UNIVERSITY OF CALIFORNIA
Los Angeles

Thesis on Social Networks

A dissertation submitted in partial satisfaction
of the requirements for the degree
Doctor of Philosophy in Economics

by

Jingyu Fan

2023

ABSTRACT OF THE DISSERTATION

Thesis on Social Networks

by

Jingyu Fan
Doctor of Philosophy in Economics
University of California, Los Angeles, 2023
Professor Moritz Meyer-ter-Vehn, Chair

This dissertation consists of three independent chapters that discuss social networks (defined in the broad sense) in different angles.

Chapter 1 studies how an agent's propensity to accept bribes depends on the organizational structure, which we model with a broad set of random networks that contains two canonical special cases. In *hierarchies*, agents' best responses exhibit *strategic substitutability*, with bribe taking being risker if others accept more bribes, for it is then easier for a corruption investigation to trace through bribe transactions to locate bribe takers. On the contrary, best responses in flat, *two-layer networks* feature *strategic complementarity*, as more bribe acceptances better protect criminal subordinates from being caught, reducing the risk of bribe taking. While incentives differ across networks, we show that for any of our random network, in equilibrium, increasing its density always deters agents from accepting bribes. Nevertheless, opposite results for hierarchies and two-layer networks are obtained if we make the number of subordinates each agent monitors more evenly distributed. We use this model to point out a corruption identification problem and propose a remedy to it.

Chapter 2 studies how parochial fairness concerns – a player's incentives to compare wages with those in the same group – affect group deviations. We propose a new theoretical framework based on the transferable-utility cooperative game through extending the

utility space to incorporate in players' other-regarding incentives. We then apply in the Fehr-Schmidt utility specification to study how parochial fairness concerns govern income redistribution outcomes after coalitional deviations and the structures of core allocations. We find that while both disadvantageous and advantageous inequality aversion exacerbate income inequality after a coalition deviates, advantageous inequality inclination ameliorates it. In addition, if players are moderately averse to advantageous inequality, the grand-coalition allocation most robust to coalitional deviations is the "tyranny-of-the-majority" allocation that gives the single poorest player indefinitely small amount and equates the other players' incomes.

Chapter 3 studies the rewards-based crowdfunding industry. We ask how a *creator* of a crowdfunding project optimally designs the pricing strategies for rewards to maximize the total fund raised. We build a structural model based on [Bre87] and estimate the project and reward values and the distribution of *backers'* preferences with data on reward prices and the numbers of backers buying each reward. We find that backers' preference distribution heavily weighs towards the lower side – most backers are of low types; and that a creator optimally employs a bow-shaped pricing strategy – they extract most of the surplus from low- and high-type backers and charge those in the middle little to no premiums.

The dissertation of Jingyu Fan is approved.

Simon Adrian Board

Tomasz Marek Sadzik

Robert Zeithammer

Moritz Meyer-ter-Vehn, Committee Chair

University of California, Los Angeles

2023

To my friends and colleagues
who supported me emotionally
during the hardest time

CONTENTS

LIST OF FIGURES

ACKNOWLEDGMENTS

First of all, I would like to thank my main advisor and committee chair, Prof. Moritz Meyer-ter-Vehn. This dissertation would not be here without his insightful and meticulous advice over the past few years. I am also grateful to my committee member Prof. Simon Board for his tremendous help in revising the dissertation and spotting potential research directions.

I thank my other committee members, Prof. Tomasz Sadzik and Prof. Robert Zeithammer, for reviewing my dissertation and providing literature suggestions, as well as other professors in my department, Prof. Alexander W Bloedel, Prof. Jay Lu, Prof. Maurizio Mazzocco, Prof. Ichiro Obara, and Prof. William R Zame (named in alphabetical order by last name), for regularly attending my presentations and providing useful advice.

At different times, various individuals have left comments that contribute substantially to my research, including Prof. Attila Ambrus, Prof. Alessandro Pavan, and Prof. Yuichi Yamamoto (named in alphabetical order by last name). I would like to express my gratitudes for their support.

Lastly, I thank the participants at the LBS Trans-Atlantic Doctoral Conference, the Conference on Mechanism and Institution Design, the Stony Brook International Conference on Game Theory, and the European Winter Meeting of the Econometric Society, for sitting through my talks and helping with the dissemination of my idea.

CURRICULUM VITAE

2017	B.A. (Joint Major in Economics and Mathematics), New York University, New York, New York.
2018	M.A. (Economics), UCLA, Los Angeles, California.
2018–present	Teaching Assistant, Department of Economics, UCLA, Los Angeles, California.

PRESENTATIONS

Trans-Atlantic Doctoral Conference, London, UK (virtual), June 2022.

Conference on Mechanism and Institution Design, Singapore (virtual), July 2022.

33rd Stony Brook International Conference on Game Theory, Stony Brook, New York, July 2022.

European Winter Meeting of the Econometric Society, Berlin, Germany, December 2022.

CHAPTER 1

Corruption Networks

1.1 Introduction

Corruption is a huge problem. Apart from wasting public resources, it hinders crime detection, as oftentimes corruptible monitors are bribed into underreporting criminal behaviors. Bribe exchanges of this type occur systematically in social networks. For instance, consider the uncovering of rampant corruption in the New York Police Department (NYPD) by the Knapp Commission in the 1970s. Immediately after the police officer Frank Serpico published a New York Times article revealing the NYPD's corruption, the Knapp Commission was established to investigate into it. Within several years, it gathered testimonies from dozens of witnesses (some were corrupt themselves), and implicated policemen from all layers accustomed to receive favors from criminals and lower-rank officers, while also covering up crimes and each other's bribe taking routine.[1] This story implies that in large bureaucracies, corruption is systematically organized into networks.[2]

This paper studies an individual's incentive to take bribes, and how it is jointly determined by the network structure and the general corruption level. Our contribution is to study a common social behavior in a novel angle, and to propose a new game form played

[1]See Wikipedia (`https://en.wikipedia.org/wiki/FrankSerpico` and `https://en.wikipedia.org/wiki/Knapp_Commission`) for a detailed description of this series of events.

[2]There exist rich documentations on police corruption (see, for instance, [Pun00], [Pun09] and [Ver99]), which is both systematic and recurrent. The above example was by no means the only corruption scandal on the NYPD – comparable scandals broke out every 20 years (1895, 1913, 1932, 1954, 1973 and 1994). Perhaps not surprisingly, 20 years after the 1970s incident, another investigative commission – the Mollen Commission – was summoned only to discover that pervasive corruption in the NYPD was regenerated in novel forms [Pun00].

on a large set of networks.

We consider a group of agents connected by a directed *monitoring network*, each link pointing from a *supervisor* to a *subordinate*. The monitoring network can be any large, random network with finite components (which implies acyclicity). All agents are *criminal* and offer bribes to their supervisors.[3] A supervisor can costlessly verify if her subordinates are criminal and, upon observing a criminal subordinate, decides whether to accept the subordinate's bribe offer and in return withhold the crime report. All bribe acceptance decisions are made simultaneously. An agent is *caught* by an external law enforcement agency if she is directly detected by it to be criminal, reported by a supervisor, or discovered through accepting a subordinate's bribe.

Since a caught agent is obliged to pay a fixed punishment cost and surrender all accepted bribes, while each additional bribe acceptance brings in constant marginal value, it imposes increasing marginal cost. An agent thus employs a cutoff bribe acceptance strategy, that is, accepting bribes up to an optimal desired number of bribes.

The structure of the monitoring network is the key to how others' bribe acceptance behaviors translate into an agent's bribe taking risk, thereby determining her bribe acceptance propensity. In the first result, we analyze two canonical network structures and demonstrate that they offer contrasting implications on that (Proposition 1.1).

One is *hierarchies* (the left panel of figure 1.1), where each agent has at most one supervisor, and so there is no co-supervision. In a hierarchy, agents' best responses exhibit *strategic substitutability*, i.e., a rise in others' corrupt behaviors renders bribe taking riskier and thus less appealing to an agent. Intuitively, an agent's bribe taking risk is only affected by the actions of her direct and indirect subordinates. The more bribes they accept, the easier it is for a corruption investigation to percolate up to reach the agent, should one lower-rank opponent be directly caught as criminal. Hence, the agent accepts fewer bribes to mitigate the higher risk. Decreasing best responses also imply the existence of a unique equilibrium.

[3]Later, we will consider an extended model where some agents are criminal, and some are *innocent*.

The other is *two-layer networks* (the right panel of figure 1.1), where the agents are divided into pure supervisors and pure subordinates, thus a supervisor has no indirect subordinates. As opposed to a hierarchy, in a two-layer network, agents' best responses feature *strategic complementarity*. This is because now an agent's bribe taking risk is determined by her co-supervisors' actions. More corrupt co-supervisors tend to cover up subordinates' crimes, creating a safer environment for the agent to accept bribes. Increasing best responses[4] sometimes lead to multiple equilibria – both high and low corruption levels can be sustainable.[5]

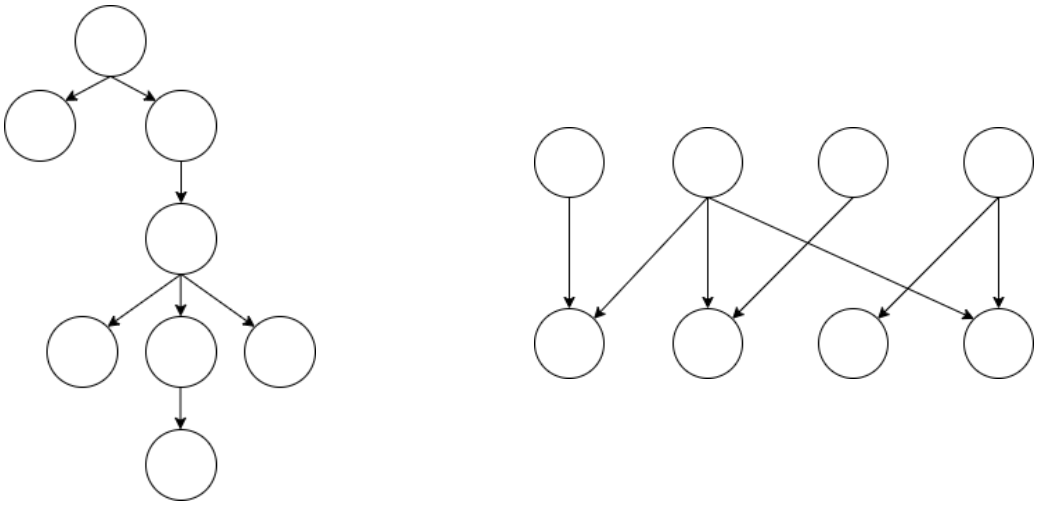


Figure 1.1: Hierarchy (Left) and Two-Layer Network (Right)

The following case helps understand how these results fit into reality:

In November 2011, a robbery at Peizhong Bai's villa accidentally uncovered that this chairman of Shanxi Coking Coal Group had embezzled ¥50 million. The case was originally closed in haste. However, years later, it was dug out by the Standing Com-

[4]Though the comparison between strategic substitutability and complementarity is presented here with best response variations, in a model extension with *incorruptible* agent types, it manifests itself in equilibrium in that depending on the network structure, *corruptible* agents' propensity to accept bribes rises or falls with the proportion of incorruptible agents (Section 1.5.1, Corollary 1.1).

[5]This outcome accords with reality: various studies on police corruption point out that in a generally corrupt environment, a new police officer would be easily coaxed into conforming to the corrupt norm. Thus corruption maintains prevalent over the long run. However, in a generally incorrupt environment, threatened by the high exposure rate, he would be deterred from corruption at the first place. Hence integrity sustains.

mittee of the National People's Congress (NPCSC) and prosecuted again. This time several officials and businessmen involved were arrested, among whom were four police officers from the local Public Security Bureau (PSB) initially in charge of the robbery case, including even the Deputy Chief Laiwei Dai. These police officers were accused of underreporting Bai's amount of loss in the robbery in exchange for bribes counted in millions. Notably, one of them, Yongping Li, also bribed his superior Dai [KW16].

One can identify two monitoring networks in this example. First, since the PSB and the NPCSC separately monitor Bai, their relationship forms a two-layer network. The PSB's corruption was detected because the NPCSC reported Bai truthfully (instead of taking a bribe), indicating strategic complementarity between their corrupt behaviors. Another network is the hierarchy consisting of the Deputy Chief Dai, his subordinate Li, and the criminal Bai that Li monitors. Dai's outcome implies strategic substitutability, as he was investigated and caught as corrupt only because Li was caught first through taking Bai's bribe. This case also suggests the prevalence of the two network structures in social organizations: hierarchies are standard structures that organize individuals within an institution; two-layer networks naturally arise among different institutions and divisions where co-supervision is common.

Despite the different incentives across networks, our second result shows that, given any monitoring network, increasing its density always deters agents from accepting bribes in equilibrium[6] (Proposition 1.3). We argue that the intuitions however vary for a hierarchy and a two-layer network. When a hierarchical monitoring network becomes denser, an agent has more subordinates and hence receives more bribe offers, enabling her to accept more bribes. Since corruption in hierarchies features strategic substitutability, it makes bribe taking riskier and thus less appealing. As for a denser two-layer network, since a subordinate is monitored by more supervisors, each receiving more bribe offers and so accepting a smaller proportion of them, she is more vulnerable to crime detection. Therefore, bribe taking carries higher risk and is thus less attractive.

[6]Wherever multiple equilibria exist, we select the most corrupt equilibrium for comparative statics, which is Pareto optimal.

While changes in the network density – first-order variations in the network degree distributions – lead to consistent results across hierarchies and two-layer networks, opposite outcomes for them are brought about by second-order variations in agents' out-degree distributions (Proposition 1.4).

In hierarchies, agents' propensity to accept bribes falls given a more even distribution of the number of subordinates each agent monitors. In particular, agents are the most deterred from accepting bribes when they are arranged into a *linear hierarchy*, where each agent monitors exactly one subordinate (except for the bottom agent). To see that, notice that the risk of being caught only transmits from an agent to her direct and indirect supervisors, and not those in the same tier. Hence, by positioning only one agent in each tier and in turn elongating the hierarchy, the bribe taking risk is maximized.

In two-layer networks, agents are less inclined to accept bribes when the number of subordinates each supervisor monitors is, on the contrary, less evenly distributed. Intuitively, while some agents now receive fewer bribe offers and so are more constrained in bribe taking, others receive more and so accept a smaller proportion of them. Overall, fewer bribe offers are accepted, indicating larger likelihood a subordinate is reported. Hence, bribe taking becomes riskier and so less appealing.

This model also has empirical implications. Specifically, we argue that the *corruption measure* defined based on the number of bribe taking cases detected by law enforcements is non-monotonic in and thus an unreliable indicator of the actual *corruption level* (Proposition 1.5). It is because the *corruption detection rate* is endogenous and depends on agents' bribe acceptance decisions, thus rampant corruption may accompany weak detection, leading to meagre observed corruption cases (likewise, the opposite can also happen). This discovery casts doubt on the validity of corruption measures constructed with law enforcement data in evaluating anti-corruption policies. For instance, employing such measures, mixed empirical results are obtained as to whether higher salaries for public officials curb or aggravate corruption.[7]

[7]While evidence exists that corruption is alleviated if public officials are better paid, as shown by [GR89]

Lastly, we consider a model extension where some agents are noncriminal (*innocent*). Compared to a criminal agent, an innocent agent is more reluctant to accept bribes, as corruption exposes her to the risk of being caught by the external law enforcement agency. The extended model allows us to predict how the prevalence of crime affects corruption in different network structures. First, in a linear hierarchy, a rise in criminal agents lowers one's bribe acceptance propensity. This is because criminal agents are strategically more corrupt, and so their vaster presence makes a corruption investigation easier to percolate up to endanger a bribe taker. Conversely, in a two-layer network, an increase in criminal supervisors encourages agents to accept bribes, as criminal supervisors tend to be corrupt and underreport subordinates' crimes, making bribe acceptance less risky (Proposition 1.7). These results echo the contrast between strategic substitutability and complementarity.

Literature Review

This research is most closely related to two network formation papers. Both center on the role of a network in spreading the risk of detection by an external force among a group of agents, and analyze individual network formation choices in this setting.

[BB08] study optimal network formation in a terrorist organization, where each agent chooses the set of colleagues to reveal his identity to. While disclosing personal information improves group efficiency, it renders an agent more vulnerable to external threat – once the anti-terrorist agency detects an agent, it also detects those he holds information on. Our work differs from theirs in both approaches and results. First, while they consider completely free network formation, we fix the monitoring network and let the *corruption subnetwork* form endogenously through bribe taking. As for results, in their model, only small networks – singletons, binary cells or two-tier hierarchies – form to minimize risk contagion. This is consistent with our strategic substitutability in hierarchies, where a larger corruption subnetwork raises the bribe taking risk. But we also bring in the novel strategic complementarity

and [GN98] for the US and by [SSZ16] for Russia, [AL12] find no obvious relationship between corruption and public officials' salaries using the same US dataset. Moreover, in a case study on the Mississippi state, [KRS06] find that corruption is in fact more rampant when public officials' salaries rise.

in two-layer networks, where more corruption reduces the contagious reporting risk and so encourages agents to accept bribes.

[AMO16] consider the setting where a network of agents is threatened by contagious cyberattack. Though they mainly focus on how agents invest in self immunity given an exogenous network structure, one section allows them to sever their links, each generating positive payoffs yet facilitating risk contagion. Similar to [BB08], their model brings about strategic substitutability in that an agent optimally minimizes the size of the component he is attached to. But strategic complementarity is absent. Another distinction is that while their network is perfectly observable, our agents only know how many supervisors and subordinates they have, and base their decisions on network parameters. This approach allows us to compare agents' strategies across structurally different networks.

