model has a surprising prediction: a technological advance that is useful to too many people may actually trigger a downturn in production. This is the main result of the paper. Consider a new technology which is useful to many people and will work well most of the time, represented in the model by a string of realizations of μ which are all higher than $\tilde{\mu}$ and a string of realizations of χ which are mostly one.

Table 12: A (potentially) useful technology

t	1	2	3	4
μ_t	$> \tilde{\mu}$	$> \tilde{\mu}$	$> \tilde{\mu}$	$> \tilde{\mu}$
χ_t	1	1	1	0

Since $\mu_t > \tilde{\mu}$, which corresponds to a technological innovation relevant to a large fraction of workers, the firm is unable to credibly threaten to fire workers who fail to produce because the cost of replacing them is too high. So in these states, workers know they can shirk with impunity, as the firm can't fire everyone. Production in this example will be high for the first three periods but fail in the fourth for the large fraction of workers who are using the new technology. It is important to note that such states (with $\mu_t > \tilde{\mu}$) may arise rarely. The firm may in fact think that it is not subject to such agency frictions simply because a useful enough technology (high enough μ_t) has not come along in a while. The model can thus explain the boom and bust of the 2008 financial crisis as owing to the introduction of widely applicable new financial technology which was not vetted as rigorously as it might have been.

Furthermore, consider a worker who somehow learns that the new, widely useful technology is flawed ($\chi_t = 0$). Even with this knowledge, he will still choose to use the new technology if everyone else is, as it nets him a higher wage, and there is no chance of punishment. Research effort in this model has the flavor of a public good, since the technology is common. If it were possible for just one worker to exert the research effort and then share the signal with the others before technology choices were made, that would fix the problem. But once everyone is using a technology, it may be difficult to dissuade them, even if its flaws are made known. This may explain why even as doubts regarding the soundness of the credit boom arose, industry practices did not change.

3.4.2 The financial crisis of 2008

So coordinated shirking could have produced macro-level phenomena like those seen in the financial crisis. But did it? This section musters anecdotal evidence that not only were the conditions of the model met, but the forces at work in the model indeed contributed to key actors' decisions.

In the model, technology arrives exogenously and is publicly known. The development of financial technology that permitted the rapid growth of collateralized debt obligation (CDO) markets (including securities backed by subprime mortgages) quickly became public knowledge in the finance community. MacKenzie and Spears (2012)[4] report that "Although such techniques might originally have been proprietary, they quickly became common knowledge amongst investment bank quants. People moved from bank to bank, carrying knowledge

[4]MacKenzie and Spears (2012) present a fascinating "predominantly oral-history account of the development of the Gaussian copula family of models, which are used in finance to estimate the probability distribution of losses on a pool of loans or bonds, and which were centrally involved in the credit crisis." They conducted 95 interviews of financial services workers, of which 29 "took place before the onset of the crisis in July 2007."

of models with them, and quants – typically educated to PhD level or beyond – retained something of an academic habitus, talking about their work to their peers, and seeking opportunities to publish it." Thus not only was the new technology commonly available, as the model requires, but the model can arguably be applied at an industry level rather than just a firm level. A quant using industry-standard technology could reasonably expect this choice not to affect his future employability, even if said technology turned out to be unsound.

In particular, Gaussian copula models were ubiquitous in pricing CDOs; MacKenzie and Spears offer two reasons for this. By using an industry standard, traders could communicate and deal more easily with other firms. Furthermore, standardization made it easier for traders to claim the full expected future profits of such deals as justification for (current) bonuses, since the deals were priced using standard formulae. This corresponds well to the timing of the model, in which workers receive payment for prospective output before output is actually realized. Furthermore, it corroborates the model's result that employees are not punished for using technology that everyone else is using. The strategic productive complementarities of using a ubiquitous technology are not modeled in this paper, but might make an interesting extension.

"For all its deep entrenchment, however," MacKenzie and Spears write, "pervasive dissatisfaction with the Gaussian copula was expressed by interviewees even in our earliest interviews in 2006." As one of their interviewees explained in January 2007, "'So, the thing that is very interesting on credit [derivatives] is ... almost for the first time in finance we have models which are not so robust and are almost there as a kind of consensus, and that's all very well until something changes and something can change quite dramatically.' " So despite uncertainty or even pessimism about the use of a new financial technology, its use continued, as in this paper's model. In fact, MacKenzie and Spears summarize that "If a trader started to use a model substantially different from the Gaussian copula, then a position that was properly hedged on that model would not look properly hedged to risk controllers, unless the trader could succeed in the difficult task of persuading them to stop using the market-standard model." Thus while using the new technology was essentially free, *not* using it was made even more difficult by risk policies than this paper assumes.

Of course, many financial services employees were indeed fired or laid off during the financial crisis. This paper's model predicts total amnesty for large enough groups of shirkers, but it is simplistic in that the firm size remains the same — if the firm size shrank during recessions, for example, there could be firing during a downturn. And since workers would not individually be able to affect the probability of a downturn, the core mechanism of the model would remain intact.

3.4.3 Policy implications

Following the financial crisis of 2008, capping bankers' bonuses was widely discussed as a means of controlling risk-taking. In 2013, the European Union restricted banker bonuses to 100% of pay (or up to 200% with shareholder approval). (PricewaterhouseCoopers LLP 2013) This shifts the balance of bankers' compensation from variable (bonus) to fixed (salary). Can such a policy reduce risk-taking?

In this paper, the vital friction arises precisely because firms cannot punish workers by paying them less, only by firing them. Restricting variable compensation may thus exacerbate this friction, and increase the chance of a downturn. Variable compensation aligns workers' incentives with the returns of their projects. If in the model firms were allowed to pay workers their realized production rather than their potential production (similar to end-of-year bonuses), workers would be motivated to put forth research effort all the time, and the friction would disappear.

Of course, timing is key. If the outcomes of workers' projects are not realized for years after they are paid, then end-of-year variable compensation would be no different from fixed compensation during the year. In this case, something stronger would be needed, such as clawbacks (in which the firm is able to reclaim bonuses if long-term performance objectives are not met) or payment in stock or stock options (which would lose value if the projects failed). These prescriptions are not new, but this paper illuminates another reason they may be important: the inadequacy of firing workers as a punishment to induce effort.

Finally, the example of speeding (discussed in Appendix A.8) suggests another possible solution. If firms are able to somehow identify a salient individual for any possible subset of workers and commit to punishing that individual, coordinated shirking can be avoided. For example, the firm could commit to firing the most senior worker (in terms of tenure at the firm) in a group whenever their production fails. Of course, this may not always be possible, or enough. Workers may know that a lot of other people are using a new technology but remain unsure exactly who. Or a group of workers may have all started at the same time, making seniority an insufficiently distinguishing characteristic. Nonetheless, to the extent that firms can create a complete, strict ordering of workers that is known by all, it may be able to prevent instances of coordinated shirking.

Appendix A

Supplementary material

A.1 Ancillary tables for Chapter 1

Table 13 shows that the conditional wage premium for jobs found through friends differs across occupations, as does the frequency with which jobs are found through friends. The wage premia are calculated by a fixed effects regression akin to that in Table 2, but also including dummies for the interaction of occupation with whether a job was found through friends.

Table 13: Job market importance of friends by occupation

Occupation	# obs.	% found through friends	Wage premium (log pts)	Std. error
Professional and Technical	2,086	17.5	0.350	0.040
Sales	3,396	19.7	0.209	0.033
Operatives ex. Transport	892	25.9	0.201	0.056
Clerical and Kindred	2,055	23.8	0.201	0.038
Craftsmen and Kindred	1,768	26.2	0.181	0.043
Transport Equipment Operatives	1,299	24.4	0.134	0.049
Managers and Admin. ex. Farm	833	18.2	0.083	0.065
Farm Laborers and Foremen	3,073	20.6	0.064	0.038

Table 14 shows the results of log wages of jobs found directly and jobs found through friends separately regressed on age-adjusted, normalized AFQT score as well as controls for age, age squared, and year. The results for jobs found directly are similar to those of O'Neill (1990) and Neal and Johnson (1996), but the coefficient on race is much larger for jobs found through friends.

Table 14: Race, AFQT, and jobs found through friends

	Regression of log weekly earnings			
	Found directly		Found through friends	
Variable	Coefficient	Std. Err.	Coefficient	Std. Err.
AFQT	0.1601	0.0076	0.1479	0.0117
$AFQT^2$	0.0258	0.0068	0.0018	0.0106
Black	-0.0266	0.0182	-0.1155	0.0282
Age	0.2379	0.0179	0.3131	0.0220
Age^2	-0.0032	2.62 E-4	-0.0045	3.38 E-4
# obs.	19,052		5,381	

Note: Unreported coefficients include year dummies and a constant.

Table 15: Targeted statistics, data vs. simulated

Dependent variable	Indep. var./subset	Data	Simulated	95% CI
Log wage w_{it}	x_{it}	0.0430	0.0557	[0.0266, 0.0716]
	${x_{it}}^2$	-0.0006	-0.0022	[-0.0013, 0.0004]
	f_{it}	5.3717	5.9116	[5.7858, 6.0931]
	$f_{it} \cdot r_i$	-0.1780	-0.2037	[-0.3466, -0.1606]
	$f_{it} \cdot x_{it}$	0.0596	-0.0015	[-0.0279, 0.0143]
	$f_{it} \cdot x_{it}^2$	-0.0023	-0.0002	[-0.0008, 0.0012]
w_{it}, job through friends	$\hat{\eta}_i$	0.8980	0.9784	[0.9481, 1.0196]
Estimated fixed effect $\hat{\eta}_i$	1	5.5154	5.8099	[5.7311, 5.9488]
	r_i	-0.1211	-0.2044	[-0.2930, -0.1666]
Std. dev. of $\hat{\eta}_i$	Non-black	0.6262	0.6018	[0.5741, 0.6449]
	Black	-0.1031	-0.1799	[-0.2186, -0.1254]
Std. dev. of $(w_{it} - \hat{\eta}_i)$	Direct	0.5145	0.4376	[0.3982, 0.4652]
	Through friends	0.5042	0.4127	[0.3639, 0.4455]
Unemployment u_{it}	1	0.1675	0.3382	[0.2931, 0.4418]
	r_i	0.0243	0.0148	[0.0007, 0.0334]
	$\hat{\eta}_i$	-0.0128	-0.0491	[-0.0634, -0.0393]
	x_{it}	-0.0051	0.0004	[-0.0013, 0.0007]
Log tenure at current job	1	1.1801	2.4416	[2.0447, 2.7240]
	r_i	-0.0710	-0.0663	[-0.1621, -0.0062]
	$\hat{\eta}_i$	0.3964	0.2755	[0.2210, 0.3394]
	x_{it}	0.0865	0.0695	[0.0656, 0.0754]
	f_{it}	0.0923	-0.0678	[-0.1586, 0.0186]
Job found through friends f_{it}	1	0.4066	0.2645	[0.1004, 0.4401]
	r_i	0.0169	0.0041	[-0.0448, 0.0320]
	$\hat{\eta}_i$	-0.0176	-0.0097	[-0.0367, 0.0171]
	x_{it}	-0.0062	0.0010	[-0.0003, 0.0036]
Employed when found job	1	0.4439	0.5638	[0.5656, 0.6095]
	f_{it}	0.4306	0.5029	[0.5054, 0.5875]

A.2 Correcting error in measurement of fixed effects

Let $\hat{\eta}_i$ be the individual fixed effects estimated from Equation 1.10. Using $\langle \cdot, \cdot \rangle$ to denote covariance, the correct parameter values here are

$$\beta = \begin{bmatrix} \langle \eta, \eta \rangle & \langle X, \eta \rangle \\ \langle \eta, X \rangle' & \langle X, X \rangle \end{bmatrix}^{-1} \begin{bmatrix} \langle \eta, w^f \rangle \\ \langle X, w^f \rangle' \end{bmatrix} \tag{A.1}$$

whereas simply running a regression using the estimated fixed effects gives you

$$\hat{\beta} = \begin{bmatrix} \langle \hat{\eta}, \hat{\eta} \rangle & \langle X, \hat{\eta} \rangle \\ \langle \hat{\eta}, X \rangle' & \langle X, X \rangle \end{bmatrix}^{-1} \begin{bmatrix} \langle \hat{\eta}, w^f \rangle \\ \langle X, w^f \rangle' \end{bmatrix}. \tag{A.2}$$

Let o_i be the number of directly-obtained wage observations recorded for worker i. Then the estimated fixed effect is just the true fixed effect plus the average wage shock:

$$\hat{\eta}_i = \eta_i + \frac{\sum_{t:f_{ij(t)}=0} \epsilon^d_{ij(t)}}{o_i}. \tag{A.3}$$

Note that the covariance between $\epsilon^d_{ij(t)}$ and all right-hand side variables in Equation 1.11 is zero by assumption, so $\text{cov}\,(\eta, X) = \text{cov}\,(\hat{\eta}, X)$. And under the assumption that $\left\langle \epsilon^f_{ij(t)}, \epsilon^d_{ij(t' \neq t)} \right\rangle = 0$, $\langle \hat{\eta}, w^f \rangle = \langle \eta, w^f \rangle$. So the only element of Equation A.2 that requires correction is $\langle \hat{\eta}, \hat{\eta} \rangle$—we just need to find the true variance of η.