The presence of corruptible monitors is well-documented in empirical literature ([DGP13], [BBK21][8]). In theory literature, this problem is traditionally studied as collusion between two parties – a monitor and an agent. [Tir86] first recognizes it as a moral hazard problem. Following that, [KL93] and [FLM03] explore various ways a principal can design contracts to efficiently prevent such collusions. Outside the contract theory sphere, several papers discuss welfare-improving policies in face of corruptible monitors. [CW92] and [BM93] make money transfer designs to maximize tax revenues when evasive taxpayers can bribe their auditors. [MP95] conduct similar studies in the context of a polluting factory and a pollution inspector. More recently, [OC18] randomize a monitor's wage to introduce information asymmetry between the potentially collusive parties.

This paper is the first to study corruptible monitors in a systematic, structural way using networks, breaking away from the confinement to two-player interactions. We also shift attention from optimal contract and policy designs to how monitors make corrupt decisions conditional on the network structure and each other's strategy. While previous papers generally assume that the probability a monitor is caught as corrupt is exogenous,

[8][DGP13] find that pollution auditors paid by plants frequently underreport their pollution levels in India. [BBK21] observe that procurement officers in the Pakistanian government report higher prices when they are monitored.

in our paper, it is the key endogenous variable through which the network structure alters agents' corrupt incentives.

As this paper points out the non-monotonic relationship between actual corruption levels and corruption measures obtained from law enforcement data, it also contributes to the applied economics literature on corruption where such measures are extensively used. To cite a few examples, [MH92] and [GS06] adopt the number of convicted corrupt individuals to study corruption in the US; [DT13] and [Zak19] utilize the number of registered corrupt cases to measure corruption in China and Russia, respectively. Our structural approach can help obtain more accurate measures of the underlying corruption in the future.

The paper is organized as follows. Section 1.2 delineates the baseline model. Section 1.3 provides equilibrium analysis. In 1.3.1, we derive agents' best responses, and use them to discuss strategic substitutability and complementarity across different network structures. In 1.3.2, we derive the equilibria and perform comparative statics on network parameters. Section 1.4 evaluates the performance of a commonly used corruption measure and raises concerns for its validity. Section 1.5 considers two model extensions. Section 1.6 concludes.

1.2 Model

1.2.1 Primitives

A set of criminal agents are connected by a directed monitoring network, each edge representing a crime monitoring relationship pointing from the supervisor to the subordinate. Per monitoring pair (edge), the supervisor has observed the subordinate's crime, who has offered a bribe with fixed value $b > 0$ back. The supervisor chooses either to *accept* it, or to decline it and *report* the subordinate's crime to the law enforcement agency. All bribe acceptance decisions are made simultaneously.

There are three ways through which an agent can get caught: (i) being reported by at least one supervisor; (ii) being directly caught as criminal by the law enforcement agency,

which happens on each agent with independent probability $q \in (0,1)$; (iii) being detected through accepting bribes: once an agent is caught, the law enforcement agency initiates an investigation over her supervisors; any of them who has accepted the agent's bribe is detected independently with probability $\eta \in (0,1]$.

A caught agent pays a punishment cost $c > 0$ and has all the bribes confiscated, including those she accepted from her subordinates and those she offered to her supervisors but were declined[9]. Hence, if an agent with $l \geq 0$ supervisors accepts $n \geq 0$ bribes, her payoff is $(-c - lb)$ if she is caught and $(n - l)b$ if not.

All the parameters mentioned above are public knowledge, as well as the network parameters to be introduced in the next section. Regarding the monitoring network, each agent only knows her own in- and out-degrees, that is, how many supervisors and subordinates she herself has.

1.2.2 Monitoring Network

We consider a large, random monitoring network with finite components. Let π be an agent's joint distribution of in- and out-degrees with finite support. The in- and out-degree distributions are thus $\lambda = \int \pi(\cdot, k)dk$ and $\mu = \int \pi(l, \cdot)dl$. Denote by $\hat{\lambda}$ the in-degree distribution for any of an agent's subordinates, and by $\hat{\mu}$ the out-degree distribution for any of her supervisors. By definition, $\hat{\lambda}(l) = \lambda(l)l/\mathbb{E}_\lambda(\tilde{l}),\ \hat{\mu}(k) = \mu(k)k/\mathbb{E}_\mu(\tilde{k})\ \forall l, k$.[10]

The underlying network can be constructed with the configuration model [Jac08]. Yet, for the purpose of this paper, it is enough to restrict attention to one component. We use the *branching process* to generate a random component – a component we uncover through randomly picking an agent in the network:

1. Start with a node, create its l supervisors and k subordinates according to π.

[9]When a supervisor reports a subordinate, she also submits the bribe the subordinate offers to the law enforcement agency. In other words, the bribes an agent offers to her supervisors are sunk cost that never gets recovered.

[10]See [Jac08] for a detailed derivation of neighbors' degree distributions.

2. (a) If no new node is generated in the last step, terminate;

   (b) otherwise, for each new subordinate, create its $\hat{l}$ new supervisors from $(\hat{\lambda}-1)$, then create its k subordinates from $\mu(\cdot|l=\hat{l}+1)$; for each new supervisor, create its $\hat{k}$ new subordinates from $(\hat{\mu}-1)$, then create its l supervisors from $\lambda(\cdot|k=\hat{k}+1)$.[11]

3. Repeat step 2.

In this paper, we are particularly interested in hierarchies and two-layer networks (see figure 1.1). The former is constructed by specifying $\operatorname{supp}\lambda = \{0,1\}$, the latter by setting $\pi(l,k)=0$ if $l,k>0$.

Lastly, the following assumptions are imposed such that a component does not grow unbounded, and the network is well-defined.

Assumption 1.1. *i (**finiteness**)* $\mathbb{E}_{\hat{\mu}}(k-1)\cdot\mathbb{E}_{\hat{\lambda}}(l-1) < \big(1-\mathbb{E}_{\hat{\mu}}\big[\mathbb{E}_{\lambda}(l|k)\big]\big)\cdot\big(1-\mathbb{E}_{\hat{\lambda}}\big[\mathbb{E}_{\mu}(k|l)\big]\big)$ *and* $\mathbb{E}_{\hat{\mu}}\big[\mathbb{E}_{\lambda}(l|k)\big] < 1$.

*ii (**consistency**)* $\mathbb{E}_{\mu}(k) = \mathbb{E}_{\lambda}(l)$.

(i) is the necessary and sufficient condition for a random component to be finite (see Appendix 1.7.1 for the formal statement and its proof). To understand that, notice that a stronger condition is $\max\big\{\mathbb{E}_{\hat{\mu}}\big[k-1+\mathbb{E}_{\lambda}(l|k)\big], \mathbb{E}_{\hat{\lambda}}\big[l-1+\mathbb{E}_{\mu}(k|l)\big]\big\} < 1$, which says that the expected number of new links generated for each node is less than one. This assumption ensures that the branching process terminates in finitely many steps, such that the local structure of the underlying network indeed converges in distribution to the component generated by the given branching process when the network gets large [Sad20]. Since the branching process generates directed trees, it also implies that a component – and so the network in general – is acyclic, which will greatly simplify the derivation of agents' best responses. (ii) states that the network has the same expected in- and out-degrees.

[11] In this step, neighbors' degree distributions $\langle\hat{\lambda},\hat{\mu}\rangle$ are employed to adjust for the *friendship paradox*, namely, compared with a random agent in the network, a random subordinate of an agent is more likely to have more supervisors, and a random supervisor of an agent is more likely to have more subordinates. For a more rigorous treatment on that, see [Sad20].

1.3 Equilibrium Analysis

The ultimate goal of this section is to answer how the monitoring network shapes agents' propensities to accept bribes. Specifically, we will contrast the relationships between agents' bribe acceptance decisions across hierarchies and two-layer networks, and explore how the equilibria are controlled by network parameters.

1.3.1 Best Response

First, we solve for the optimal bribe acceptance decision of a random agent in the monitoring network. Suppose she has $l \geq 0$ supervisors, and $k > 0$ subordinates who have offered bribes to her. We consider which of these k bribe offers she optimally accepts.

This decision boils down to how many of the k bribe offers to accept, which can be summarized with one variable. This is because all bribe offers have the same value b; and accepting each of them brings to the agent identical risk of being caught, for the agent can only observe her own degrees (similarly, the risk of being reported by each of the l supervisors is equal).

Define the risk variables as follows:

Definition 1.1. ***Downward risk*** *p is the probability an agent is caught through accepting any given subordinate's bribe;* ***upward risk*** *r is the probability an agent is reported by any given supervisor.*

They are endogenous and will be derived from network parameters and agents' strategies.

We temporarily disregard how many subordinates the agent has and just think about how many bribes she desires to accept given no upper limit. To solve for that, we first need to know her *expected utility* of accepting $n \in \mathbb{N}$ bribes. Since the monitoring network is acyclic, the risk coming from each link is independent. Therefore, if the agent accepts n bribes, the probability she stays safe from being caught is $(1-q)(1-r)^l(1-p)^n$. Thus her

expected utility is:

$$\begin{aligned} U^l(n) &= (1-q)(1-r)^l(1-p)^n \cdot (n-l)b \\ &\quad + \left[1-(1-q)(1-r)^l(1-p)^n\right] \cdot (-lb-c). \end{aligned} \tag{1.1}$$

Only when not caught can the agent keep the accepted bribes nb and avoid the punishment cost c; lb is the cost of the bribes she offers to her supervisors.

Denote by $\hat{n} \equiv \arg\max U^l(n)$ the agent's *desired number of bribes*. The following lemma presents an important property of the expected utility function $U^l(n)$ which allows $\hat{n}$ to be represented in a simple form.

Lemma 1.1. *The expected utility function $U^l(n)$ is quasi-concave on $\mathbb{R}_+$; moreover, an agent's desired number of bribes $\hat{n}$ is independent of her degrees l, k and the upward risk r, it is a nonincreasing correspondence of the downward risk p.*

$U^l(n)$ is quasi-concave because while the marginal benefit of a bribe is constant at b, its marginal cost increases with the number of accepted bribes – each additional bribe acceptance brings the same confiscation risk p to all bribes already accepted. To see why the agent's desired number of bribes $\hat{n}$ is independent of her in-degree l and the upward risk r, notice that she is subject to the same reporting risk $\left[1-(1-r)^l\right]$ regardless of how many bribes she accepts; and the bribes she hands to her supervisors lb are predetermined sunk cost. The number of subordinates the agent has k only poses a capacity constraint on how many bribes she can accept and hence is irrelevant to how many she desires. The formal proof is supplied in Appendix 1.7.2.

Lemma 1.1 implies that all agents desire the same number(s) of bribes, which we will derive as a correspondence of the downward risk p. For that, we first compute the *cutoff downward risk* $p_{n(n+1)}$ that makes an agent indifferent between accepting n and $(n+1)$ bribes by equalizing the expected utilities $U^l(n)$ and $U^l(n+1)$:

$$p_{n(n+1)} = \frac{b}{c+(n+1)b}$$

Since $p_{n(n+1)}$ strictly decreases with n, the number of bribes an agent desires $\hat{n}$ is a nonincreasing correspondence of the downward risk p (the riskier bribe taking is, the fewer bribes one accepts):

$$\hat{n}(p) = \begin{cases} 0 & p \in (p_{01}, 1] \\ n & p \in (p_{n(n+1)}, p_{(n-1)n}) \\ \{n, n+1\} & p = p_{n(n+1)} \\ \infty & p = 0 \end{cases} \tag{1.2}$$

$\hat{n}$ is nondecreasing in the bribe value b, as an agent desires more bribes when they are more valuable.

(1.2) implies that at optimum, an agent mixes at most between accepting two adjacent numbers of bribes. Therefore, to capture mixed strategies, it is enough to extend the number of bribes to accept n to the real line: let $n \in \mathbb{R}_+$ stand for accepting $\lfloor n \rfloor$ bribes with probability $(\lfloor n \rfloor + 1 - n)$ and $(\lfloor n \rfloor + 1)$ bribes with probability $(n - \lfloor n \rfloor)$. For example, $n = 3.6$ means accepting 3 bribes with probability 0.4 and 4 bribes with probability 0.6. Next, extend the desired number of bribes correspondence $\hat{n}(p)$ to the real line accordingly by letting $\hat{n}(p) = [n, n+1]$ whenever $p = p_{n(n+1)}$ for some n.

Since an agent with k subordinates only receives k bribe offers, the optimal number(s) of bribes she actually accepts is $\min\left\{k, \hat{n}(p)\right\}$. As shown in figure 1.2[12], variations in agents' optimal strategies come solely from their different capacity constraints.

In principle, the probabilities with which an indifferent agent (whose capacity constraint is not binding) mixes between accepting $n \in \mathbb{N}$ and $(n+1)$ bribes could depend on her degrees. In the spirit of symmetric equilibria we assume they do not by imposing the following assumption:

Assumption 1.2. ***(Strong Symmetry Assumption)*** *There is $n \in \mathbb{R}_+$ such that an agent with k subordinates accepts $\min\{k, n\}$ bribes.*

[12]Parameter values: $b = 2$. In this figure and all subsequent analyses, the punishment cost c is normalized to 1.

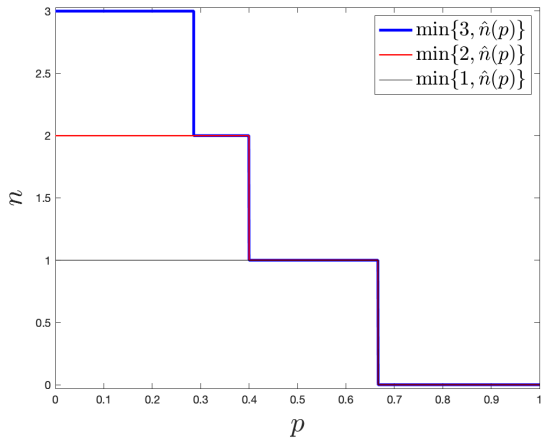


Figure 1.2: Optimal Strategies for Agents with Different Numbers of Subordinates ($k = 1, 2, 3$)

Assumption 1.2 allows us to capture all agents' strategies with a one-dim variable n (henceforth referred to as agents' *propensity to accept bribes*). Besides, it is without loss of generality.[13]

Risk Functions

Having mapped the downward risk p to agents' propensity to accept bribes n through the desired number of bribes correspondence $\hat{n}(p)$, we now do the reverse by defining the downward and upward risk functions: $P(n), R(n) \in [0, 1]$. They crucially rely on the network structure.

Due to information asymmetry, an agent treats her subordinates equally and randomly selects a set of bribe offers to accept. We thus obtain the upward risk – the probability an agent is reported by any given supervisor – as follows:

$$R(n) = \sum_{k \geq 1} \Big(1 - \frac{n \wedge k}{k}\Big) \hat{\mu}(k), \tag{1.3}$$

[13]To establish the generality of *strongly symmetric equilibria*, we define a *symmetric equilibrium* as an equilibrium at which agents with the same in-and out-degrees (l, k) play the same mixed strategy, and show that fixing the parameters, any symmetric equilibrium shares the same upward risk and expected utilities with one and only one strongly symmetric equilibrium. We relegate the formal statement, its proof, and an illustrative example to Appendix 1.7.3.1.

where $\big(1-(n\wedge k)/k\big)$ is the probability that a supervisor with k subordinates declines the agent's bribe offer and reports her.

Given $R(n)$, the downward risk function $P(n)$ is recursively defined[14] by the following equation:

$$\begin{aligned} p = \eta \cdot \sum_{k} \sum_{l\geq 1} \Big[& 1-(1-q)\big(1-R(n)\big)^{l-1} \\ & \cdot (1-p)^{\lfloor n\wedge k\rfloor}\big(1-(n\wedge k-\lfloor n\wedge k\rfloor)p\big)\Big]\hat{\lambda}(l)\mu(k|l). \end{aligned} \tag{1.4}$$

An agent is caught through accepting any given subordinate's bribe if and only if that subordinate is caught, and their bribe transaction is consequently detected (with probability η). Specifically, $\big[1-(1-q)(1-R(n))^{l-1}(1-p)^{\lfloor n\wedge k\rfloor}\big(1-(n\wedge k-\lfloor n\wedge k\rfloor)p\big)\big]$ is the probability a subordinate with degrees (l,k) is caught conditional on the agent's having accepted her bribe. It generalizes the similar expression in equation (1.1) to mixed strategies.

Strategic Substitutability/Complementarity

We now discuss whether an agent optimally acts against (strategic substitutability) or follows (strategic complementarity) her opponents' bribe acceptance decisions in two network structures – hierarchies and two-layer networks (see figure 1.1). We show that as each network class features a distinctive driving force, they provide opposite answers to this question:

Proposition 1.1. *In a hierarchy, the downward risk function $P(n)$ is strictly* ***increasing****; in a two-layer network, it is strictly* ***decreasing****.*

The proof is supplied in Appendix 1.7.5. The rising downward risk function $P(n)$[15] in a hierarchy implies strategic substitutability between agents' bribe acceptance decisions – an agent accepts **fewer** bribes when her opponents accept **more**. Intuitively, if a subordinate

[14] $P(n)$ is well-defined because any component in the monitoring network is finite, as ensured by Assumption 1.1(i). For the formal proof, see Appendix 1.7.4.