Taking the variance of Equation A.3 yields, assuming i.i.d. $\{\epsilon^d\}$,

$$\langle \hat{\eta}_i, \hat{\eta}_i \rangle = \langle \eta, \eta \rangle + \frac{1}{o_i} \langle \epsilon^d, \epsilon^d \rangle . \tag{A.4}$$

Next, observe that $\eta + \epsilon^d = \hat{\eta} + \hat{\epsilon}^d$, where $\hat{\epsilon}^d$ is the estimated residual from Equation 1.10. So

$$\langle \hat{\epsilon}^d_i, \hat{\epsilon}^d_i \rangle = \langle \eta, \eta \rangle + \langle \epsilon^d, \epsilon^d \rangle + \langle \hat{\eta}_i, \hat{\eta}_i \rangle - 2 \langle \epsilon^d, \hat{\eta}_i \rangle - 2 \langle \eta, \hat{\eta}_i \rangle$$

$$\langle \hat{\epsilon}^d_i, \hat{\epsilon}^d_i \rangle = \langle \eta, \eta \rangle + \langle \epsilon^d, \epsilon^d \rangle + \langle \hat{\eta}_i, \hat{\eta}_i \rangle - \frac{1}{o_i} \langle \epsilon^d, \epsilon^d \rangle - 2 \langle \eta, \eta \rangle .$$

Combining this with Equation A.4 yields an expression for the true variance of η in terms of observables

$$\langle \eta, \eta \rangle = \langle \hat{\eta}_i, \hat{\eta}_i \rangle - \frac{1}{o_i - 1} \left\langle \hat{\epsilon}_i^d, \hat{\epsilon}_i^d \right\rangle . \tag{A.5}$$

Using this in place of $\langle \hat{\eta}_i, \hat{\eta}_i \rangle$ in Equation A.2 will yield the true parameter estimates β.

A.3 Potential sources of bias

This section examines potential sources of bias in the estimation of key coefficients. Under the most likely scenarios, this paper underestimates the racial difference in social capital and its effect on wages.

A.3.1 Within-race correlation between social and human capital

Suppose that the log wage of a job found through friends is:

$$w_{ij(t)}^f = \beta_0 + \beta_{\text{hcap}} \eta_i + \beta_{\text{exp}} \left(\alpha_1 x_{it} + \alpha_2 x_{it}^2 \right) + \beta_s s_i + \epsilon_{ij(t)}^f . \tag{A.6}$$

This is essentially the same wage equation as Eq. 1.2, but social capital can now differ by individuals as well as race. Without loss of generality, decompose i's social capital into an average for i's race as well as an idiosyncratic component ξ_i:

$$s_i = r_i + \xi_i .$$

Using the orthogonality of $\bar{s}_{r(i)}$ with $\epsilon_{ij(t)}$, ξ_i, and the experience terms (and using $\langle \cdot, \cdot \rangle$ to denote covariance), this yields

$$\beta_s = \frac{\left\langle r_i, w_{ij(t)}^f \right\rangle - \beta_{\text{hcap}} \langle \eta_i, r_i \rangle}{\langle r_i, r_i \rangle} .$$

Similarly,

$$\beta_{\text{hcap}} = \frac{\left\langle \eta_i, w^f_{ij(t)} \right\rangle - \beta_s \left\langle \eta_i, r_i \right\rangle - \beta_s \left\langle \eta_i, \xi_i \right\rangle}{\left\langle \eta_i, \eta_i \right\rangle}.$$

Combining these yields

$$\beta_s = \frac{\left\langle r_i w^f_{ij(t)} \right\rangle \left\langle \eta_i, \eta_i \right\rangle - \left\langle \eta_i, r_i \right\rangle \left\langle \eta_i, w^f_{ij(t)} \right\rangle}{\left\langle r_i, r_i \right\rangle \left\langle \eta_i, \eta_i \right\rangle - \left\langle \eta_i, r_i \right\rangle^2 - \left\langle \eta_i, r_i \right\rangle \left\langle \eta_i, \xi_i \right\rangle}. \tag{A.7}$$

Compare this to the estimate if we simply project $w_{ij(t)}$ on a constant, r_i, η_i, a_{it}, and a^2_{it} :

$$\hat{\beta}_s = \frac{\left\langle r_i w^f_{ij(t)} \right\rangle \left\langle \eta_i, \eta_i \right\rangle - \left\langle \eta_i, r_i \right\rangle \left\langle \eta_i, w^f_{ij(t)} \right\rangle}{\left\langle r_i, r_i \right\rangle \left\langle \eta_i, \eta_i \right\rangle - \left\langle \eta_i, r_i \right\rangle^2}. \tag{A.8}$$

Assume $\langle \eta_i, r_i \rangle$ and $\beta_s > 0$ (this is flipping the race dummy from its interpretation in the rest of the paper, but makes the comparison much easier). If human capital η_i is correlated with within-race social capital ξ_i as well as race r_i, then the true value of β_s will be larger than an estimate which ignores ξ_i (and this is assuming we knowη_i). The intuition is that correlation between ξ_i and η_i inflates the apparent dependence of wages on human capital, resulting in an underestimate of the role of social capital.

A.3.2 Discrimination in direct hiring

Now suppose there is discrimination in direct hiring as well as in jobs found through friends, such that

$$w^d_{ij(t)} = \left(\eta_i - \kappa^d r_i\right) + \alpha_1 x_{it} + \alpha_2 x^2_{it} + \epsilon^d_{ij(t)}. \tag{A.9}$$

$$w^f_{ij(t)} = \beta_{\text{hcap}} \left(\eta_i - \kappa^f r_i\right) + \beta_{\text{exp}} \left(\alpha_1 x_{it} + \alpha_2 x^2_{it}\right) + \beta_0 + \beta_s s_i + \epsilon^f_{ij(t)}. \tag{A.10}$$

Here black workers wages of jobs found through friends are reduced by κ^d and those of jobs found through friends are reduced by κ^f. If we estimate individual fixed effects from the direct wage equation, we will be getting an estimate of $\eta_i - \kappa^d r_i$ rather than true human capital η_i. Now, notice that Equation A.10 can be rewritten

$$w^f_{ij(t)} = \beta_{\text{hcap}} \left(\eta_i - \kappa^d r_i\right) + \beta_{\text{exp}} \left(\alpha_1 x_{it} + \alpha_2 x^2_{it}\right) + \beta_0 + \beta_s s_i + \beta_{\text{hcap}} \left(\kappa^d - \kappa^f\right) r_i + \epsilon^f_{ij(t)}.$$

So if we project $w^f_{ij(t)}$ on $\eta_i - \kappa^d r_i$, the experience terms, a constant, and r_i, instead of β_s we will get as the estimated coefficient on race

$$\beta_s + \beta_{\text{hcap}} \left(\kappa^d - \kappa^f\right). \tag{A.11}$$

So if $\beta_s < 0$, $\beta_{\text{hcap}} > 0$, and there is more discrimination in direct hiring than hiring through friends ($\kappa^d > \kappa^f$), we will underestimate the magnitude of β_s.

A.4 Empirical results for Chapter 2

This section empirically establishes two stylized facts explained by our model. First, we verify correlation between degree and prosocial behavior: individuals with higher degree (more friends) are more likely to vote, and more likely to exhibit other prosocial behaviors such as blood donation, organ donor registration, and community service.

Next, we document homophily in voting behavior. Specifically, individuals who vote are more likely to have friends who vote, and partisans are more likely to have friends of the same political persuasion. Our model gives a principled reason to expect such behavior, in contrast to models of prosocial behavior in which individuals value civic behavior for its own sake. While we are unable to rule out alternative explanations for such homophily, it does seem robust to the controls available in our data.

A.4.1 Data

We use data from the Add Health survey, which began with a nationally representative sample of students in grades 7–12 in the 1994–1995 school year. Our data are from Wave III, conducted in 2001–02, when respondents were aged 18–26. Crucially, some Wave III survey respondents had an opportunity to identify other respondents (from a randomly generated list of ten names) as friends, yielding a sample of 3,572 respondents for whom we have data on both the respondent and up to ten friends. This allows us to determine the extent to which friends share similar prosocial behaviors. All results, including summary statistics, are weighted by the provided cross-sectional sampling weights. The data do not include friendship weights, so we simply assume an unweighted network.

A.4.2 Degree and prosocial behavior

In our model, individuals with insufficient social influences cannot be induced to vote. This does not necessarily imply correlation between degree (number of friends) and voting for all networks, but the logic of the model does provide an economic reason for a relationship between the two. In the data, we are unable to observe the actual degree (number of friends) of the respondents, so we use as a proxy the number of people they identify as friends when presented with a list of ten names. We then use this proxy for degree along with other relevant information to predict the likelihood that given individual will vote. Our dependent variable is self-reported voting in the most recent presidential election. In the data, individuals with higher degrees are more likely to vote: voters in our sample had an average degree of 3.4, compared to 2.9 for nonvoters. Table 1 presents the estimates of a logit regression of voting on degree and a variety of demographic controls, which confirms the robustness of the

Table 1: Logit estimates—voting

Variable	**Coefficient**	**(Std. Err.)**
Degree	0.037	(0.015)
Male	-0.016	(0.092)
Black	0.398	(0.109)
Asian	-0.407	(0.221)
Hispanic	-0.483	(0.159)
Age	-0.038	(0.055)
Intercept	0.120	(1.094)
N	3337	
Log-likelihood	-4222259.147	
$\chi^2_{(6)}$	36.651	

correlation.[1]

Furthermore, this correlation is present for other prosocial behaviors as well. Higher degree is associated with blood donation[2], organ donor registration, and community service[3]. These results are discussed in Section 2.5.

A.4.3 Homophily in prosocial behavior

In our model, friendships are more likely to be realized between individuals who behave similarly—homophily (Remark 2.3). Friendship weights are not included in the Add Health data, so we assume equal weights and test whether or not people tend to behave like their friends. Add Health survey respondents were asked whether or not they had voted in the last

[1]The set of individuals from which the names of friends are drawn contains only respondents from previous waves of the survey, so our proxy of degree is likely biased lower for those who pursue higher education. To keep this bias from affecting our results, education variables are omitted from the set of controls when degree is an explanatory variable.

[2]The survey asks, “Have you donated blood, plasma, or platelets during the last 12 months?”

[3]The survey asks, “During the last 12 months did you perform any unpaid volunteer or community service work?”

Table 2: Logit estimates—blood donation

Variable	**Coefficient**	**(Std. Err.)**
Degree	0.049	(0.020)
Male	-0.181	(0.114)
Black	-0.369	(0.158)
Asian	-0.422	(0.260)
Hispanic	-0.315	(0.191)
Age	-0.039	(0.065)
Intercept	-0.644	(1.310)
N	3337	
Log-likelihood	-4222259.147	
$\chi^2_{(6)}$	36.651	

Table 3: Logit estimates—organ donor registration

Variable	**Coefficient**	**(Std. Err.)**
Degree	0.041	(0.016)
Male	-0.508	(0.097)
Black	-0.939	(0.124)
Asian	-0.238	(0.227)
Hispanic	-0.695	(0.173)
Age	-0.084	(0.057)
Intercept	1.333	(1.144)
N	3337	
Log-likelihood	-4222259.147	
$\chi^2_{(6)}$	36.651	

Table 4: Logit estimates—community service

Variable	**Coefficient**	**(Std. Err.)**
Degree	0.078	(0.016)
Male	-0.072	(0.097)
Black	-0.339	(0.122)
Asian	0.313	(0.206)
Hispanic	-0.181	(0.162)
Age	-0.260	(0.060)
Intercept	4.171	(1.186)
N	3337	
Log-likelihood	-4222259.147	
$\chi^2_{(6)}$	36.651	

presidential election; we use this as a proxy for $|v_i|$. Voter turnout in our sample was 37%. Table 5 displays the results of a logit with voting as the dependent variable. The results confirm that voters' friends tend to vote as well.

More precisely, our model features politically like-minded individuals exhibiting homophily in voting. Unfortunately, the fraction of respondents who identified with a political party is too low to test this specific prediction directly. But since friends tend to share political affiliation (see Section A.4.4), our result unconditional on affiliation is supportive of our model. Furthermore, it raises another possibility: if an individual is unaware of his friends' specific political preferences but aware that they are likely to be correlated with his own, he will prefer to encourage turnout amongst his friends, since on average that will result in a positive net vote differential in favor of his favored party. Thus even if party affiliations are unobserved, people may support turnout amongst their friends in service to their own political leanings.

As before, similar results hold for blood donation, organ donor registration, and community service; these results are withheld for brevity's sake but are available upon request.

Table 5: Logit estimates—voting

Variable	**Coefficient**	**(Std. Err.)**
# friends voting - # friends abstaining	0.069	(0.038)
Male	0.096	(0.096)
Black	0.580	(0.118)
Asian	-0.646	(0.230)
Hispanic	-0.287	(0.163)
Years' schooling	0.173	(0.043)
Age	0.090	(0.058)
Full-time student	0.919	(0.122)
Part-time student	0.587	(0.188)
Intercept	-4.986	(1.262)

N	3334
Log-likelihood	-3966011.079
$\chi^2_{(9)}$	180.204

A.4.4 Homophily in party affiliation

Table 6 documents homophily of party affiliation in the network of friends. A respondent self-identifying as a Republican has an average friend group containing 11% fellow Republicans but just 7% Democrats[4], while a self-identified Democrat knows more than twice as many fellow Democrats as Republicans.