[15] More precisely, $P(n)$ is strictly monotone on $[0,\bar{k}]$ and stays constant on $[\bar{k},\infty)$, where $\bar{k}\equiv \operatorname{supp}\mu$ is the largest number of bribes an agent can accept.

accepts more bribes, the chance she gets caught through bribe taking rises, and so does the risk of accepting her bribe. This effect is further intensified by risk transmission through the *corruption subnetwork*[16], namely, the fact that an indirect subordinate of an agent accepts more bribes also increases the agent's bribe taking risk.[17]

In a two-layer network, since the downward risk function $P(n)$ is decreasing[15], agents' bribe acceptance decisions exhibit strategic complementarity – an agent accepts **more** bribes when her opponents accept **more**. This is because when an agent's co-supervisors accept more bribes, the risk her subordinates are reported, thus caught, falls. Hence, accepting bribes becomes less dangerous for her.[18]

While the two opposite forces – strategic substitutability/complementarity – are demonstrated with best responses here, with slight modifications on the model, they can be easily transformed into equilibrium results. See Corollary 1.1, Section 1.5.1.

1.3.2 Equilibrium

Fixing a set of parameters, the equilibria are found at the intersecting points of the downward risk function $P(n)$ and the desired number of bribes correspondence $\hat{n}(p)$. The *Intermediate Value Theorem* then implies:

Proposition 1.2. *An equilibrium $n^* \in \mathbb{R}_+$ exists.*

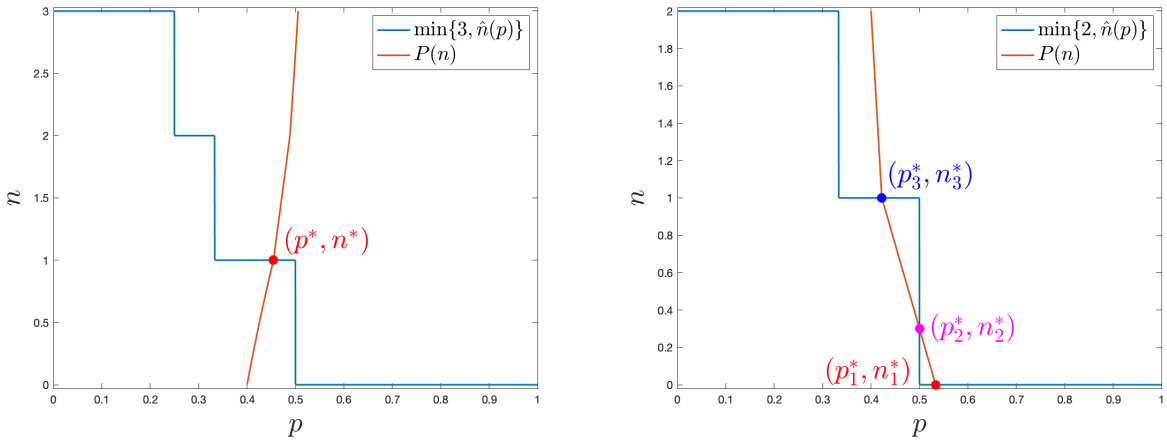
The shape of the downward risk function $P(n)$ has implications on equilibrium multiplicity. In a hierarchy, since $P(n)$ is strictly increasing while the desired number of bribes correspondence $\hat{n}(p)$ is weakly decreasing, they must intersect exactly once, producing a

[16]The *corruption subnetwork* is derived from the monitoring network by preserving only the links where bribe transactions successfully occur.

[17]To understand that, let us pick three agents in the network. Suppose agent 1 monitors agent 2, who monitors and colludes with agent 3. When 3 accepts more bribes, 2 becomes more likely to get caught through colluding with her. As a result, the risk for 1 to accept 2's bribe also increases.

[18]We have analyzed two special cases. In a more general network, both forces are present: an agent's bribe acceptance strategy is substitutable to the strategies of those in lower tiers, and complementary to the strategies of those in the same tier. Hence, the downward risk function $P(n)$ is generally non-monotonic.

unique equilibrium (see figure 1.3a). On the contrary, in a two-layer network, the strictly decreasing downward risk function $P(n)$ may cross the desired number of bribes correspondence $\hat{n}(p)$ more than once. Figure 1.3b illustrates that in this case, we may have multiple equilibria.



(a) Hierarchy[19] (b) Two-Layer Network[20]

Figure 1.3: Finding the Equilibria

Comparative Statics on the Network Parameters

In this section, we analyze how agents' equilibrium bribe acceptance propensity n^* is affected by both the network density – first-order differences in the network degree distributions – and constrained allocations of monitoring resources – second-order differences in the network degree distributions. We show that while the former generates consistent results across hierarchies and two-layer networks, the latter brings about opposite implications for them.

To tackle the equilibrium multiplicity problem, we restrict attention to the largest bribe acceptance propensity n^*, which is the one that maximizes any agent's expected utility (see Appendix 1.7.6). We also make the simplifying assumption: $\mu(\cdot|\tilde{l} = l) = \mu(\cdot|\tilde{l} \geq 1)\ \forall l \geq 1$,

[19]Parameter values: $\lambda(0) = 0.1, \lambda(1) = 0.9,\ \mu(0) = 0.7, \mu(3) = 0.3,\ \lambda \perp \mu;\ q = 0.5, \eta = 0.8, b = 1$. In this network, an agent has at most 3 subordinates and so can accept at most 3 bribes. Hence, to find the equilibrium, it suffices to truncate the desired number of correspondence $\hat{n}(p)$ at 3. We will apply the similar truncation to all following graphs.

[20]Parameter values: $\pi(1,0) = \pi(0,1) = 0.4, \pi(2,0) = \pi(0,2) = 0.1;\ q = 0.5, \eta = 0.8, b = 1$.

such that all agents monitored by at least one supervisor share the same conditional out-degree distribution. Notice that this assumption is satisfied by the two special cases – hierarchies and two-layer networks.

We define network density as follows:

Definition 1.2. *Network 1 is* ***denser*** *than network 2 if their degree distributions satisfy:* $\lambda_1 \succeq_{MLRP} \lambda_2$, $\mu_1 \succeq_{MLRP} \mu_2$, $\mu_1(\cdot|l \geq 1) \succeq_{FOSD} \mu_2(\cdot|l \geq 1)$.

Notice that the in- and out-degree distributions $\langle \lambda, \mu \rangle$ are connected via Assumption 1.1(*ii*). Here we compare them in the MLRP (*monotone likelihood ratio property*) manner, as MLRP ordering implies FOSD ordering for both the base degree distributions and their corresponding neighbors' degree distributions: $\lambda_1 \succeq_{MLRP} \lambda_2$ implies $\lambda_1 \succeq_{FOSD} \lambda_2$ and $\hat{\lambda}_1 \succeq_{FOSD} \hat{\lambda}_2$.[21]

Proposition 1.3. *If the monitoring network becomes* ***denser****, agents' propensity to accept bribes* n^* ***decreases****.*

The proof is supplied in Appendix 1.7.7. While Proposition 1.3 holds for general networks, the underlying intuition depends on the network structure. To gain more insight into it, consider first a hierarchy. When the network becomes denser, an agent has more subordinates and so can accept more bribes – the capacity constraint is relaxed. She is thus more likely to get caught through bribe taking, which in turn makes her bribes riskier to accept. Hence, agents tend to accept fewer bribes. Mathematically, an FOSD shift of monitored agents' out-degree distribution $\mu(\cdot|l \geq 1)$ shifts the downward risk function $P(n)$ right, pushing down the bribe acceptance propensity n^*.

Next, we look at a two-layer network. When the network becomes denser, not only is a subordinate monitored by more supervisors, but her risk of getting reported by each of them rises. To see the second point, notice now a supervisor has more subordinates and

[21] See [BV21] for other ways of treating comparative statics on network parameters when the friendship paradox is present.

so receives more bribe offers. She thus accepts a smaller proportion of them.[22] Again, a subordinate becomes more likely to get caught, deterring supervisors from accepting bribes. At a mathematical level, FOSD shifts of neighbors' degree distributions $\langle \hat{\lambda}, \hat{\mu} \rangle$ shifts the downward risk function $P(n)$ right, thus lowering agents' propensity to accept bribes n^*.

The same applies for any network, for it preserves the characteristics of both hierarchies and two-layer networks, namely, a subordinate of an agent can get caught either through accepting bribes (as manifested in hierarchies), or through being reported by others (as manifested in two-layer networks). As the network becomes denser, both events are more likely to happen, making the subordinate's bribe riskier to accept.

Notice that the bribe acceptance propensity n^* represents how many bribes one desires to take, but an agent with k subordinates can actually only accept $\min\{k, n^*\}$ bribes. Hence, in subsequent studies, we also consider a more objective corruption indicator – the *corruption level* $\kappa^* \equiv \mathbb{E}_\mu\big[\min\{k, n^*\}\big]$, which measures the average number of bribes an agent accepts.

Proposition 1.3 suggests the bribe acceptance propensity n^* falls with the network density, yet the same does not necessarily apply for the corruption level κ^*, as it is confounded by the direct increase in the out-degree distribution μ. In other words, although densifying the monitoring network facilitates corruption detection and thus deters people from accepting bribes, it simply creates more bribe taking opportunities. Hence, whether it reduces the corruption level κ^* is ambiguous.

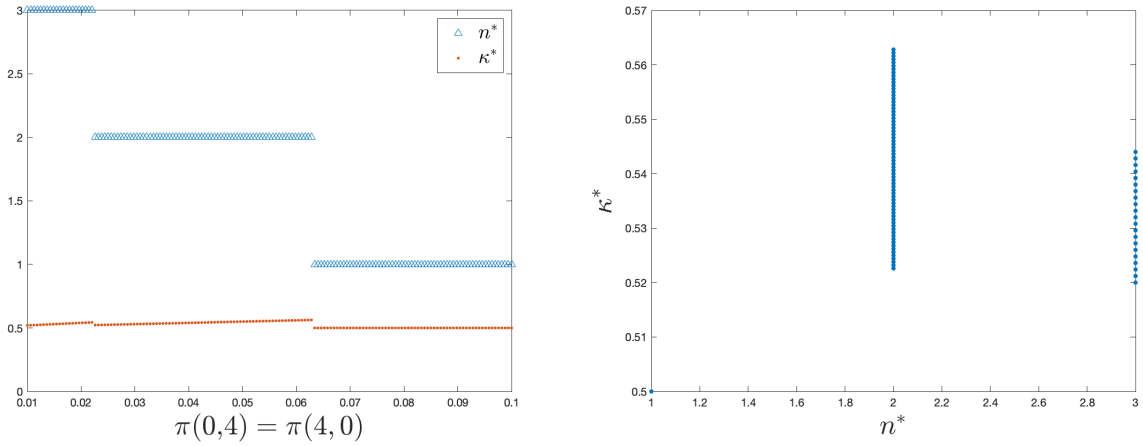
In particular, since the desired number of bribes correspondence $\hat{n}(p)$ is a step function, when the network becomes denser, agents' propensity to accept bribes n^* either drops or stays constant. In both two-layer networks (figure 1.4[23]) and hierarchies (figure 1.5[24]), a rise in the network density that does not affect the bribe acceptance propensity n^* raises

[22]For example, fix agents' bribe acceptance propensity $n = 1$. If each supervisor has one subordinate, then all bribes are accepted. Hence the upward risk $r = 0$. Now let each supervisor monitor two subordinates, then she randomly accepts one out of the two bribes. r thus rises to $1/2$.

[23]Parameter values: $\operatorname{supp} \pi = \big\{(1,0),(4,0),(0,1),(0,4)\big\}$; $q = 0.3, \eta = 0.8, b = 1$. In this example, the network finiteness assumption (Assumption 1.1(i)) is violated when the network density becomes large. But it is innocuous for the model performance.

[24]Parameter values: $\operatorname{supp} \lambda = \{0,1\}, \operatorname{supp} \mu = \{0,2\}$, $\lambda \perp \mu$; $q = 0.4, \eta = 0.8, b = 1$.

the corruption level κ^* through relaxing agents' capacity constraints $\big($1.4a and 1.5a plot the bribe acceptance propensity n^* (blue) and the corruption level κ^* (orange) against the network density; 1.4b and 1.5b plot the corruption level κ^* against the bribe acceptance propensity $n^*\big)$. Yet increases in the network density that do reduce the bribe acceptance propensity n^* also lower the corruption level κ^* – it is obvious for two-layer networks, as the bribe acceptance propensity n^* drops discontinuously; in a hierarchy, the corruption level κ^* falls continuously along with n^*, for agents' reduced propensity to accept bribes always dominates their expanded freedom in doing so.[25]



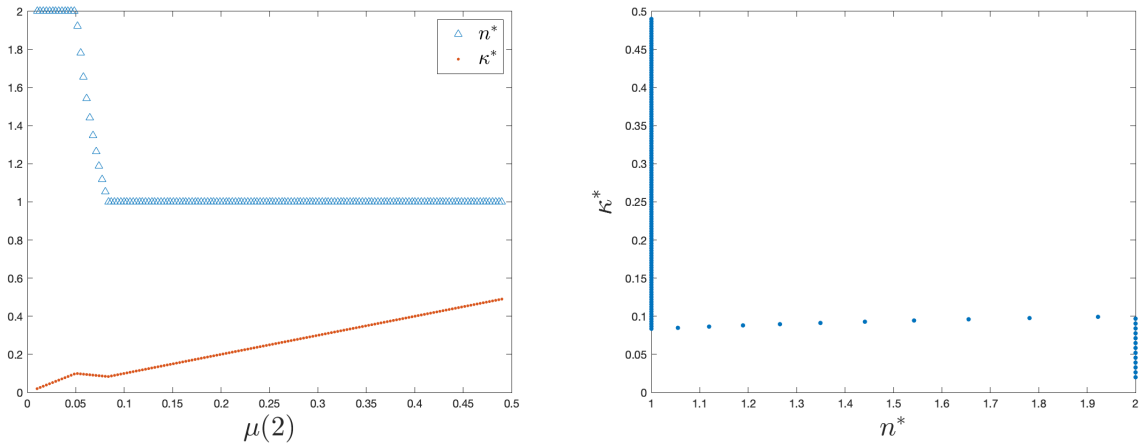
(a) n^*, κ^* on the Network Density (b) κ^* on n^*

Figure 1.4: Comparing n^* with κ^* when Varying the Network Density in a Two-Layer Network

Since first-order increases in the network degree distributions raise the average number of subordinates each agent monitors $\mathbb{E}_\mu(k)$, Proposition 1.3 cannot speak to constrained allocations of monitoring resources. We now fill in this gap by performing second-order analyses on the network degree distributions.

Our goal is best illustrated with a network design question: given a large set of agents, suppose we have some fixed number of monitoring links $\big($fixing $\mathbb{E}_\mu(k)\big)$, and would like to allocate them to monitor a fixed proportion of the agents $\big($fixing $1-\lambda(0)\big)$, how do we better

[25]In Appendix 1.7.8, we formalize this result for hierarchies (Proposition 1.12) and generalize it to any network where the in- and out-degrees are independent ($\lambda \perp \mu$).



(a) n^*, κ^* on the Network Density (b) κ^* on n^*

Figure 1.5: Comparing n^* with κ^* when Varying the Network Density in a Hierarchy

design the network structure to reduce corruption?

In the following proposition, we show that when the number of subordinates each agent monitors is less evenly distributed, corruption is alleviated in a two-layer network, yet aggravated in a hierarchy.

Proposition 1.4. *i In a two-layer network, a mean-preserving spread of unmonitored agents' out-degree distribution $\mu(\cdot|l=0)$* ***decreases*** *agents' bribe acceptance propensity n^* and the corruption level κ^*.*

ii In a hierarchy with independent in- and out-degrees ($\lambda \perp \mu$), a mean-preserving spread of the out-degree distribution μ ***increases*** *agents' bribe acceptance propensity n^*.*

Proof: See Appendix 1.7.9.

We explain the intuitions with small networks. Consider first a two-layer network with two supervisors, three subordinates and four monitoring links (figure 1.6). When the number of subordinates each supervisor monitors becomes less evenly distributed (from 1.6*a* to 1.6*b*), the bribe taking risk rises. To see that, suppose each agent desires to take $n = 2$ bribes. Then all bribes are accepted in 1.6a, making a subordinate unlikely to be reported. In comparison, in 1.6*b*, while agent 2's capacity constraint is binding, agent 1 receives three bribe offers and

so reports one subordinate. Hence, accepting bribes from subordinates becomes riskier, thus less appealing. Mathematically, a mean-preserving spread of unmonitored agents' out-degree distribution $\mu(\cdot|l = 0)$ shifts the downward risk function $P(n)$ right, thus lowering agents' bribe acceptance propensity n^*.

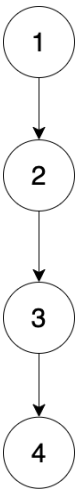
(a) More Even Allocation (b) Less Even Allocation

Figure 1.6: Constrained Allocation of Out-Links in a Two-Layer Network

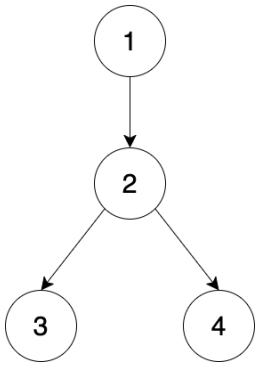
Next, we look at a four-agent hierarchy with three monitoring links (figure 1.7). Compared with the linear hierarchy (1.7a), in 1.7b, the number of subordinates each agent monitors becomes less evenly distributed, reducing the bribe taking risk. Intuitively, suppose each agent desires to accept $n = 1$ bribe. In 1.7a, agent 1 can be caught through bribe taking if any of 2, 3 and 4 is caught directly. In comparison, in 1.7b, she is in danger only if either 2 or 3 is caught directly (w.l.o.g. assume 2 accepts 3's bribe and reports 4). Thus bribe taking becomes less risky for 1. A similar analysis reveals that it also becomes less risky for 2. Mathematically, contrary to a two-layer network, a mean-preserving spread of the out-degree distribution μ shifts the downward risk function $P(n)$ left, raising agents' bribe acceptance propensity n^*. This example suggests that in a linear hierarchy, risk percolates up most smoothly, implying maximum deterrence on corruption.