Table 6: Homophily in party affiliation

	Fraction of friends (%)			
	Democrats	Republicans	Other	Pop. avg. (%)
Democrats	10	4	86	16
Republicans	7	11	82	11

Sample size: 3,402

An alternative explanation for this phenomenon might be that friendships form based

[4]The balance includes those who identified with another party, refused to answer the question, or did not identify with any party.

on a correlate of party affiliation such as race or income. The list of the potential correlates available in our data is far from exhaustive, and we make no claims that these results distinguish our model from all other reasonable explanations for homophily in party affiliation. But to the potential correlates available, party homophily remains robust. Table 7 displays the estimates of a logit regression with Republican party affiliation as the dependent variable (the sample here includes only respondents who self-identified as either Republican or Democrat).

Table 7: Logit estimates—Republican Party affiliation

Variable	**Coefficient**	**(Std. Err.)**
# Rep. friends - # Dem. friends	0.256	(0.117)
Male	0.203	(0.195)
Black	-2.962	(0.324)
Asian	-1.607	(0.514)
Hispanic	-0.863	(0.349)
Years' schooling	0.103	(0.089)
Age	-0.015	(0.127)
Full-time student	0.090	(0.239)
Part-time student	-0.054	(0.405)
Intercept	-0.984	(2.610)
N	1026	
Log-likelihood	-999536.067	
$\chi^2_{(9)}$	114.148	

A.5 Individual incentives and group welfare

Depending on players' benefits of turnout g and costs of voting c, an equilibrium with voting may provide much higher aggregate utility than one without. It is even possible that a voting equilibrium with punishment Pareto dominates a non-voting equilibrium in which everyone cooperates fully in the bilateral games. This will be the case if those players who are punished in equilibrium value the provision of the public good more than the full cooperation of their friends.

Consider a party leader for a given political party who has the power to publicly announce z and thus effect a given level of turnout for that party. While a high z might raise the stakes of social punishment and increase turnout (by Proposition 2.4), a lower z would let more people off the hook, improving payoffs in the cooperation game. Crucially, in our model individuals would follow such a leader's recommendation out of pure self-interest, simply using the announced punishment level z to coordinate strategies.

How would such a leader behave? This question has already been studied, and the results match well with the observed aggregate dynamics of voting. Shachar and Nalebuff (1999) develop and estimate a model in which a party leader is able to induce voter turnout through costly personal effort. They show that the leader will increase effort (and thus turnout) in elections which are expected to be closer, matching empirical evidence.

However, it is not obvious why people would follow the leader's instructions. Even if the leader were simply bribing potential voters, they would have incentive to defect and stay home after receiving their bribes, since in their model the leader's efforts at generating turnout occur before the election. Our paper resolves this critical question of individual rationality. In our model, the leader's personal costs can be interpreted as the costs of communicating z to scattered, disconnected communities, say by canvassing. Individuals will follow the leader's recommendation because they expect others to, and risk social punishment otherwise.

Another interpretation is that the costs of publicly announcing z to all are negligible, but that the leader is benevolent and simply internalizes the group's costs of voting and benefits of victory. She will consider the benefits of victory as well as individuals' costs of getting to the polls and social costs of punishment, and announce a z that maximizes group expected utility.

Our model does not, however, literally require a leader to coordinate voters' intensity of punishment z. As long as like-minded individuals share a common expectation of relative party strength (say, by reading the newspaper) and a common expectation that higher punishment is expected in closer or more important elections, they will behave as if led by a benevolent planner. In this vein, Feddersen and Sandroni (2006) develop a two-party model of voting behavior in which some agents' "preferences over social outcomes reflect the choices the agent would make *if* he were a social planner," in the spirit of Harsanyi's rule-based utilitarianism. These "ethical" agents weigh the increased social cost of higher turnout against the chance of victory in the election, as would our social planner. Their model also produces an inverse correlation between margin of victory and voter turnout.[5,6]

These important papers establish that in the context of voting many people seem to behave as if they internalize the concerns of their group or the concerns of a leader, and our model provides a mechanism which aligns individual incentives with group welfare. Our work is thus consistent with both the leader-follower approach and the group-utilitarian approach, and the stylized facts explained by these models.

A.6 Non-anonymous strategies

Up until now, we have only allowed i's treatment of j to depend on j's action in period 1, but not on j's identity, network position, etc. The appeal of such anonymous strategies is simplicity—the level of information required is low, and the equality of treatment matches the

[5] Their model also predicts turnout increasing in the importance of the election and decreasing in voting costs.

[6] Feddersen and Sandroni (2006) predict different optimal turnout rates for the majority and minority parties; this could be achieved in our model by having different punishment levels for the different parties, as long as there were one publicly accepted level of punishment for interparty punishment.

universality of a social norm. Nevertheless, we have seen that such strategies are sufficiently rich to engender a variety of interesting phenomena.

However, anonymous strategies may not be sufficient to achieve the welfare-optimal equilibrium. Since anonymous strategies require punishment on the equilibrium path, ad hoc forgiveness for some players may improve aggregate welfare. No punishment occurs on the equilibrium path in the welfare-optimal equilibrium.

Proof. Let S be the strategy profile in an equilibrium with associated period-1 actions v and period-2 actions ρ. Assume that i punishes j in this profile: $\rho_{ij} = \rho_{ji} < 1$. Let S' be a strategy profile identical to S except that it prescribes full cooperation between i and j (that is, $\rho_{ij} = \rho_{ji} = 1$) if actions v are taken in period 1. This allocation clearly has higher welfare, since all actions have remained the same but the welfare of i and j has increased. Furthermore, it is an equilibrium, since v_i and v_j have only become more attractive, and full cooperation is a mutual best response in period 2. Thus S is not the welfare-optimal equilibrium. □

Ad hoc forgiveness is not merely academic. In 2010, the Affordable Care Act was signed into law in the United States. Every Republican in the House of Representatives voted against the bill, as did thirty-four of the 253 Democrats. Speculation ran that some of these dissenting, but nonpivotal, Democrats had been permitted to break party ranks to improve the chance of reelection in districts opposed to the ACA. Ben Chandler, then a Democratic representative from Kentucky, said "I'm quite confident that had I voted for the Affordable Care Act, I would have lost in 2010" (Haberkorn (2014)). This kind of scenario can be accommodated in our model by the use of non-anonymous strategies. A liberal might find himself torn between strategies that will result in either alienating political allies or opponents. When alienating one's allies is a best response in anonymous strategies, there could be a welfare improvement were such a player given special treatment and forgiven for his breach of orthodoxy. The following strategy dictates that all is forgiven if some target voting or contribution level y is attained. However this strategy fails to satisfy anonymity, per Definition 2, because a player's treatment of a friend j will depend on $\sum_{i \in I} v_i$ (and thus potentially the actions of other players), instead of simply θ_i, θ_j, and $\hat{v}_j$.