Notice that a mean-preserving spread of the out-degree distribution μ also has a direct negative effect on the corruption level κ^*, as then more agents' capacity constraints bind, forcing agents to accept fewer bribes on average.[26] Hence, in a two-layer network, it causes the corruption level κ^* to drop together with the bribe acceptance propensity n^*; nevertheless, in a hierarchy where the bribe acceptance propensity n^* rises with a mean-preserving spread

[26]The intuition can be illustrated with figure 1.6. When each agent desires to accept $n = 2$ bribes, in 1.6a, both 1 and 2 have relaxed capacity constraints, and so the corruption level $\kappa = 4/5$. In comparison, in 1.6b, 2's capacity constraint binds, leading to $\kappa = 3/5$. The same intuition plays out in figure 1.7.



(a) More Even Allocation



(b) Less Even Allocation

Figure 1.7: Constrained Allocation of Out-Links in a Hierarchy

of the out-degree distribution μ, changes in the corruption level κ^* is ambiguous. Indeed, in a linear hierarchy where agents are the most discouraged from bribe taking (figure 1.7a), full corruption is nevertheless easily arrived at whenever an agent desires to accept at least one bribe.

This issue for hierarchies can be circumvented by shifting monitored agents' out-degree distribution $\mu(\cdot|l \geq 1)$ instead of the unconditional one. In Appendix 1.7.10, we illustrate its intuition and supply the formal statement.

1.4 Identifying Corruption

We now delve into the identification of the corruption level κ^*, which can be interpreted as the per person number of corruption cases, where a *corruption case* refers to an incident of successful bribe exchange. Since it is hard to observe the corruption level κ^* in reality, many empirical studies instead adopt the per person number of corruption cases **detected** by law enforcements (denoted by $\hat{\kappa}^*$) to measure corruption.[27] As the probability each corruption case is detected is given by the downward risk p^* (henceforth referred to as the *detection*

[27]Such measures are particularly popular for studying corruption in non-US countries. To cite a few examples: [DT13], [KS14], [SSZ16], [MO19], [Zak19], etc.

rate), we have

$$\hat{\kappa}^* = p^* \cdot \kappa^*.$$

An immediate problem of the *corruption measure* $\hat{\kappa}^*$ is underestimation ($\hat{\kappa}^* < \kappa^*$), as corruption detection is rarely perfect ($p^* < 1$). This issue is widely acknowledged in applied economics literature and formally addressed by [KS14].[28] Another problem is the non-monotonic relationship between the corruption measure $\hat{\kappa}^*$ and the corruption level κ^*, so that the former does not even capture the trend of the latter:

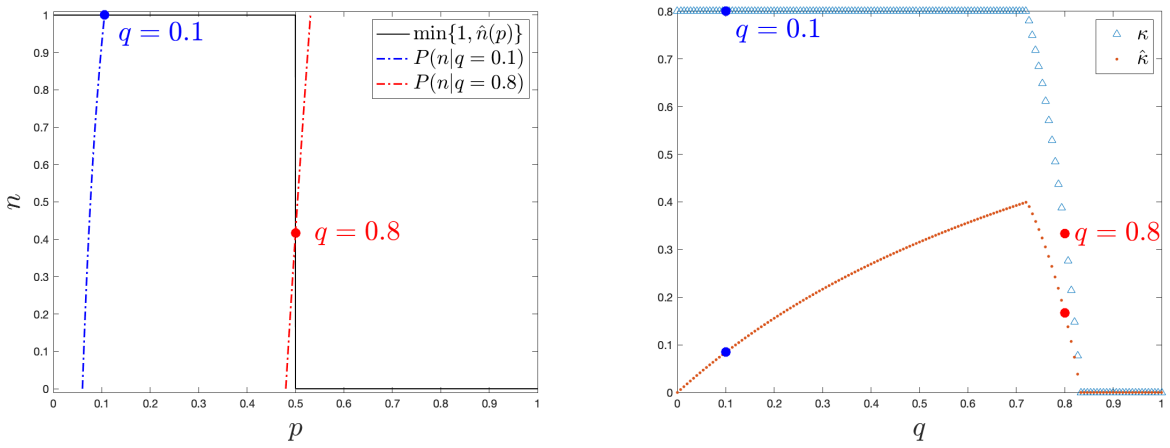
Proposition 1.5. *The corruption measure $\hat{\kappa}^*$ is **not** generally monotonic in the corruption level κ^*: a change in exogenous parameter values can raise (lower) κ^* and lower (resp., raise) $\hat{\kappa}^*$.*

We prove Proposition 1.5 by examples. Figure 1.8a illustrates what happens in a hierarchy when the external monitoring success rate q is enhanced from 10% to 80%. Since now one's subordinates are more likely to be caught directly, it becomes more dangerous to accept bribes, i.e., the downward risk function $P(n)$ shifts right. Hence, agents' bribe acceptance propensity n^* drops, resulting in reduction in the corruption level κ^* from 0.8 to 0.33 (figure 1.8b). However, thanks to the higher corruption detection rate p^*, the corruption measure $\hat{\kappa}^*$ rises from 0.08 to 0.17. This example is consistent with the finding in [GN11] that the corruption conviction rate in the US is positively and significantly correlated with law enforcement strength.

More generally, figure 1.8b plots the corruption level κ^* (blue) and corruption measure $\hat{\kappa}^*$ (orange) against the external monitoring success rate q when it varies between 0 and 1. We can see that only when the corruption level κ^* strictly decreases with q does the corruption measure $\hat{\kappa}^*$ follow its trend, for the detection rate p^* is constant in this range; otherwise, if the corruption level κ^* stays constant, the detection rate p^* rises with q, and so does the corruption measure $\hat{\kappa}^*$.

[28] They estimate the "reporting rate of corruption" and divide the observed number of corruption cases by it to obtain a relatively unbiased corruption measure.

[29] Parameter values: $\lambda(0) = 0.2, \lambda(1) = 0.8$, $\mu = \lambda$, $\lambda \perp \mu$; $\eta = 0.6, b = 1$.



(a) Equilibria at Different q's[29]

(b) $\kappa^*, \hat{\kappa}^*$ on q

Figure 1.8: Discrepancies between κ^* and $\hat{\kappa}^*$ when Varying q in a Hierarchy

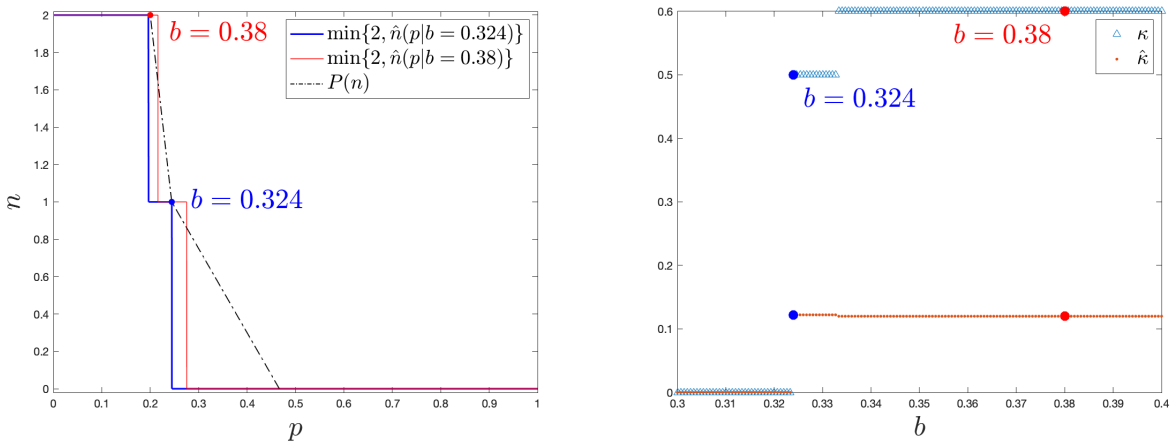
To cite another example, consider changing the bribe value b in a two-layer network (it is equivalent to varying the punishment cost c). In figure 1.9a, when the bribe value b increases from 0.324 to 0.38, bribes become more desirable – the desired number of bribes correspondence $\hat{n}(p)$ shifts right. Thus, agents' propensity to accept bribes n^* rises from 1 to 2, as well as the corruption level κ^* from 0.5 to 0.6 (figure 1.9b). Nevertheless, since the downward risk function $P(n)$ is decreasing (strategic complementarity), the detection rate p^* drops with the bribe acceptance propensity n^*, resulting in falling corruption measure $\hat{\kappa}^*$ from 0.122 to 0.12. In a more general manner, figure 1.9b demonstrates the non-monotonic relationship between the corruption level κ^* (blue) and the corruption measure $\hat{\kappa}^*$ (orange) when the bribe value b varies on a larger scale.

Lastly, when network parameters change, while the corruption measure $\hat{\kappa}^*$ follows exactly the same trend as the corruption level κ^* in both hierarchies and two-layer networks (see figure 1.4a for the "discrete jumps" in two-layer networks and figure 1.5a for the "zigzag pattern" in hierarchies[31]), wrong prediction can still arise around the turning points.[32] For

[30]Parameter values: $\pi(1,0) = \pi(0,1) = 0.4, \pi(2,0) = \pi(0,2) = 0.1$; $q = 0.2, \eta = 1$.

[31]Wherever the corruption level κ^* increases, the detection rate p^* also rises, and so does the corruption measure $\hat{\kappa}^*$; wherever the corruption level κ^* drops, the detection rate p^* stays constant, thus the corruption measure $\hat{\kappa}^*$ drops along with κ^*.

[32]We use a hypothetical example to illustrate the intuition. In a hierarchy, suppose we keep increasing



(a) Equilibria at Different b's[30]

(b) $\kappa^*, \hat{\kappa}^*$ on b

Figure 1.9: Discrepancies between κ^* and $\hat{\kappa}^*$ when Varying b in a Two-Layer Network

instance, in a hierarchy where one has either 0 or 2 subordinates, if the network becomes denser, that is, if the probability of having 2 subordinates, $\mu(2)$, rises from 0.044 to 0.087, the corruption level κ^* drops from 0.0872 to 0.0868, whereas the corruption measure $\hat{\kappa}^*$ rises from 0.0289 to 0.0290.

The measurement exercise fails because the corruption detection rate p^* is endogenous: a corruption case is detected only if the involved criminal subordinate is caught, which depends on the bribe acceptance strategies of the agents elsewhere in the network.

One tentative approach to solve this problem is to trim off the endogenous part of the detection rate p^*. Suppose we know the sources of detection for all observed corruption cases, that is, given any observed corruption case, we know whether the briber is (1) reported by a supervisor, (2) detected directly by the law enforcement agency, or (3) caught through accepting bribes (notice that the three cases are not mutually exclusive). Then we can construct the *per person number of corruption cases detected through source (2)* (denoted by $\hat{\kappa}_e^*$) as a new measure for the corruption level κ^*.

---

the network density and find three equilibrium points along the way: $(\kappa^*, \hat{\kappa}^*) = (0.4, 0.2)$, $(0.43, 0.21)$, $(0.41, 0.18)$, where (0.43,0.21) is a local maximum, then the surrounding points (0.4,0.2) and (0.41,0.18) form an instance of non-monotonicity.

Define the *probability any given corruption case is detected through source (2)*:

$$p_e = \eta q,$$

which says that a corruption case is detected directly if and only if the subordinate is directly caught as criminal (with probability q), and the supervisor who has accepted her bribe is caught through the subsequent corruption investigation (with probability η). The new corruption measure can thus be expressed as

$$\hat{\kappa}_e^* = p_e \cdot \kappa^* = \eta q \cdot \kappa^*.$$

Since p_e is exogenous, so long as the law enforcement strength (q, η) is controlled for,[33] the new measure $\hat{\kappa}_e^*$ correctly reflects the trend of the corruption level κ^*.[34]

1.5 Extensions

1.5.1 A Model with Corruptible and Incorruptible Agents

In Section 1.3.1, we show that while agents' bribe acceptance decisions are strategically substitutable in hierarchies, they are strategically complementary in two-layer networks. This section reformulates the discovery as comparative statics.

We achieve it by making modest extensions on the baseline model. Suppose now an agent is *corruptible* with independent probability $\gamma \in (0, 1]$, and *incorruptible* with probability $(1 - \gamma)$. A corruptible agent makes strategic bribe acceptance decisions; an incorruptible

[33]For details on how it is typically addressed in applied economics research, see, for instance, [GR89], [SSZ16], and [MO19].

[34]The term "law enforcement agency" should not be taken at face value. More precisely, it refers to any anticorruption agency exerting influence on but sufficiently independent from the object of study. For instance, if we are interested in studying police corruption, this term fails to apply as the police system itself is a law enforcement agency. Instead, to circumvent the endogeneity problem, we can focus on those corruption cases independently discovered by an external investigative commission, i.e., they are detected neither thanks to a whistleblower nor through tracing up bribe transactions.

agent is a crazy type that never accepts bribes.[35]

This extended model gives the following result as a corollary for Proposition 1.1:

Corollary 1.1. *When the proportion of corruptible agents γ* ***increases****, their bribe acceptance propensity n^** ***decreases*** *in a hierarchy, and* ***increases*** *in a two-layer network.*

Proof: When the proportion of corruptible agents γ rises, in a hierarchy, the downward risk function $P(n)$ shifts right, lowering the bribe acceptance propensity n^*; in a two-layer network, the upward risk function $R(n)$ decreases, hence $P(n)$ shifts left, elevating the bribe acceptance propensity n^*.

Q.E.D.

1.5.2 A Model with Criminal and Innocent Agents

Each agent is criminal with independent, publicly known probability $s \in (0, 1)$, so the chance she is innocent is $(1 - s)$. An agent is *guilty* if she is either criminal or *corrupt* – having accepted at least one bribe. Guilty agents are subject to conviction by the external law enforcement agency and offer bribes to their supervisors.

Agents simultaneously choose how many bribes to accept maximally $n \geq 0$. The game then clears starting from the agents with no subordinate.

1.5.2.1 Best Response

Criminal agents' incentives are the same as before. Their expected utilities remain unaltered $\big($equation (1.1)$\big)$, and so do their optimal strategies depending only on the downward risk p (figure 1.2).

In comparison, innocent agents are less inclined to accept bribes, as keeping away from corruption protects them against being caught. An innocent agent's optimal strategy can

[35]Equilibria are derived in the same way as those for the baseline model, though the risk functions $P(n), R(n)$ are slightly different, as displayed in Appendix 1.7.11.1.

be derived from that of a criminal agent with the same degrees. Suppose they both have l supervisors and $j > 0$ guilty subordinates. Their expected utilities when accepting any positive number of bribes coincide; if accepting 0 bribe, the innocent agent is not guilty and so derives strictly higher expected utility than the criminal agent does $\big(0 > U^l(0) = (1-q)(1-r)^l \cdot (-lb) + [1-(1-q)(1-r)^l] \cdot (-lb-c)\big)$. Hence, while the criminal agent accepts $\min\{j, n\}$ bribes at optimum (recall that $n \in \mathbb{R}_+$ is a criminal agent's bribe acceptance propensity), the innocent agent accepts either 0 or $\min\{j, n\}$ bribes depending on which option generates larger payoff.[36]

This binary choice can be captured by a convenient expression. Suppose the bribe acceptance propensity $n \geq 1$. Since the expected utility function U^l is quasi-concave (Lemma 1.1), before the optimal number of bribes to accept n is reached, the more bribes one accepts, the higher her payoff is. Construct the *cutoff policy* $\underline{j}^l \in \big\{1, ..., \lfloor n \rfloor, \infty\big\}$ to be the smallest number of bribes an innocent agent with l supervisors is willing to accept. Given that her utility of accepting no bribe is 0, $\underline{j}^l$ must be the smallest number of bribes that generates nonnegative expected utility:

$$\underline{j}^l = \begin{cases} \min\limits_{U^l(j) \geq 0,\ j \in \{1,...,\lfloor n \rfloor\}} j & \text{if } U^l(\lfloor n \rfloor) \geq 0 \\ \infty & \text{if } U^l(\lfloor n \rfloor) < 0 \end{cases} \tag{1.5}$$

So the optimal strategy for an innocent agent with l supervisors and $j > 0$ guilty subordinates is to accept $\min\{j, n\}$ bribes if $\min\{j, n\} \geq \underline{j}^l$, and to decline all bribes otherwise. Notice that when the bribe acceptance propensity $n < 1$, the agent optimally accepts no bribe – the cutoff policy $\underline{j}^l = \infty$, as the fact that criminal agents are indifferent between accepting 0 and 1 bribe suggests innocent agents strictly prefer the former $\big(0 > U^l(0) = U^l(1)\big)$.

The following lemma presents some illustrative properties of the cutoff policy $\underline{j}^l$:

Lemma 1.2. *i An innocent agent's cutoff policy $\underline{j}^l$* ***increases*** *in the number of supervisors*

[36] We insist on the Strong Symmetry Assumption, that is, all agents share the same bribe acceptance propensity $n \in \mathbb{R}_+$. For the formal statement, see Definition 1.4(i), Appendix 1.7.3.2.

she has l and the downward and upward risks p, r.

ii An innocent agent strictly prefers not to accept bribes if the number of her supervisors is at least the same as that of her guilty subordinates ($l \geq j > 0$).