Given an integer y, consider the following strategy $\hat{s}^\star(y)$:

1. Vote for your party.
2. If $\sum_{i \in I} v_i = y$, cooperate fully with all friends. If not, cooperate with no one ($\rho_{ij} = 0$ for all $i, j : a_{ij} > 0$).

Let the strategy $\hat{s}^0$ be identical, except with abstention in the first period. While more general than the anonymous strategies considered throughout the paper, these are still fairly parsimonious in terms of information. Each player simply needs to know whether she's supposed to vote, and what the net vote differential is supposed to be. These may not be reasonable for a general election, but are plausible for the case of the House of Representatives, in which party leaders may be able to grant individual exceptions to those whose reelection depends on occasional departure from the party line.

Proposition A.1 establishes that the welfare-optimal equilibrium can be achieved using only these two strategies. Other strategies which utilize richer information about the network structure, etc. are certainly conceivable, but they cannot provide higher aggregate utility in equilibrium.

Proposition A.1. Let v, ρ be the actions taken in a given welfare-optimal equilibrium. There exists a subgame perfect equilibrium with identical actions in which $s_i \in \{\hat{s}^\star, \hat{s}^0\}$ for all $i \in I$.

Proof. By Lemma A.6, there is no punishment in the welfare-optimal equilibrium. Let $v^\star$

be the vector of period-1 actions taken in this equilibrium with $y^\star \equiv \sum_{i \in I} v_i^\star$. Consider the strategy profile S, in which everyone who votes in $v^\star$ plays $\hat{s}^\star(y^\star)$ and all others play $\hat{s}^0(y^\star)$. Clearly S effects the same actions as the welfare-optimal equilibrium; we now show that S is an equilibrium.

By assumption, there is an equilibrium $\tilde{S}$ with action profile (v, ρ), all players best responding. We show that if all players best respond according to $\tilde{S}$, they also best respond according to S. A player who chooses $v_i = 0$ according to S attains the highest utility possible given others' strategies, $U_i(S) = \tilde{f}(A_i e) + g_i(v)$, so we know they will have no profitable deviation. A player who chooses $v_i \neq 0$, will have no profitable deviation to some s' with $v_i' = -v_i$,

$$U_i(S) - U_i(s_i', S_{-i}) \geq \tilde{f}(A_i e) - \tilde{f}(0) \geq 0.$$

Now we check if there can be a profitable deviation to $v_i' = 0$. This saves i the cost of voting, but she sacrifices the full benefit of friendship in the second period,

$$U_i(S) - U_i(s_i', S_{-i}) \geq -c + \tilde{f}(A_i e) - \tilde{f}(0).$$

By assumption, there is no profitable deviation to abstention for i under $\tilde{S}$ — that is, $U_i(\tilde{S}) - U_i(s_i', \tilde{S}_{-i}) \geq 0$. Note that $U_i(S) - U_i(s_i', S_{-i})$ maximizes the utility difference between a voting strategy and a non-voting strategy, by ensuring full cooperation in the second period when i votes and zero cooperation when i deviates to not voting. Thus, if $U_i(\tilde{S}) - U_i(s_i', \tilde{S}_{-i}) \geq 0$, then $U_i(S) - U_i(s_i', S_{-i}) \geq -c + \tilde{f}(A_i e) - \tilde{f}(0) \geq 0$.

Finally, all players are best responding in the second period according to S, because $\rho_{ij} = \rho_{ji}$ in any subgame for any $i, j \in I$. By Lemma 2.3, this is a best response. We've now shown that no player has a profitable deviation, so S is an equilibrium.

□

In previous sections, we assumed that a player can only observe the decisions of their friends. Substituting one informational requirement for another, the strategies introduced above do not require a player to observe any individual decisions, but only an aggregate statistic.

If a lack of information makes personal responsibility impossible, justice requires penal communion. Even without these information constraints, many groups, especially sports teams, use a similar punishment scheme. For example, a college football team may be forced to run when one player skips a class.[7]

A.7 Graph spectra and equilibrium properties

Several important contributions to the study of games on networks use properties of the eigenvalues computed from the adjacency matrix A. Elliott and Golub (2015) characterize the Pareto-optimal action profile of a public goods game in terms of the highest eigenvalue and Bramoullé et al. (2014) show the importance of the lowest eigenvalue in a game of substitutes. In our setting, however, the spectrum of a graph does not determine the set of equilibria. Figure 13 shows two graphs which are isospectral (sharing the same multiset of eigenvalues) but not isomorphic. Assuming all nonzero capacities are one, the graph on the right has an equilibrium with voting (shown) but the one on the left does not, assuming as before common costs of $\frac{3}{2}$. Additionally, the eigenvectors of a matrix will not be sufficient to determine the set of equilibria. We can scale the capacities without changing the eigenvectors of a matrix.

[7] This was part of former head coach Charlie Strong's policy at the University of Texas.

Figure 13: Two isospectral graphs

Two isospectral graphs with qualitatively different equilibria. Filled-in nodes indicate players who are voting, and hollow nodes indicate abstention.

So, voting equilibria in a given network can be eliminated by scaling down the capacities to be sufficiently below the cost of voting.

A.8 Herding in crime

This model can be used to explain herding behavior in crime, such as looting, or speeding on the highway. Consider the latter example, with the agents cast as motorists and the principal in the role of highway patrol officer. Let μ_t represent the number of drivers who, by chance, choose to traverse a certain stretch of highway in period t. Those drivers travelling the highway (those for whom $\mu_{it} = 1$) can either drive slowly (effort) or quickly (shirk). Driving slowly is more costly (c) as it takes longer, but allows a driver to see (most of the time, $1 - \epsilon$ percent) whether the patrol officer is there ($\chi_t = 0$) or not ($\chi_t = 1$). If the patrol officer is there, she'd prefer to punish the driver. The officer may have an extremely convex punishment function $r(\cdot)$, since she may only be able to pull over a fixed number of drivers per period, $\tilde{\mu}$. So a driver may obey the speed limit unless he finds himself driving with a large pack of other drivers ($\mu > \tilde{\mu}$), at which point the chance of getting pulled over decreases and he may choose to speed.