To see (i), notice that if an innocent agent has more supervisors – l gets larger, then once she becomes corrupt, not only does she need to pay more bribes, but she is also more likely to get reported. Similarly, larger upward risk r elevates her chance of getting reported through each link. Both deter her from engaging in corruption. Larger downward risk p makes accepting bribes riskier and thus less attractive. (ii) can be understood with a simple reasoning: suppose an innocent agent has no fewer supervisors than guilty subordinates ($l \geq j > 0$). If she decides to engage in corruption, her most optimistic outcome is to accept all the bribes while remaining safe from being caught, which however gives her nonpositive payoff $(j - l)b \leq 0$. Since the agent has a strictly positive chance of getting caught through accepting bribes (a subordinate is caught directly with probability $q > 0$), she would rather stay away from corruption.

Lemma 1.2(i) implies that compared with a criminal agent whose bribe acceptance decision only depends on the behaviors of her co-supervisors (strategic complementarity) and direct and indirect subordinates (strategic substitutability), an innocent agent also bases the bribe acceptance decision on her direct supervisors' behaviors – the more inclined they are to accept bribes, the less likely the innocent agent is reported should she engage in corruption, and the more she tends to do so (the cutoff policy $\underline{j}^l$ decreases when the upward risk r gets smaller), suggesting a new complementary force between agents' strategies.

1.5.2.2 Equilibrium

Since only guilty agents offer bribes, we prune non-guilty agents from the network. Let g be the probability any given subordinate of an agent is guilty, and μ_g be the distribution of the number of guilty subordinates a random agent has. g and μ_g are interdependent and can be jointly solved as functions of agents' strategies $\langle n, (\underline{j}^l)_l \rangle$. Given these statistics, we can then

define the risk functions $P\big(n,(\underline{j}^l)_l\big), R\big(n,(\underline{j}^l)_l\big) \in [0,1]$.[37]

Since innocent agents' optimal strategies depend on both risk variables (p,r), in finding the equilibria, it is necessary to employ a more advanced technique: we construct a self-map $\Gamma : (p,r) \mapsto (p,r)$ and show that the equilibria are found at its fixed points: $(p^*,r^*) \in \Gamma(p^*,r^*)$. Specifically, given the downward and upward risks (p,r), we solve for all optimal strategies with the form $\langle n,(\underline{j}^l)_l\rangle$ through the desired number of bribes correspondence $\hat{n}(p)$ and condition (1.5) that characterizes innocent agents' cutoff policies $(\underline{j}^l)_l$. We then map the optimal strategies back to the risk variables (p,r) using the risk functions P,R.

For the self-map Γ to be well-defined, we need to incorporate in innocent agents' mixed strategies between accepting zero and some positive number of bribes: if at some downward and upward risks (p,r) an innocent agent is indifferent between these two choices, that is, if $U^l\big(\lfloor \hat{n}(p) \wedge j \rfloor\big)|_{p,r} = 0$ for some l,j, we extend the domain of the risk functions P,R to include in all the induced optimal mixed strategies.[38]

Figure 1.10[39] illustrates how the equilibria are reached in a two-layer network. In this case, since supervisors who make bribe acceptance decisions are not monitored, their optimal strategies, thus the self-map Γ, are defined on the downward risk p alone. 1.10a depicts agents' optimal strategies. Innocent agents' cutoff policy $\underline{j}^0$ (red) increases with the downward risk p, for they are more reluctant to engage in corruption when accepting bribes is more dangerous. If p is sufficiently large, $\underline{j}^0$ goes to infinity, indicating they never accept bribes. 1.10b visualizes the self map $\Gamma : p \mapsto p$ (blue) and characterizes the equilibrium p^* at its fixed point.

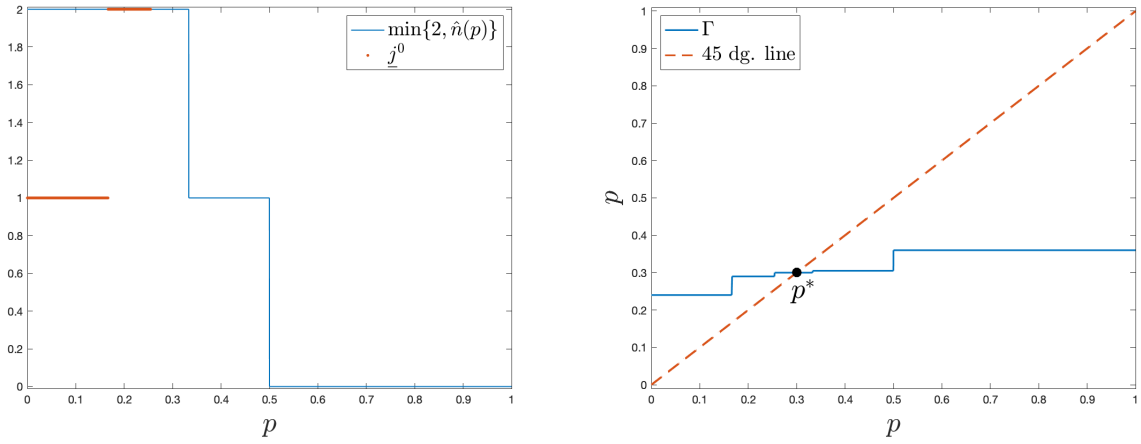
Figure 1.11[40] demonstrates agents' optimal strategies in a hierarchy. Now, since innocent, monitored agents base their bribe acceptance decisions on both the downward and upward

[37]For the formal expressions of the statistics for the network of guilty agents $\langle g,\mu_g\rangle$ and the risk functions P,R, see Appendix 1.7.11.2.

[38]As part of the Strong Symmetry Assumption, we impose that all innocent agents indifferent between accepting bribes and not mix them with the same probabilities (Definition 1.4(ii), Appendix 1.7.3.2).

[39]Parameter values: $\pi(1,0)=\pi(0,1)=0.4, \pi(2,0)=\pi(0,2)=0.1$; $q=0.4, \eta=0.6, b=1, s=0.5$.

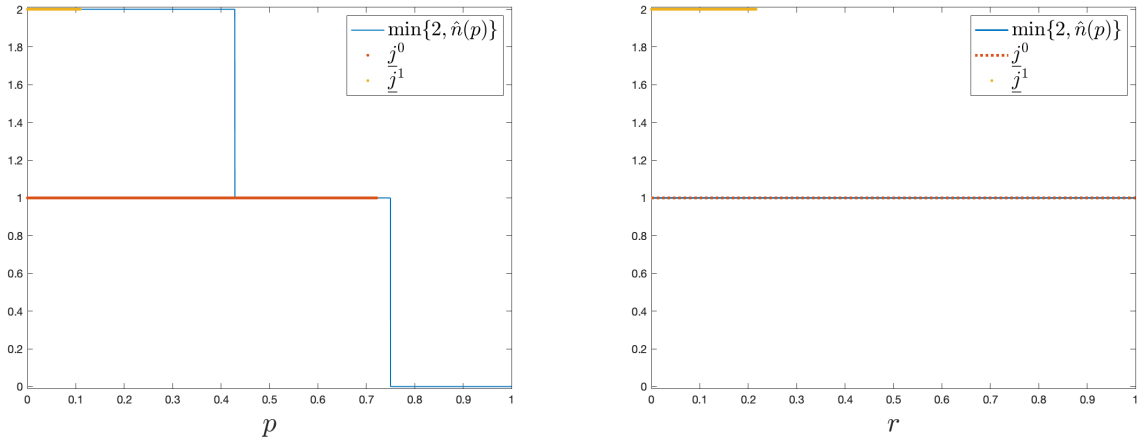
[40]Parameter values: $\max \operatorname{supp} \mu = 2$; $q=0.1, b=3$.



(a) Optimal Strategies (b) Self-Map Γ and the Equilibrium

Figure 1.10: Equilibrium for a Two-Layer Network

risks (p, r), in each of 1.11a and 1.11b, we fix one risk variable and plot the optimal strategies on the other. As reflected in both graphs, among the innocent agents, compared with those at the top, monitored agents incur the additional risk of being reported and thus adopt a more stringent cutoff policy ($\underline{j}^1 \geq \underline{j}^0$, $\underline{j}^1$ in yellow, $\underline{j}^0$ in red). This result is consistent with Lemma 1.2(i) that the cutoff policy $\underline{j}^l$ increases with the number of supervisors one has l.



(a) Optimal Strategies on p (Fixing $r = 0.2$) (b) Optimal Strategies on r (Fixing $p = 0.1$)

Figure 1.11: Optimal Strategies in a Hierarchy

The self-map Γ is upper hemicontinuous, and convex and closed at each point (p, r), thus equilibrium existence is proved with *Kakutani's Fixed Point Theorem*.

Proposition 1.6. *An equilibrium exists.*

1.5.2.3 Comparative Statics on the Crime Rate

In this model, corruption is triggered by crime, hence a natural question to ask is how the prevalence of crime influences agents' bribe acceptance decisions. We focus on two network structures – linear hierarchies and two-layer networks – and show that changes in the *crime rate* s have different implications within and across networks.

We analyze agents' *propensities to accept bribes* $\langle n^*, (\underline{j}^{l*})_l \rangle$, as well as the corruption level κ^* defined again as the average number of bribes an agent accepts.[41] A linear hierarchy produces a unique equilibrium,[42] yet multiple equilibria[43] can exist for a two-layer network. We thus select the largest equilibrium as before.[44]

Unlike hierarchies in general, in a linear hierarchy, only top agents' cutoff policy $\underline{j}^{0*}$ matters, as innocent, monitored agents never accept bribes – they have weakly more supervisors than guilty subordinates $\big($Lemma 1.2$(ii)\big)$. The irrelevance of monitored agents' cutoff policies is necessary in producing clear-cut results, justifying our restriction.

In a linear hierarchy, when the crime rate s **increases**, agents' propensities to accept bribes **fall**: n^* decreases, $\underline{j}^{0*}$ increases. This is because when there are more criminals who

[41] See Appendix 1.7.11.2 for the formal expression of the corruption level κ in this extended model.

[42] In a linear hierarchy, the downward risk function P depends only on criminal agents' bribe acceptance propensity n (see Footnote 45). Since $P(n)$ is increasing in a hierarchy, we have equilibrium uniqueness.

[43] Remember that we impose the Strong Symmetry Assumption. In Appendix 1.7.3.2, we argue that even in this extended model, in a two-layer network, it is without loss of generality to restrict attention to the strongly symmetric equilibria.

[44] The largest equilibrium is the one with the smallest downward risk p^*, which must accompany the largest bribe acceptance propensity for criminal agents n^* (the desired number of bribes correspondence $\hat{n}(p)$ is nonincreasing), and the smallest cutoff policy for innocent agents $\underline{j}^{0*}$ (the smaller the downward risk p^* is, the more incentivized innocent agents are to accept bribes). We claim that **in this extended model, in a two-layer network, the largest equilibrium maximizes any agent's expected utility**. The proof is similar to that for Proposition 1.11, Appendix 1.7.6, except that the upward risk r^* is now irrelevant to supervisors' expected utilities, and unlike in the baseline model, innocent supervisors now have the option to decline all bribes and achieve 0 payoff. Multiple equilibria at the smallest downward risk p^* can occur if both criminal and innocent agents employ mixed strategies, though they share the same expected utilities and are thus virtually equivalent. In this case, we make the further refinement by selecting the equilibrium with the largest bribe acceptance propensity for criminal agents n^*.

offer and (may) accept bribes, the corruption subnetwork expands, making a corruption investigation easier to percolate up to endanger a bribe taker. Thus, agents are deterred from corrupt behaviors.[45] This result echoes the strategic substitutability in a hierarchy.

The comparative statics for the corruption level κ^* is ambiguous, as it is confounded by two forces that counteract the reduction in agents' bribe acceptance propensities – more criminals induce more bribe taking opportunities, relaxing agents' capacity constraints; besides, criminal agents who are more prone to bribe taking now constitute a larger population.

Now look at two-layer networks. We distinguish between the crime rate among supervisors and that among subordinates, for they generate different results.

Proposition 1.7. *In a two-layer network, when supervisors' crime rate* ***increases****, agents' propensities to accept bribes and the corruption level both* ***rise****:* n^* *increases,* $\underline{j}^{0*}$ *decreases,* κ^* *increases; when subordinates' crime rate* ***increases****, agents' propensities to accept bribes* ***fall****:* n^* *decreases,* $\underline{j}^{0*}$ *increases.*

Proof: When supervisors' crime rate increases, the upward risk function R shifts down, and so does the downward risk function P as well as the self-map $\Gamma : p \mapsto p$. Hence, the equilibrium downward risk p^* falls, implying larger bribe acceptance propensities and, consequently, larger corruption level κ^*. The opposite happens when subordinates' crime rate increases, leading to smaller bribe acceptance propensities.

Q.E.D.

Intuitively, if there are more criminals among the supervisors who tend more to accept bribes, criminal subordinates are less likely to be reported. Hence, accepting bribes from them is less risky and thus more attractive. This result is reminiscent of the strategic complementarity in a two-layer network. This increase in agents' propensity to accept bribes,

[45]Mathematically, since innocent, monitored agents never accept bribes, the downward risk function P depends only on criminal agents' bribe acceptance propensity n. When the crime rate s increases, $P(n)$ shifts right, resulting in smaller bribe acceptance propensity for criminal agents n^* and larger downward risk p^*, which in turn raises innocent, top agents' cutoff policy $\underline{j}^{0*}$.

together with the rising population of criminal supervisors more susceptible to bribe taking, elevates the corruption level κ^*.

In contrast, if there are more criminals among the subordinates, more bribes are offered and thus fewer are accepted. As a result, it is more dangerous and so less appealing to accept criminal subordinates' bribes. Nevertheless, just like in a linear hierarchy, increases in the supply of bribe offers also relax supervisors' capacity constraints. Hence, it is hard to say whether the corruption level κ^* rises or falls.

1.6 Conclusion

In this paper, we study the bribe acceptance decisions of corruptible monitors when they are placed in a monitoring network that propagates bribe taking risk. All networks are in between two extreme cases – hierarchies and two-layer networks – that predict opposite relationships between an agent's bribe taking risk and her opponents' bribe acceptance behaviors. In equilibrium, while first-order increases in network degree distributions (densifying the network) deter agents from accepting bribes in any network, second-order increases in out-degree distributions (allocating monitoring resources more evenly) generate contrasting outcomes across hierarchies and two-layer networks.

For simplicity, we assume away bribers' strategic decisions, cyclic networks (which allow mutual monitoring), and the possibility of being caught through bribing a supervisor. All are meaningful directions future research could advance in. Besides, it remains to be tested whether this model can be implemented in empirical studies, and whether the issue it uncovers on corruption measurement is evident in reality.

1.7 Appendix

1.7.1 Network Finiteness Assumption

In this section, we prove that Assumption 1.1(i) is a necessary and sufficient condition for a component of the monitoring network to be finite. The formal statement is given below:

Proposition 1.8. *Any component of a random monitoring network is finite in expectation iff Assumption 1.1(i) is satisfied.*

Proof: Define by n_1 the expected number of (direct and indirect) neighboring nodes we get through tracing down an out-link, and by n_2 that through tracing up an in-link. By definition, they are recursively expressed as follows:

$$n_1 = 1 + \mathbb{E}_{\hat{\lambda}}\Big[(l-1)n_2 + \mathbb{E}_{\mu}(k|l) \cdot n_1\Big] \tag{1.6}$$

$$n_2 = 1 + \mathbb{E}_{\hat{\mu}}\Big[(k-1)n_1 + \mathbb{E}_{\lambda}(l|k) \cdot n_2\Big] \tag{1.7}$$

When we trace down an out-link, we first count in the direct subordinate. Suppose she has degrees (l, k), we add in all the nodes we get by tracing through each of her other in-links $(l-1)n_2$, and each of her out-links kn_1. We then take expectation of her degrees (l, k) to get an expression for n_1. That for n_2 is similarly defined.

Rearranging (1.6) and (1.7), we get explicit expressions for n_1, n_2:

$$n_1 = \frac{1 - \mathbb{E}_{\hat{\mu}}\big[\mathbb{E}_{\lambda}(l|k)\big] + \mathbb{E}_{\hat{\lambda}}(l-1)}{\Big(1 - \mathbb{E}_{\hat{\mu}}\big[\mathbb{E}_{\lambda}(l|k)\big]\Big)\Big(1 - \mathbb{E}_{\hat{\lambda}}\big[\mathbb{E}_{\mu}(k|l)\big]\Big) - \mathbb{E}_{\hat{\mu}}(k-1) \cdot \mathbb{E}_{\hat{\lambda}}(l-1)} \tag{1.8}$$

$$n_2 = \frac{1 - \mathbb{E}_{\hat{\lambda}}\big[\mathbb{E}_{\mu}(k|l)\big] + \mathbb{E}_{\hat{\mu}}(k-1)}{\Big(1 - \mathbb{E}_{\hat{\mu}}\big[\mathbb{E}_{\lambda}(l|k)\big]\Big)\Big(1 - \mathbb{E}_{\hat{\lambda}}\big[\mathbb{E}_{\mu}(k|l)\big]\Big) - \mathbb{E}_{\hat{\mu}}(k-1) \cdot \mathbb{E}_{\hat{\lambda}}(l-1)} \tag{1.9}$$

A random component is finite in expectation iff n_1, n_2 are well-defined, that is, positive and finite. From equations (1.8) and (1.9), it is straightforward to see that Assumption 1.1(i) implies n_1, n_2 are well-defined. We now show the other direction.

First, suppose $\mathbb{E}_{\hat{\mu}}(k-1)\cdot\mathbb{E}_{\hat{\lambda}}(l-1) > \big(1-\mathbb{E}_{\hat{\mu}}\big[\mathbb{E}_{\lambda}(l|k)\big]\big)\cdot\big(1-\mathbb{E}_{\hat{\lambda}}\big[\mathbb{E}_{\mu}(k|l)\big]\big)$. Then n_1, n_2 have negative denominators. So they must have negative nominators to be well-defined, which requires:

$$1-\mathbb{E}_{\hat{\mu}}\big[\mathbb{E}_{\lambda}(l|k)\big] < -\mathbb{E}_{\hat{\lambda}}(l-1) \leq 0 \quad \text{and} \quad 1-\mathbb{E}_{\hat{\lambda}}\big[\mathbb{E}_{\mu}(k|l)\big] < -\mathbb{E}_{\hat{\mu}}(k-1) \leq 0.$$

It then implies $\mathbb{E}_{\hat{\mu}}(k-1)\cdot\mathbb{E}_{\hat{\lambda}}(l-1) < \big(1-\mathbb{E}_{\hat{\mu}}\big[\mathbb{E}_{\lambda}(l|k)\big]\big)\cdot\big(1-\mathbb{E}_{\hat{\lambda}}\big[\mathbb{E}_{\mu}(k|l)\big]\big)$, contradicting our assumption.

Now suppose $\mathbb{E}_{\hat{\mu}}(k-1)\cdot\mathbb{E}_{\hat{\lambda}}(l-1) = \big(1-\mathbb{E}_{\hat{\mu}}\big[\mathbb{E}_{\lambda}(l|k)\big]\big)\cdot\big(1-\mathbb{E}_{\hat{\lambda}}\big[\mathbb{E}_{\mu}(k|l)\big]\big)$. We only need to check the case when the numerator for n_2 is zero:

$$1-\mathbb{E}_{\hat{\lambda}}\big[\mathbb{E}_{\mu}(k|l)\big] = -\mathbb{E}_{\hat{\mu}}(k-1), \tag{1.10}$$

for otherwise n_2 explodes. Rearranging equation (1.6) and substituting in equation (1.10), we obtain:

$$\begin{aligned}\Big(1-\mathbb{E}_{\hat{\lambda}}\big[\mathbb{E}_{\mu}(k|l)\big]\Big)\cdot n_1 - \mathbb{E}_{\hat{\lambda}}(l-1)\cdot n_2 = 1 \quad &\Rightarrow \\ -\mathbb{E}_{\hat{\mu}}(k-1)\cdot n_1 - \mathbb{E}_{\hat{\lambda}}(l-1)\cdot n_2 = 1 \quad &\end{aligned} \tag{1.11}$$

Apparently, for equation (1.11) to hold, n_1, n_2 cannot be both positive.

Lastly, suppose $\mathbb{E}_{\hat{\mu}}(k-1)\cdot\mathbb{E}_{\hat{\lambda}}(l-1) < \big(1-\mathbb{E}_{\hat{\mu}}\big[\mathbb{E}_{\lambda}(l|k)\big]\big)\cdot\big(1-\mathbb{E}_{\hat{\lambda}}\big[\mathbb{E}_{\mu}(k|l)\big]\big)$, but $\mathbb{E}_{\hat{\mu}}\big[\mathbb{E}_{\lambda}(l|k)\big] \geq 1$. Together, they imply $\mathbb{E}_{\hat{\mu}}\big[\mathbb{E}_{\lambda}(l|k)\big], \mathbb{E}_{\hat{\lambda}}\big[\mathbb{E}_{\mu}(k|l)\big] > 1$. For n_1, n_2 to be positive, their numerators must be positive:

$$0 > 1-\mathbb{E}_{\hat{\mu}}\big[\mathbb{E}_{\lambda}(l|k)\big] > -\mathbb{E}_{\hat{\lambda}}(l-1) \quad \text{and} \quad 0 > 1-\mathbb{E}_{\hat{\lambda}}\big[\mathbb{E}_{\mu}(k|l)\big] > -\mathbb{E}_{\hat{\mu}}(k-1).$$

However, it implies $\mathbb{E}_{\hat{\mu}}(k-1)\cdot\mathbb{E}_{\hat{\lambda}}(l-1) > \big(1-\mathbb{E}_{\hat{\mu}}\big[\mathbb{E}_{\lambda}(l|k)\big]\big)\cdot\big(1-\mathbb{E}_{\hat{\lambda}}\big[\mathbb{E}_{\mu}(k|l)\big]\big)$. We run into contradiction again.

Hence, overall, if Assumption 1.1(i) is not satisfied, then n_1, n_2 are not well-defined.

Q.E.D.

1.7.2 Proof of Lemma 1.1

Suppose the downward and upward risks $p, r < 1$. We first show that the expected utility function $U^l(n)$ extended on $\mathbb{R}_+$ is quasi-concave. For that, we compute its derivative:

$$\frac{\partial U^l(n)}{\partial n} = (1-q)(1-p)^n(1-r)^l\big[b + (nb+c)\log(1-p)\big].$$

If $p = 0$, it is positive, and so $U^l(n)$ is an increasing function. When $p > 0$, if $c \geq -b/\log(1-p)$, it is nonpositive, thus $U^l(n)$ is nonincreasing; otherwise, it is positive when $n < -1/\log(1-p) - c/b$ and negative when $n > -1/\log(1-p) - c/b$, and so $U^l(n)$ first increases, then decreases. In any case, $U^l(n)$ is quasi-concave.

We now derive an agent's desired number of bribes correspondence $\hat{n}(p)$, which furnishes the proof of the rest of Lemma 1.1. Since the expected utility function $U^l(n)$ is quasi-concave, accepting $n \in \mathbb{N}_+$ bribes is optimal iff $U^l(n) \geq \max\big\{U^l(n+1), U^l(n-1)\big\}$, which yields

$$\frac{b}{c+(n+1)b} \leq p \leq \frac{b}{c+nb}$$

Similarly, declining all bribes is optimal iff $U^l(0) \geq U^l(1)$, which gives $p \geq b/(c+b)$. These conditions characterize the desired number of bribes $\hat{n}$ as a nonincreasing correspondence of the downward risk p. They also suggest that $\hat{n}$ is irrelevant of the degrees k, l and the upward risk r.

Lastly, we incorporate in the left-out boundary cases: $p = 1$ or $r = 1$. Notice that the downward risk $p = 1$ if and only if any subordinate of an agent is monitored by someone else, any agent desires to accept zero bribe (such that the upward risk $r = 1$), and the contagion rate $\eta = 1$. Hence, we must have $\hat{n}(1) = 0$.

Now consider the case when the upward risk $r = 1$, but the downward risk $p < 1$. $r = 1$ if and only if any agent desires to accept no bribe: $\hat{n}(p) = 0$. Hence, this case is captured by the

desired number of bribes correspondence $\hat{n}(p)$ when the downward risk is large: $p \geq b/(c+b)$.

Q.E.D.

1.7.3 Strong Symmetry Assumption

1.7.3.1 Baseline Model

In this section, we divide the set of mixed-strategy *symmetric equilibria* into different groups, each including one and only one *strongly symmetric equilibrium* defined in Assumption 1.2. We then show that all equilibria in the same group share the same expected utility for any agent, such that it is without loss of generality to select only the strongly symmetric equilibrium.

Define a symmetric equilibrium by $\mathbf{n}^* \equiv (n^*_{lk})$, where $n^*_{lk} \in \mathbb{R}_+$ is the number of bribes an agent with l supervisors and k subordinates accepts. We only consider the symmetric equilibria where some agents adopt strictly mixed strategies (pure-strategy symmetric equilibria are strongly symmetric), and classify them in the following way:

Definition 1.3. *An* ***equivalence class*** $\mathcal{C}(m, x_m)$ *is the set of symmetric equilibria* $\mathbf{n}^*$ *for which the associated downward risk* $p^* = p_{m(m+1)}$ *and* $\mathbb{E}_\pi[n^*_{lk}|k > m] = x_m \in (m, m+1)$.

In words, at any equilibrium in an equivalence class $\mathcal{C}(m, x_m)$, agents are indifferent between accepting m and $(m+1)$ bribes, and the expected number of bribes agents with relaxed capacity constraints accept is x_m. $\mathcal{C}(m, x_m)$ must include a unique strongly symmetric equilibrium $n^* = x_m$.

The following proposition establishes the similarity between all equilibria in an equivalence class, justifying our sole selection of the strongly symmetric one.

Proposition 1.9. *All equilibria in the same equivalence class entail the same upward risk and thus the same expected utilities.*

Proof: We first show that all equilibria in an equivalence class $\mathcal{C}(m, x_m)$ share the same

upward risk. Consider any equilibrium $\mathbf{n}^* \in \mathcal{C}(m, x_m)$. Its upward risk is expressed as:[46]

$$\begin{aligned} R(\mathbf{n}^*) &= \sum_l \sum_{k\geq 1} \Big(1 - \frac{n^*_{lk}}{k}\Big)\hat{\mu}(k)\lambda(l|k) \\ &= \sum_l \sum_{k\geq 1} \frac{k - n^*_{lk}}{k} \cdot \frac{k\mu(k)}{\bar{k}}\lambda(l|k) \quad \big(\bar{k} \equiv \mathbb{E}_\mu(\tilde{k})\big) \\ &= \sum_l \sum_{k>m} \frac{k - n^*_{lk}}{k} \cdot \frac{k\mu(k)}{\bar{k}}\lambda(l|k) \quad (n^*_{lk} = k \ \forall k \leq m) \\ &= \frac{\sum_{k>m} k\mu(k) - x_m \cdot \Pr(k > m)}{\bar{k}} \end{aligned}$$

Since it only relies on (m, x_m), all equilibria in the equivalence class $\mathcal{C}(m, x_m)$ share the same upward risk. Because they also share the same downward risk $p^* = p_{m(m+1)}$, the expected utility function $U^l(n)$ $\big($equation (1.1)$\big)$ is the same across them. An agent with degrees (l, k) thus obtains the same expected utility $\max_{n=0}^{k} U^l(n)$ at any equilibrium in $\mathcal{C}(m, x_m)$.

Q.E.D.

Here we present a numerical example where multiple equilibria exist in an equivalence class. Consider a hierarchy with the out-degree distribution given by $\mu(0) = 1/2$ and $\mu(1) = \mu(2) = 1/4$. Set parameter values: $\eta = 1$, $q = 1/2$, $b = 4/3$, $c = 1$. This game has an equivalence class $\mathcal{C}(0, 1/2)$, which includes, for instance, three equilibria: $(n^*_{01}, n^*_{11}, n^*_{02}, n^*_{12}) = (1, 1, 0, 0)$, $(0, 0, 1, 1)$ and $(1/2, 1/2, 1/2, 1/2)$ (apparently, $n^*_{00} = n^*_{10} = 0$), where the last one is strongly symmetric. It is easy to verify that they share the same downward risk $p^* = p_{01} = 4/7$ and upward risk $r^* = 1/3$. Consequently, agents' expected utilities are also the same across them: -1/2 for those with no supervisor and -2 for those with one supervisor.

1.7.3.2 Model with Criminal and Innocent Agents

This section presents the definition of strongly symmetric equilibrium for the extended model introduced in Section 1.5.2, and show that it is without loss of generality to restrict attention

[46]This expression is adapted from the upward risk function $R(n)$ defined by equation (1.3).

to the strongly symmetric equilibria in a two-layer network – the main focus of study in our comparative statics.

Definition 1.4. *In a strongly symmetric equilibrium,*

i there is $n^* \in \mathbb{R}_+$ *such that an agent with* $j > 0$ *guilty subordinates accepts either* $\min\{j, n^*\}$ *or 0 bribe;*

ii all innocent agents indifferent between accepting bribes and not mix them with the same probabilities.

To establish the generality of strongly symmetric equilibria in a two-layer network, we first group the symmetric equilibria[47] in the following way:

Definition 1.5. *An* ***equivalence class*** $\mathcal{C}(p^*)$ *is the set of symmetric equilibria with the same downward risk* p^**.*

Since in a two-layer network, supervisors who make bribe acceptance decisions are not monitored, their expected utilities are irrelevant to the upward risk r. Thus all equilibria in the same equivalence class $\mathcal{C}(p^*)$ produce the same expected utility for any agent, indicating that they are virtually equivalent. Hence, the following proposition implies that it is without loss of generality to focus only on the strongly symmetric equilibria:

Proposition 1.10. *In a two-layer network, each equivalence class* $\mathcal{C}(p^*)$ *contains at least one strongly symmetric equilibrium.*

Proof: In a two-layer network, subordinates do not make bribe acceptance decisions, so any given subordinate of an agent is guilty if and only if she is criminal: $g = s$, and the distribution of the number of guilty subordinates an agent has μ_g is fixed at μ_s. Hence, the more prone supervisors are to accept bribes, the smaller the upward risk r is, and the smaller the downward risk p is $\big($for this derivation, the fixation of $\langle g, \mu_g \rangle$ at $\langle s, \mu_s \rangle$ is

[47] A symmetric equilibrium is an equilibrium in which agents of the same type (criminal/innocent) and with the same numbers of supervisors and guilty subordinates (l, j) adopt the same mixed strategy.

essential; otherwise, if they vary with agents' strategies, changes in the risk variables (p, r) are ambiguous$\big)$. Therefore, among the optimal symmetric strategies at some equilibrium downward risk p^*, the function P that maps from agents' strategies to the downward risk p obtains its maximum $p_M \geq p^*$ (minimum $p_m \leq p^*$) when any innocent agent indifferent between accepting bribes and not declines all bribes (resp., accepts bribes) with certainty, and all criminal agents and innocent ones who accept bribes desire $\min \hat{n}(p^*)$ $\big($resp., $\max \hat{n}(p^*)\big)$ bribes. Since both the maximizer and the minimizer are strongly symmetric, if they do not coincide $\big($otherwise, they define the unique element of the equivalence class $\mathcal{C}(p^*)\big)$, we can easily construct a continuum of optimal strongly symmetric strategies at the downward risk p^* for which they are the two boundary points and on which the mapping P is continuous.[48] Since the downward risk $p^* \in [p_m, p_M]$, the *Intermediate Value Theorem* suggests that it must be achieved by some optimal strongly symmetric strategies in this domain. By definition, it belongs to the equivalence class $\mathcal{C}(p^*)$.

Q.E.D.

1.7.4 Downward Risk Function $P(n)$

In this section, we show that the downward risk function $P(n)$ solved implicitly by equation (1.4) is well-defined. Denote the RHS of (1.4) as a function:

$$f(n,p) \equiv \eta \cdot \sum_k \sum_{l \geq 1} \Big[1 - (1-q)\big(1 - R(n)\big)^{l-1} \cdot (1-p)^{\lfloor n \wedge k \rfloor} \big(1 - (n \wedge k - \lfloor n \wedge k \rfloor)p\big) \Big] \hat{\lambda}(l)\mu(k|l).$$

[48] For this construction, restriction to a subset of the optimal strongly symmetric strategies at the downward risk p^* is only necessary if the following two types of mixed strategies can coexist: criminal agents may mix between accepting two adjacent numbers of bribes, i.e., $\min \hat{n}(p^*) < \max \hat{n}(p^*)$ and some agents' capacity constraints do not bind; and some innocent agents may mix between accepting bribes and not – there exists some j such that $U^0(\lfloor \hat{n}(p^*) \wedge j \rfloor)|_{p^*} = 0$. In this case, we reduce the space to the optimal strongly symmetric equilibria for which the two types of mixed strategies share the same probabilities: there is $r \in [0,1]$ such that all innocent agents indifferent between accepting bribes and not accept bribes with probability r, and all agents likely to accept bribes desire $\max \hat{n}(p^*)$ bribes with probability r and $\min \hat{n}(p^*)$ bribes with probability $(1-r)$.

Thus the downward risk function $P(n)$ is implicitly solved by

$$f(n,p) - p = 0. \tag{1.12}$$

We first show given any bribe acceptance propensity $n \geq 0$, there must exist a downward risk $p \in (0,1)$ such that equation (1.12) holds. Notice that the function $f(n,p) - p$ is continuous. Since $f(n,p) \in [0,1]$ for any $n \geq 0$ and $p \in [0,1]$, we have $f(n,0) - 0 \geq 0$ and $f(n,1) - 1 \leq 0$ at any $n \geq 0$. Hence, the Intermediate Value Theorem suggests for any $n \geq 0$, there exists $p \in [0,1]$ such that $f(n,p) - p = 0$.

Next, we prove this solution is unique. For that, we need to show $\partial f/\partial p - 1 < 0$ for any $n \geq 0$ and $p \in [0,1]$, such that given any $n \geq 0$, $f(n,p) - p$ strictly decreases on $p \in [0,1]$, and so must cross 0 exactly once.

First, we show $\partial f/\partial p$ decreases on $p \in [0,1]$. Define function

$$g(k,n,p) \equiv \frac{\partial \Big[(1-p)^{\lfloor n \wedge k \rfloor} \big(1 - (n \wedge k - \lfloor n \wedge k \rfloor)p\big) \Big]}{\partial p}$$

Then

$$\frac{\partial f}{\partial p} = -\eta(1-q) \cdot \sum_k \sum_{l \geq 1} \big(1 - R(n)\big)^{l-1} \cdot g(k,n,p) \hat{\lambda}(l) \mu(k|l). \tag{1.13}$$

Now notice that

$$\frac{\partial g}{\partial p} = \begin{cases} 2(n-1) & 1 \leq n < 2 \leq k \\ -\big(\lfloor n \rfloor - 1\big)\big[-n + \big(\lfloor n \rfloor + 1\big)\big(n - \lfloor n \rfloor\big)p\big] \cdot \\ (1-p)^{\lfloor n \rfloor - 2} + \big(\lfloor n \rfloor + 1\big)\big(n - \lfloor n \rfloor\big)(1-p)^{\lfloor n \rfloor - 1} & 2 \leq n < k \\ k(k-1)(1-p)^{k-2} & 2 \leq k \leq n \\ 0 & \text{o.w.} \end{cases} \tag{1.14}$$

where

$$\begin{aligned}
-n + \big(\lfloor n \rfloor + 1\big)\big(n - \lfloor n \rfloor\big)p &\leq -n + \big(\lfloor n \rfloor + 1\big)\big(n - \lfloor n \rfloor\big) \\
&= n \cdot \lfloor n \rfloor - \big(\lfloor n \rfloor\big)^2 - \lfloor n \rfloor \\
&= \big(n - \lfloor n \rfloor - 1\big) \cdot \lfloor n \rfloor \\
&\leq 0.
\end{aligned}$$

Therefore, we have $\partial g/\partial p \geq 0$ for any $k, n \geq 0$ and $p \in [0,1]$. Thus equation (1.13) suggests $\partial^2 f/\partial p^2 \leq 0$ for any $n \geq 0$, $p \in [0,1]$, and so give any $n \geq 0$, $\partial f/\partial p$ decreases on $p \in [0,1]$.