In this application, it is perhaps unrealistic to assume that the patrol officer will not punish any drivers when they are numerous ($\mu > \tilde{\mu}$). Because it is her job (or more cynically, to fulfill a quota), she may still pull over as many speeding drivers as she can. Note, however, that as long as she pulls over speeders randomly, this will not induce any change of behavior on the part of the drivers, since any one driver's expected chance of punishment will be less than $\underline{\gamma}$ when $\mu > \tilde{\mu}$.

Policy implications

Speeding drivers have an important characteristic not included in this paper's model: they are ordered. And crucially, if the first driver in a pack of speeders slows down, he ceases to be first. So the patrol officer only needs to commit to pulling over the first car in any speeding pack; this will motivate the first driver to slow down until he is no longer the first. Then the second driver will become the first and do the same, and so on, until no one is speeding.

This example highlights how coordinated shirking can be avoided if given any set of shirkers, one is somehow salient. Even when the principal is unable to credibly punish a large group, she can credibly threaten to always punish the salient individual. As long as everyone knows who the salient individual is for any possible subset of agents and the principal is able to condition punishment on the characteristic that distinguishes him, this will unravel groups of shirkers and allow for a full-effort equilibrium. The easiest way to do this is one that arises naturally in the speeding example: a complete, strict order over all agents.[8]

[8]Technically, all that is needed is a map $f(\cdot)$ from the power set of agents to agents such that $f(a) \in a$; a complete strict order is simply an easy example.

Bibliography

(2013). EU Bonus Cap Update — Cap to Apply More Broadly. FS Regulatory Brief, PricewaterhouseCoopers LLP.

Abbink, K., Gangadharan, L., Handfield, T., and Thrasher, J. (2017). Peer punishment promotes enforcement of bad social norms. *Nature Communications*, 8(1):609.

Abrams, S., Iversen, T., and Soskice, D. (2011). Informal social networks and rational voting. *British Journal of Political Science*, 41(02):229–257.

Acharya, V. V. and Richardson, M. (2009). Causes of the financial crisis. *Critical Review*, 21(2-3):195–210.

Acharya, V. V. and Yorulmazer, T. (2007). Too many to fail — An analysis of time-inconsistency in bank closure policies. *Journal of Financial Intermediation*, 16(1):1–31.

Acharya, V. V. and Yorulmazer, T. (2008). Cash-in-the-Market Pricing and Optimal Resolution of Bank Failures. *Review of Financial Studies*, 21(6):2705–2742.

Aizer, A. and Currie, J. (2017). Lead and Juvenile Delinquency: New Evidence from Linked Birth, School and Juvenile Detention Records. Working Paper 23392, National Bureau of Economic Research. DOI: 10.3386/w23392.

Ali, S. N. and Bénabou, R. (2016). Image versus information: Changing societal norms and optimal privacy. Technical report, National Bureau of Economic Research.

Ali, S. N. and Lin, C. (2013). Why people vote: Ethical motives and social incentives. *American Economic Journal: Microeconomics*, 5(2):73–98.

Altonji, J. G. and Blank, R. M. (1999). Race and Gender in the Labor Market. *Handbook of Labor Economics*, 3:3143–3259.

Ambrus, A., Mobius, M., and Szeidl, A. (2014). Consumption risk-sharing in social networks. *American Economic Review*, 104(1):149–82.

Arrow, K. J. and Borzekowski, R. (2004). Limited Network Connections and the Distribution of Wages. FEDS Working Paper 2004-41, Board of Governors of the Federal Reserve System.

Becker, G. S. (1968). Crime and Punishment: An Economic Approach. *Journal of Political Economy*, 76(2):169.

Becker, G. S. and Tomes, N. (1979). An Equilibrium Theory of the Distribution of Income and Intergenerational Mobility. *Journal of Political Economy*, 87(6):1153–1189.

Bernstein, R., Chadha, A., and Montjoy, R. (2001). Overreporting voting: Why it happens and why it matters. *Public Opinion Quarterly*, 65(1):22–44.

Bertrand, M. and Mullainathan, S. (2004). Are Emily and Greg More Employable Than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination. *American Economic Review*, 94(4):991–1013.

Blais, A. (2000). *To vote or not to vote?: The merits and limits of rational choice theory*. University of Pittsburgh Pre.

Borgers, T. (2004). Costly voting. *The American Economic Review*, 94(1):57–66.

Bramoullé, Y., Kranton, R., and D'amours, M. (2014). Strategic interaction and networks. *The American Economic Review*, 104(3):898–930.

Brown, M., Setren, E., and Topa, G. (2015). Do Informal Referrals Lead to Better Matches? Evidence from a Firm's Employee Referral System. *Journal of Labor Economics*, 34(1):161–209.

Burdett, K. (1978). A Theory of Employee Job Search and Quit Rates. *The American Economic Review*, 68(1):212–220.

Calvó-Armengol, A. and Jackson, M. O. (2004). The Effects of Social Networks on Employment and Inequality. *The American Economic Review*, 94(3):426–454.

Chandrasekhar, A. G., Larreguy, H., and Xandri, J. P. (2015). Testing models of social learning on networks: Evidence from a lab experiment in the field.

Cialdini, R. B., Demaine, L. J., Sagarin, B. J., Barrett, D. W., Rhoads, K., and Winter, P. L. (2006). Managing social norms for persuasive impact. *Social influence*, 1(1):3–15.

Cialdini, R. B., Reno, R. R., and Kallgren, C. A. (1990). A focus theory of normative conduct: recycling the concept of norms to reduce littering in public places. *Journal of personality and social psychology*, 58(6):1015.

Coate, S. and Loury, G. C. (1993). Will Affirmative-Action Policies Eliminate Negative Stereotypes? *The American Economic Review*, 83(5):1220–1240.

DellaVigna, S., List, J. A., Malmendier, U., and Rao, G. (2014). Voting to tell others. Technical report, National Bureau of Economic Research.

Devenow, A. and Welch, I. (1996). Rational herding in financial economics. *European Economic Review*, 40(3-5):603–615.

DiPasquale, D. and Glaeser, E. L. (1998). The Los Angeles Riot and the Economics of Urban Unrest. *Journal of Urban Economics*, 43(1):52–78.

Downs, A. (1957). An economic theory of voting.

Durlauf, S. N. (1996). A theory of persistent income inequality. *Journal of Economic Growth*, 1(1):75–93.

Elliott, M. and Golub, B. (2015). A network approach to public goods. *Available at SSRN 2436683*.

Fang, H. and Moro, A. (2011). Theories of Statistical Discrimination and Affirmative Action: A Survey. In Benhabib, J., Jackson, M. O., and Bisin, A., editors, *Handbook of*

Social Economics, volume 1A, pages 133–200. Elsevier.