It remains to show $(\partial f/\partial p)|_{p=0} < 1$. Equation (1.14) implies

$$g(k,n,0) = \begin{cases} -k & k \leq n \\ -n & k > n \end{cases}$$

Hence,

$$\begin{aligned}
\left.\frac{\partial f}{\partial p}\right|_{p=0} &= \eta(1-q) \cdot \sum_k \sum_{l \geq 1} \big(1 - R(n)\big)^{l-1} (n \wedge k) \cdot \hat{\lambda}(l) \mu(k|l) \\
&\leq \eta(1-q) \cdot \sum_k \sum_{l \geq 1} k \cdot \hat{\lambda}(l) \mu(k|l) \\
&< \eta(1-q) \\
&< 1,
\end{aligned}$$

where the second inequality holds because Assumption 1.1(i) (the network finiteness assumption) implies $\mathbb{E}_{\hat{\lambda}}\big[\mathbb{E}_{\mu}(k|l)\big] < 1$.

Hence, $\partial f/\partial p - 1 < 0$ for all $n \geq 0$ and $p \in [0,1]$. So $f(n,p) - p = 0$ has a unique solution $p \in [0,1]$ at any $n \geq 0$, and thus the downward risk function $P(n)$ is well-defined.

1.7.5 Proof of Proposition 1.1

We show that the downward risk function $P(n)$ is strictly monotonic on $[0, \bar{k}]$ given a hierarchy or a two-layer network, where $\bar{k} \equiv \operatorname{supp} \mu$ is the largest number of bribes an agent can accept.

In a two-layer network, from equation (1.4), the downward risk function can be explicitly expressed as:

$$P(n) = \eta \cdot \sum_{l \geq 1} \Big[1 - (1-q)\big(1 - R(n)\big)^{l-1}\Big] \hat{\lambda}(l).$$

Since the upward risk function $R(n)$ is strictly increasing on $[0, \bar{k}]$, as easily seen from equation (1.3), so is the downward risk function $P(n)$.[49]

In a hierarchy, the downward risk function $P(n)$ is implicitly solved by $p = f(n, p)$, where the function $f(n, p)$ is given by:

$$f(n,p) = \eta \cdot \sum_{k} \Big[1 - (1-q)(1-p)^{\lfloor n \wedge k \rfloor}\big(1 - (n \wedge k - \lfloor n \wedge k \rfloor)p\big)\Big] \mu(k|l=1). \tag{1.15}$$

Given any two bribe acceptance propensities n_1, n_2 with $0 \leq n_1 < n_2 \leq \bar{k}$, we want to show $P(n_1) < P(n_2)$. Denote them as p_1, p_2. From equation (1.15), we get $f(n, 0) = \eta \cdot q > 0$ (remember the contagion rate $\eta > 0$, and the probability of being directly caught $q > 0$). Hence, $P(n) > 0$ for any $n \geq 0$,[50] and so $p_1 > 0$. Since (1.15) also implies $f(n, p)$ strictly increases on $n \in [0, \bar{k}]$ at any $p \in (0, 1)$, we have $f(n_2, p_1) > f(n_1, p_1) = {p_1}$[50]. Since $f(n, p)$ is continuous, and $f(n_2, 1) - 1 \leq 0$, the Intermediate Value Theorem suggests there exists some $p \in (p_1, 1]$ such that $f(n_2, p) = p$, which by definition must be equal to p_2.[50] In other words, $P(n_1) < P(n_2)$.

Q.E.D.

---

[49] We rule out the trivial network where each supervisor monitors one subordinate, which gives $\hat{\lambda}(1) = 1$.

[50] Here we are using the fact that $p = f(n, p)$ has a unique solution $p \in [0, 1]$ for any $n \geq 0$. See Appendix 1.7.4 for the proof.

1.7.6 Equilibrium Selection

In the baseline model, in face of equilibrium multiplicity, we select the largest equilibrium. More precisely, denote by N^* the set of all strongly symmetric equilibria given certain parameter values. The largest equilibrium $n_M^* = \max N^*$. Here we justify this selection by showing the following:

Proposition 1.11. *The largest equilibrium n_M^* maximizes any agent's expected utility out of all the equilibria in N^*.*

Proof: When there are multiple equilibria, pick any two $n_1^*, n_2^* \in N^*$ that satisfy $n_1^* > n_2^*$. We want to show that any agent obtains higher expected utility at n_1^* than at n_2^*. Denote the downward and upward risks at n_i^* by p_i^*, r_i^*. Since the desired number of bribes correspondence $\hat{n}(p)$ is nonincreasing, we have $p_1^* \leq p_2^*$. Moreover, the strictly decreasing upward risk function $R(n)$ implies $r_1^* < r_2^*$. Therefore, the expected utility for an agent with l supervisors and k subordinates satisfies $U^l\big(\lfloor n_1^* \wedge k \rfloor\big)|_{p_1^*, r_1^*} \geq U^l\big(\lfloor n_2^* \wedge k \rfloor\big)|_{p_1^*, r_1^*} \geq U^l\big(\lfloor n_2^* \wedge k \rfloor\big)|_{p_2^*, r_2^*}$, such that the larger equilibrium n_1^* generates larger expected utility than n_2^* does. The first inequality holds because n_1^* is the optimal choice at downward risk p_1^*. The second is true for $p_1^* \leq p_2^*$ and $r_1^* < r_2^*$.

Q.E.D.

1.7.7 Proof of Proposition 1.3

Given two network degree distributions π_1, π_2 such that π_1 is denser than π_2, we want to show their downward risk functions satisfy $P_1(n) \geq P_2(n)$.

We first show the upward risk function $R_1(n) \geq R_2(n)$. Notice that $\mu_1 \succeq_{MLRP} \mu_2$ implies $\hat{\mu}_1 \succeq_{FOSD} \hat{\mu}_2$. Since $1 - (n \wedge k)/k$ is an increasing function of k, equation (1.3) suggests $R_1(n) \geq R_2(n)$ at any $n \geq 0$.

The downward risk function $P(n)$ is implicitly solved by equation (1.4), which we rear-

range as:

$$p = \eta - \eta(1-q) \cdot \mathbb{E}_{\hat{\lambda}}\big(1 - R(n)\big)^{l-1} \cdot \mathbb{E}_{\mu}\big[h(k,n,p)|l \geq 1\big] \tag{1.16}$$
$$\text{where} \quad h(k,n,p) \equiv (1-p)^{\lfloor n \wedge k \rfloor}\big(1 - (n \wedge k - \lfloor n \wedge k \rfloor)p\big).$$

Notice that we have used the assumption $\mu(\cdot|\tilde{l} = l) = \mu(\cdot|\tilde{l} \geq 1)\ \forall l \geq 1$.

Since the function $h(k,n,p)$ decreases in k and $\mu_1(\cdot|l \geq 1) \succeq_{FOSD} \mu_2(\cdot|l \geq 1)$, we have

$$\mathbb{E}_{\mu_1}\big[h(k,n,p)|l \geq 1\big] \leq \mathbb{E}_{\mu_2}\big[h(k,n,p)|l \geq 1\big] \tag{1.17}$$

at any $n \geq 0$ and $p \in [0,1]$. Since $\lambda_1 \succeq_{MLRP} \lambda_2$ implies $\hat{\lambda}_1 \succeq_{FOSD} \hat{\lambda}_2$, and $\big(1 - R(n)\big)^{l-1}$ is a decreasing function of l, we have

$$\mathbb{E}_{\hat{\lambda}_1}\big(1 - R_1(n)\big)^{l-1} \leq \mathbb{E}_{\hat{\lambda}_2}\big(1 - R_1(n)\big)^{l-1} \leq \mathbb{E}_{\hat{\lambda}_2}\big(1 - R_2(n)\big)^{l-1} \tag{1.18}$$

at any $n \geq 0$, where the second inequality holds because $R_1(n) \geq R_2(n)$. Denote the RHS of equation (1.16) as function $f(n,p)$. Then (1.17-1.18) suggest $f_1(n,p) \geq f_2(n,p)$ at any $n \geq 0$ and $p \in [0,1]$.

Fix any $n \geq 0$. Denote the corresponding downward risks for the two network degree distributions $P_1(n), P_2(n)$ as p_1, p_2. Since $f_1(n,p_2) \geq f_2(n,p_2) = {p_2}^{50}$, $f_1(n,1) - 1 \leq 0$, and f_1 is continuous in p, the Intermediate Value Theorem suggests there exists some $p \in [p_2, 1]$ such that $f_1(n,p) - p = 0$, which by definition must be equal to p_1.[50] In other words, $P_1(n) \geq P_2(n)$ at any n. Therefore, the equilibrium bribe acceptance propensity $n_1^* \leq n_2^*$.

Q.E.D.

1.7.8 Comovement of n^* and κ^*

In this section, we formalize the comovement of the bribe acceptance propensity n^* and the corruption level κ^* in a hierarchy in the baseline model, and generalize it to any network

with independent in- and out-degrees.

Proposition 1.12. *In a hierarchy with independent in- and out-degrees ($\lambda \perp \mu$), if the bribe acceptance propensity n^** ***strictly*** *drops with a local FOSD shift of the out-degree distribution μ, so does the corruption level κ^*.*

Proof: Given a local FOSD shift of the out-degree distribution μ, the downward risk function $P(n)$ shifts right, causing the bribe acceptance propensity n^* to drop. Observe that it strictly drops only if the downward risk p^* is fixed at some cutoff $p_{m(m+1)} \in (0,1)$, where $m \in \mathbb{N}$ and the bribe acceptance propensity $n^* \in (m, m+1]$. Therefore, employing equation (1.4) that defines the downward risk function $P(n)$, we get:

$$\begin{aligned} p^* &= \eta \cdot \sum_k \Big[1 - (1-q)(1-p^*)^{\lfloor n^* \wedge k \rfloor}\big(1 - (n^* \wedge k - \lfloor n^* \wedge k \rfloor)p^*\big)\Big]\mu(k) \\ &= \eta \sum_{k \le m} \Big[1 - (1-q)(1-p^*)^k\Big]\mu(k) \\ &+ \eta \sum_{k > m} \Big[1 - (1-q)(1-p^*)^m\big(1 - (n^* - m)p\big)\Big]\mu(k). \end{aligned}$$

Rearranging it, we obtain:

$$\sum_{k>m} n^*\mu(k) = \frac{1}{p^*}\sum_{k \le m}(1-p^*)^{k-m}\mu(k) + \frac{mp^*+1}{p^*}\sum_{k>m}\mu(k) + C, \tag{1.19}$$
$$\text{where} \quad C = \frac{p^* - \eta}{\eta(1-q)p^*(1-p^*)^m}$$

Since the corruption level $\kappa^* = \mathbb{E}_\mu\big[n^* \wedge k\big] = \sum_{k \le m} k\mu(k) + \sum_{k>m} n^*\mu(k)$, plugging (1.19) in, it can be expressed as:

$$\begin{aligned} \kappa^* &= \mathbb{E}_\mu\big[g(k)\big] + C, \\ \text{where} \quad g(k) &= \mathbb{1}\{k \le m\} \cdot \Big(k + \frac{(1-p^*)^{k-m}}{p^*}\Big) + \mathbb{1}\{k > m\} \cdot \Big(m + \frac{1}{p^*}\Big) \end{aligned}$$

It is easy to verify that the function $f(k) \equiv k + (1-p^*)^{k-m}/p^*$ is strictly decreasing on $[0, m]$ and $f(m) = m + 1/p^*$, such that $g(k)$ is decreasing. Hence, a local FOSD shift of the out-

degree distribution μ decreases the corruption level κ^*. κ^* strictly decreases because $f(k)$ is strictly decreasing and $\mu(k) > 0$ for some $k \leq m$. To see that, notice that agents' bribe acceptance propensity n^* strictly drops with an FOSD shift of the out-degree distribution μ only if some agents have binding capacity constraints, such that they indeed accept more bribes when offered more, rendering bribe acceptance strictly more dangerous.

Q.E.D.

Proposition 1.12 can be generalized to any network with independent in- and out-degrees: $\lambda \perp \mu$. Formally, if some equilibrium bribe acceptance propensity n^* (not necessarily the largest one) is in $(m, m+1)$ for some $m \in \mathbb{N}$, and the derivative of the downward risk function $P'(n^*) > 0$, then given a local FOSD shift of the out-degree distribution μ, both n^* and the corresponding corruption level κ^* fall. κ^* falls strictly whenever n^* does.

The bribe acceptance propensity n^* falls because the downward risk function $P(n)$ shifts right when the out-degree distribution μ increases.[51] To see why the corruption level κ^* falls, notice that, just like in a hierarchy, it can be expressed as $\kappa^* = \mathbb{E}_{\mu}\big[g(k)\big] + C(n^*)$, where $g(k)$ is the same as before, and

$$C(n^*) = \frac{p^* - \eta}{\eta(1-q)p^*(1-p^*)^m \cdot \mathbb{E}_{\hat{\lambda}}\big[1 - R(n^*)\big]^{l-1}}$$

Since the upward risk function $R(n)$ is decreasing, and the downward risk p^* is by definition no larger than the contagion rate η, $C(n^*) \leq 0$ falls given a smaller bribe acceptance propensity n^*. The rest of the proof follows that for Proposition 1.12.

1.7.9 Proof of Proposition 1.4

Proof for 4(i): Given a joint degree distribution for two-layer networks π_1, we perform a mean-preserving spread on unmonitored agents' out-degree distribution $\mu_1(\cdot|l = 0)$ to obtain

[51] For this result, the assumption $P'(n^*) > 0$ – the downward risk function $P(n)$ is strictly increasing at n^* – is necessary. Otherwise, if $P'(n^*) < 0$, then the bribe acceptance propensity $n^* \in (m, m+1)$ in fact rises when the downward risk function $P(n)$ shifts right.

a new degree distribution π_2. We want to show the upward risk function increases after the shift: $R_2(n) \geq R_1(n)$.

First, notice that the out-degree distribution for a supervisor of an agent can be written as:

$$\hat{\mu}(k) = \frac{k \cdot \mu(k)}{\mathbb{E}_\mu(\tilde{k})} = \frac{k \cdot \Pr\left(\tilde{k} = k | \tilde{k} > 0\right)}{\mathbb{E}_\mu\big(\tilde{k} | \tilde{k} > 0\big)}$$

where the second equality holds because $\mu(k) = \Pr\left(\tilde{k} = k | \tilde{k} > 0\right) \cdot \Pr\left(\tilde{k} > 0\right)$ for any $k > 0$, and $\mathbb{E}_\mu(\tilde{k}) = \mathbb{E}_\mu\big(\tilde{k} | \tilde{k} > 0\big) \cdot \Pr\left(\tilde{k} > 0\right)$.

Hence, we can rearrange the upward risk function as follows:

$$\begin{aligned} R(n) &= \sum_{k \geq 1} \Big(1 - \frac{n \wedge k}{k}\Big) \hat{\mu}(k) \\ &= 1 - \sum_{k \geq 1} \frac{n \wedge k}{k} \cdot \frac{k \cdot \Pr\left(\tilde{k} = k | \tilde{k} > 0\right)}{\mathbb{E}_\mu\big(\tilde{k} | \tilde{k} > 0\big)} \\ &= 1 - \frac{\mathbb{E}_\mu\big(n \wedge \tilde{k} | \tilde{k} > 0\big)}{\mathbb{E}_\mu\big(\tilde{k} | \tilde{k} > 0\big)} \end{aligned} \tag{1.20}$$

Since $\mu_2(\cdot | l = 0) \preccurlyeq_{SOSD} \mu_1(\cdot | l = 0)$, by definition of SOSD, $\sum_{k=0}^{\bar{k}} \Pr\left(\tilde{k} \leq k | l = 0\right)|_{\mu_2} \geq \sum_{k=0}^{\bar{k}} \Pr\left(\tilde{k} \leq k | l = 0\right)|_{\mu_1}$ for any out-degree $\bar{k}$. Letting $\bar{k} = 0$, we also obtain $\mu_2(0 | l = 0) \geq \mu_1(0 | l = 0)$.