Farhi, E. and Tirole, J. (2012). Collective Moral Hazard, Maturity Mismatch, and Systemic Bailouts. *American Economic Review*, 102(1):60–93.

Feddersen, T. and Sandroni, A. (2006). A theory of participation in elections. *The American Economic Review*, 96(4):pp. 1271–1282.

Feddersen, T. J. (2004). Rational choice theory and the paradox of not voting. *Journal of Economic Perspectives*, pages 99–112.

Fehr, E. and Fischbacher, U. (2004). Social norms and human cooperation. *Trends in cognitive sciences*, 8(4):185–190.

Frey, B. S. and Meier, S. (2004). Social comparisons and pro-social behavior: Testing" conditional cooperation" in a field experiment. *The American Economic Review*, 94(5):1717–1722.

Funk, P. (2010). Social incentives and voter turnout: evidence from the swiss mail ballot system. *Journal of the European Economic Association*, 8(5):1077–1103.

Gerber, A. S., Green, D. P., and Larimer, C. W. (2008). Social pressure and voter turnout: Evidence from a large-scale field experiment. *American Political Science Review*, 102(01):33–48.

Golub, B. and Jackson, M. O. (2010). Naive learning in social networks and the wisdom of crowds. *American Economic Journal: Microeconomics*, pages 112–149.

Granovetter, M. S. (1973). The Strength of Weak Ties. *American Journal of Sociology*, 78(6):1360–1380.

Green, G. P., Tigges, L. M., and Diaz, D. (1999). Racial and Ethnic Differences in Job-Search Strategies in Atlanta, Boston, and Los Angeles. *Social Science Quarterly*, 80(2):263–278.

Haberkorn, J. (2014). Just 4 anti-aca house dems left. *Politico*.

Hall, R. E. (2010). Why Does the Economy Fall to Pieces after a Financial Crisis? *The Journal of Economic Perspectives*, 24(4):3–20.

Harbaugh, W. T. (1996). If people vote because they like to, then why do so many of them lie? *Public Choice*, 89(1-2):63–76.

Holzer, H. J. (1987). Informal Job Search and Black Youth Unemployment. *The American Economic Review*, 77(3):446–452.

Ioannides, Y. M. and Datcher Loury, L. (2004). Job Information Networks, Neighborhood Effects, and Inequality. *Journal of Economic Literature*, 42(4):1056–1093.

Jackson, M. O. and Zenou, Y. (2014). Games on networks. *Handbook of Game Theory*, 4.

Karlan, D., Mobius, M., Rosenblat, T., and Szeidl, A. (2009). Trust and social collateral. *The Quarterly Journal of Economics*, 124(3):1307–1361.

Kets, W. and Sandroni, A. (2015). A belief-based theory of homophily. *Working paper*.

Knack, S. (1992). Civic norms, social sanctions, and voter turnout. *Rationality and Society*, 4(2):133–156.

Knoke, D. (1990). Networks of political action: Toward theory construction. *Social Forces*, 68(4):1041–1063.

Levine, D. K. and Mattozzi, A. (2017). Voter turnout with peer punishment. Technical report, David K. Levine.

Levine, D. K. and Modica, S. (2014). Peer discipline and incentives within groups. *EUI working paper*.

Levine, D. K. and Modica, S. (2016). Peer discipline and incentives within groups. *Journal of Economic Behavior & Organization*, 123:19–30.

Loury, G. C. (1981). Intergenerational Transfers and the Distribution of Earnings. *Econometrica*, 49(4):843–867.

Loury, L. D. (2006). Some Contacts Are More Equal than Others: Informal Networks, Job Tenure, and Wages. *Journal of Labor Economics*, 24(2):299–318.

MacKenzie, D. and Spears, T. (2012). 'The Formula That Killed Wall Street'? The Gaussian Copula and the Material Cultures of Modelling. *Working Paper*.

Miller, C. (2017). The Persistent Effect of Temporary Affirmative Action. *American Economic Journal: Applied Economics*, 9(3):152–190.

Montgomery, J. D. (1991). Social Networks and Labor-Market Outcomes: Toward an Economic Analysis. *The American Economic Review*, 81(5):1408–1418.

Montgomery, J. D. (1992). Social Networks and Persistent Inequality in the Labor Market. *Working Paper*.

Mossel, E. and Tamuz, O. (2014). Opinion exchange dynamics. *arXiv preprint arXiv:1401.4770*.

Neal, D. A. and Johnson, W. R. (1996). The Role of Premarket Factors in Black-White Wage Differences. *Journal of Political Economy*, 104(5):869–895.

O'Neill, J. (1990). The Role of Human Capital in Earnings Differences Between Black and White Men. *The Journal of Economic Perspectives*, 4(4):25–45.

Panagopoulos, C. (2010). Affect, social pressure and prosocial motivation: Field experimental evidence of the mobilizing effects of pride, shame and publicizing voting behavior. *Political Behavior*, 32(3):369–386.

Perez-Truglia, R. and Cruces, G. (2016). Partisan interactions: Evidence from a field experiment in the united states. *Journal of Political Economy*.

Rajan, R. (2008). Bankers' pay is deeply flawed. *Financial Times*, 8.

Schmutte, I. M. (2015). Job Referral Networks and the Determination of Earnings in Local Labor Markets. *Journal of Labor Economics*, 33(1):1–32.

Schmutte, I. M. (2016). Labor markets with endogenous job referral networks: Theory and empirical evidence. *Labour Economics*, 42:30–42.

Shachar, R. and Nalebuff, B. (1999). Follow the leader: Theory and evidence on political participation. *American Economic Review*, pages 525–547.

Sibona, C. (2014). Unfriending on facebook: Context collapse and unfriending behaviors. In *System Sciences (HICSS), 2014 47th Hawaii International Conference on*, pages 1676–1685. IEEE.

Simon, C. J. and Warner, J. T. (1992). Matchmaker, Matchmaker: The Effect of Old Boy Networks on Job Match Quality, Earnings, and Tenure. *Journal of Labor Economics*, 10(3):306–330.

Taylor, J. B. (2009). The Financial Crisis and the Policy Responses: An Empirical Analysis of What Went Wrong. Working Paper 14631, National Bureau of Economic Research.

The Pew Research Center For The People & The Press (2006). Regular voters, intermittent voters, and those who don't: Who votes, who doesn't, and why. Technical report.

Wolpin, K. I. (1992). The Determinants of Black-White Differences in Early Employment Careers: Search, Layoffs, Quits, and Endogenous Wage Growth. *Journal of Political Economy*, 100(3):535–560.