Now, notice that

$$\begin{aligned} \Pr\left(\tilde{k} \leq k | \tilde{k} > 0\right)|_{\mu_2} &= \Pr\left(\tilde{k} \leq k | \tilde{k} > 0, l = 0\right)|_{\mu_2} \\ &= \frac{\Pr\left(0 < \tilde{k} \leq k | l = 0\right)|_{\mu_2}}{\Pr\left(\tilde{k} > 0 | l = 0\right)|_{\mu_2}} \\ &= \frac{\Pr\left(\tilde{k} \leq k | l = 0\right)|_{\mu_2} - \mu_2(0 | l = 0)}{1 - \mu_2(0 | l = 0)} \\ &= 1 + \frac{\Pr\left(\tilde{k} \leq k | l = 0\right)|_{\mu_2} - 1}{1 - \mu_2(0 | l = 0)} \end{aligned}$$

Hence, given any out-degree $\bar{k}$,

$$\begin{aligned}\sum_{k=0}^{\bar{k}} \Pr\big(\tilde{k} \leq k|\tilde{k} > 0\big)|_{\mu_2} &= \bar{k} + 1 + \frac{\sum_{k=0}^{\bar{k}} \Pr\big(\tilde{k} \leq k|l = 0\big)|_{\mu_2} - (\bar{k}+1)}{1 - \mu_2(0|l=0)} \\ &\geq \bar{k} + 1 + \frac{\sum_{k=0}^{\bar{k}} \Pr\big(\tilde{k} \leq k|l = 0\big)|_{\mu_1} - (\bar{k}+1)}{1 - \mu_1(0|l=0)} \\ &= \sum_{k=0}^{\bar{k}} \Pr\big(\tilde{k} \leq k|\tilde{k} > 0\big)|_{\mu_1}\end{aligned}$$

In other words, $\Pr\big(\tilde{k} = \cdot|\tilde{k} > 0\big)|_{\mu_2} \preccurlyeq_{SOSD} \Pr\big(\tilde{k} = \cdot|\tilde{k} > 0\big)|_{\mu_1}$. Since $(n \wedge k)$ is a concave function of the out-degree k, it implies $\mathbb{E}_{\mu_2}\big(n \wedge \tilde{k}|\tilde{k} > 0\big) \leq \mathbb{E}_{\mu_1}\big(n \wedge \tilde{k}|\tilde{k} > 0\big)$.

Also, $\mathbb{E}_{\mu_2}\big(\tilde{k}|l = 0\big) = \mathbb{E}_{\mu_1}\big(\tilde{k}|l = 0\big)$ implies

$$\begin{aligned}\mathbb{E}_{\mu_2}\big(\tilde{k}|\tilde{k} > 0\big) &= \mathbb{E}_{\mu_2}\big(\tilde{k}|\tilde{k} > 0, l = 0\big) \\ &= \frac{\mathbb{E}_{\mu_2}\big(\tilde{k}|l = 0\big)}{1 - \mu_2(0|l=0)} \\ &\geq \frac{\mathbb{E}_{\mu_1}\big(\tilde{k}|l = 0\big)}{1 - \mu_1(0|l=0)} \\ &= \mathbb{E}_{\mu_1}\big(\tilde{k}|\tilde{k} > 0\big).\end{aligned}$$

Therefore, from equation (1.20), we obtain that the upward risk function increases after the shift:

$$R_2(n) = 1 - \frac{\mathbb{E}_{\mu_2}\big(n \wedge \tilde{k}|\tilde{k} > 0\big)}{\mathbb{E}_{\mu_2}\big(\tilde{k}|\tilde{k} > 0\big)} \geq 1 - \frac{\mathbb{E}_{\mu_1}\big(n \wedge \tilde{k}|\tilde{k} > 0\big)}{\mathbb{E}_{\mu_1}\big(\tilde{k}|\tilde{k} > 0\big)} = R_1(n).$$

Thus, the downward risk function also increases: $P_2(n) \geq P_1(n)$, and so agents' bribe accep-

tance propensity drops: $n_2^* \leq n_1^*$. Additionally, the corruption level falls:

$$\begin{aligned}
\kappa_2^* &= \mathbb{E}_{\mu_2}\big(n_2^* \wedge \tilde{k}\big) \\
&\leq \mathbb{E}_{\mu_2}\big(n_1^* \wedge \tilde{k}\big) \\
&= \mathbb{E}_{\mu_2}\big(n_1^* \wedge \tilde{k}\big| l = 0\big) \cdot \lambda(0) \\
&\leq \mathbb{E}_{\mu_1}\big(n_1^* \wedge \tilde{k}\big| l = 0\big) \cdot \lambda(0) \\
&= \kappa_1^*.
\end{aligned}$$

Q.E.D.

Proof for 4(ii): Define function

$$f_n(x,y) \equiv \begin{cases} (1-x)^{n \wedge y} & n \in \mathbb{N} \\ (1-x)^{\lfloor n \rfloor \wedge y} \cdot \big(1 - (n \wedge y - \lfloor n \rfloor \wedge y)x\big) & n \in \mathbb{R}_+/\mathbb{N} \end{cases}$$

where n is the bribe acceptance propensity.

In a hierarchy with independent in- and out-degrees ($\lambda \perp \mu$), the downward risk function $P(n)$ is recursively defined by:

$$\begin{aligned}
p &= \eta \cdot \sum_k \Big[1 - (1-q) f_n(p,k)\Big] \mu(k) \\
&= \eta \cdot \Big\{1 - (1-q)\mathbb{E}_\mu\big[f_n(p,k)\big]\Big\}
\end{aligned} \tag{1.21}$$

We want to show the function $f_n(x,y)$ is convex in $y \in \mathbb{R}_+$ given any $n \in \mathbb{R}_+$ and $x \in (0,1)$. This is obvious when $n \in \mathbb{N}$. If $n \in \mathbb{R}_+/\mathbb{N}$, we compute the partial derivative of $f_n(x,y)$ w.r.t. y:

$$\frac{\partial f_n}{\partial y} = \begin{cases} \ln(1-x) \cdot (1-x)^y & y < \lfloor n \rfloor^{52} \\ -x \cdot (1-x)^{\lfloor n \rfloor} & \lfloor n \rfloor < y < n \\ 0 & y > n \end{cases}$$

It strictly increases on $\big[0, \lfloor n \rfloor\big)$ and remains constant on $\big(\lfloor n \rfloor, n\big)$ and (n, ∞). Notice that $\partial f_n / \partial y$ is not well-defined at $y = \lfloor n \rfloor$ and $y = n$. However, since $\lim_{y \to \lfloor n \rfloor^-} \partial f_n / \partial y < \lim_{y \to \lfloor n \rfloor^+} \partial f_n / \partial y$ and $\lim_{y \to n^-} \partial f_n / \partial y < \lim_{y \to n^+} \partial f_n / \partial y$, they are two wedge points. Therefore, $f_n(x, y)$ is convex in $y \in \mathbb{R}_+$, and so given any downward risk $p \in (0, 1)$, bribe acceptance propensity $n \in \mathbb{R}_+$ and two out-degree distributions μ_1, μ_2 such that $\mu_2 \preccurlyeq_{SOSD} \mu_1$, we have $\mathbb{E}_{\mu_2}\big[f_n(p, k)\big] \geq \mathbb{E}_{\mu_1}\big[f_n(p, k)\big]$.

Now we include in the two endpoints for the downward risk p. If $p = 0$, then $f_n(0, k) = 1$ for any out-degree k, thus $\mathbb{E}_{\mu_2}\big[f_n(0, k)\big] = \mathbb{E}_{\mu_1}\big[f_n(0, k)\big] = 1$. If $p = 1$, then $\mathbb{E}_{\mu}\big[f_n(1, k)\big] = \mu(0) + \big[1 - \mu(0)\big] \cdot \max\{0, 1 - n\}$. Since $\mu_2 \preccurlyeq_{SOSD} \mu_1$ implies $\mu_2(0) \geq \mu_1(0)$, we have $\mathbb{E}_{\mu_2}\big[f_n(1, k)\big] \geq \mathbb{E}_{\mu_1}\big[f_n(1, k)\big]$.

Thus $\mathbb{E}_{\mu_2}\big[f_n(p, k)\big] \geq \mathbb{E}_{\mu_1}\big[f_n(p, k)\big]$ holds for all $p \in [0, 1]$ and $n \in \mathbb{R}_+$. Equation (1.21) then implies that the downward risk function $P_2(n) \leq P_1(n)$. Hence, the equilibrium bribe acceptance propensity $n_2^* \geq n_1^*$.

Q.E.D.

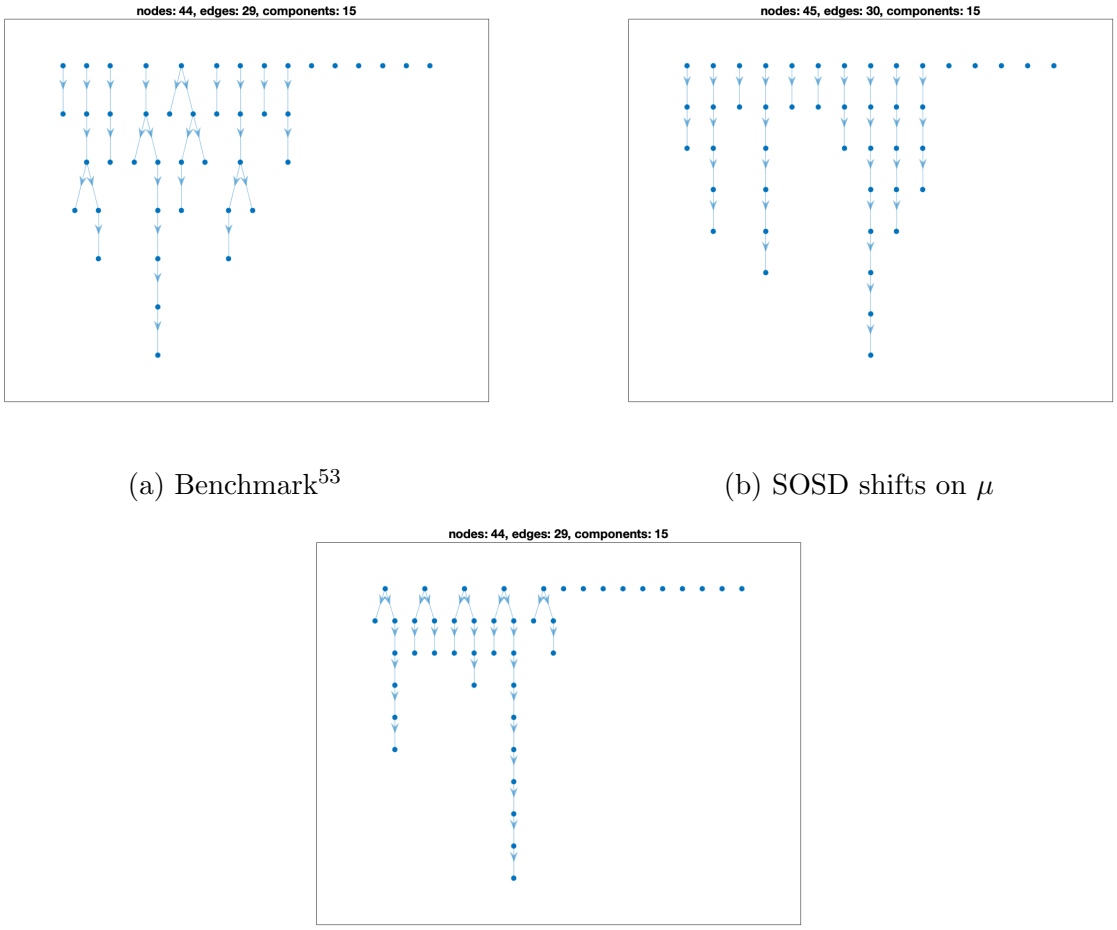
1.7.10 Second-Order Shifts of $\mu(\cdot|l \geq 1)$

This section explains why altering monitored agents' out-degree distribution $\mu(\cdot|l \geq 1)$ in a hierarchy brings about consistent changes in agents' bribe acceptance propensity n^* and the corruption level κ^*. We then supply the formal statement generalized to any networks, and argue that it offers an effective way in network design of separating the opposite effects second-order changes in out-degree distributions bring to hierarchies and two-layer networks.

We illustrate the difference between altering the unconditional out-degree distribution μ and that for monitored agents $\mu(\cdot|l \geq 1)$ in a hierarchy through the following example. Starting with a hierarchical network where an agent can monitor zero, one or two subordinates (figure 1.12a), we shift the out-degree distribution μ in the SOSD manner (the opposite way of a mean-preserving spread), such that agents are now arranged into linear hierarchies (fig-

[52] This range only exists if $n \geq 1$.

ure 1.12b). Though this alteration lowers the bribe acceptance propensity n^*, it is subject to ambiguous changes in the corruption level κ^*. Instead, a more robust improvement would be to fix the unconditional out-degree distribution μ, and only perform an SOSD shift on that for monitored agents $\mu(\cdot|l=1)$ $\big($it implies a simultaneous spread of top agents' out-degree distribution $\mu(\cdot|l=0)\big)$. In the resulting network, while a top agent monitors either zero or two subordinates, the lower-tier structures are linear which maximize risk transmission (figure 1.12c). Hence, not only does this alteration reduce agents' bribe acceptance propensity n^* to the same extent as in 1.12b, but the fixation of the out-degree distribution μ also suggests falling corruption level κ^*.



(a) Benchmark[53]

(b) SOSD shifts on μ

(c) SOSD shifts on $\mu(\cdot|l=1)$

Figure 1.12: Second-Order Variations in Out-Degree Distributions for a Hierarchical Network

This result can be generalized to any network, as exhibited below:

Proposition 1.13. *Fixing the unconditional out-degree distribution μ, a mean-preserving spread of monitored agents' out-degree distribution $\mu(\cdot|l \geq 1)$* ***increases*** *the bribe acceptance propensity n^* and the corruption level κ^*.*

Proposition 1.13 also provides an effective way of separating the opposite forces manifested in hierarchies and two-layer networks in network design. In a general network, since an agent can have both indirect subordinates and co-supervisors, while an SOSD shift of the out-degree distribution μ (the opposite way of a mean-preserving spread) makes it easier for the risk of being caught to transmit up the network (as in a hierarchy), it also renders a subordinate less likely to be reported by co-supervisors (as in a two-layer network), leading to ambiguous implications on the bribe taking risk. However, by shifting monitored agents' out-degree distribution $\mu(\cdot|l \geq 1)$ in the SOSD manner while fixing the unconditional one μ (and so a supervisor's out-degree distribution $\hat{\mu}$), we can facilitate risk transmission through the network without altering co-supervisors' incentives. Thus, bribe taking becomes unambiguously riskier – the downward risk function $P(n)$ shifts right, making corruption less rampant.

1.7.11 Expressions for Extended Models

1.7.11.1 Model with Corruptible and Incorruptible Agents

Risk Functions $P(n), R(n)$

The upward risk function is:

$$R(n) = \sum_{k \geq 1} \left\{ (1-\gamma) + \gamma\Big(1 - \frac{n \wedge k}{k}\Big) \right\} \hat{\mu}(k).$$

[53]The networks are simulated using MATLAB. The parameters are given as follows: $\lambda(0) = 1/3, \lambda(1) = 2/3,\ \mu(0) = 4/9, \mu(1) = 4/9, \mu(2) = 1/9,\ \lambda \perp \mu$ for 1.12a; $\lambda(0) = 1/3, \lambda(1) = 2/3,\ \mu = \lambda,\ \lambda \perp \mu$ for 1.12b; $\lambda(0) = 1/3, \lambda(1) = 2/3,\ \mu(0|l=0) = 2/3, \mu(2|l=0) = 1/3,\ \mu(0|l=1) = 1/3, \mu(1|l=1) = 2/3$ for 1.12c.

To understand the expression, suppose a supervisor of an agent has k subordinates. If the supervisor is incorruptible $\big($with probability $(1-\gamma)\big)$, she never accepts bribes and so reports the agent with certainty; if she is corruptible (with probability γ), she accepts $(n \wedge k)$ out of the k bribe offers, thus the chance the agent is reported by her is $\big(1-(n \wedge k)/k\big)$.

The downward risk function $P(n)$ is recursively defined by:

$$p = \eta \cdot \sum_{k} \sum_{l \geq 1} \Bigg\{ 1 - (1-q)\big(1-R(n)\big)^{l-1} \Big[(1-\gamma) + \gamma \cdot (1-p)^{\lfloor n \wedge k \rfloor} \big(1 - (n \wedge k - \lfloor n \wedge k \rfloor)p\big) \Big] \Bigg\} \hat{\lambda}(l)\mu(k|l).$$

An agent is caught through accepting a subordinate's bribe if and only if the subordinate is caught, and their bribe transaction is subsequently detected (with chance η). Consider a subordinate with degrees (l, k). Conditional on the agent's having accepted her bribe, if the subordinate is incorruptible $\big($with chance $(1-\gamma)\big)$, she remains safe if not directly caught or being reported, that is, with probability $(1-q)\big(1-R(n)\big)^{l-1}$; otherwise, if the subordinate is corruptible (with chance γ), since she may also be caught through accepting bribes, her chance of being safe is further discounted by $(1-p)^{\lfloor n \wedge k \rfloor}\big(1-(n \wedge k - \lfloor n \wedge k \rfloor)p\big)$.

As in the baseline model, the network finiteness condition $\big($Assumption 1.1$(i)\big)$ ensures that the downward risk function $P(n)$ is well-defined.

1.7.11.2 Model with Criminal and Innocent Agents[54]

Statistics for the Network of Guilty Agents $\langle g, \mu_g \rangle$

The probability any given subordinate of an agent is guilty g and the distribution of the number of guilty subordinates an agent has conditional on her having l supervisors $\mu_g(\cdot|l)$

[54]For ease of exposition, these expressions disregard the likelihood for innocent agents to mix between accepting zero and some positive number of bribes.

are jointly solved by the following two equations:

$$g = s + (1-s)\sum_{j}\sum_{l\geq 1}\mathbb{X}\{j \geq \underline{j}^{l}\}\cdot\hat{\lambda}(l)\mu_{g}(j|l), \tag{1.22}$$

$$\mu_{g}(j|l) = \sum_{k\geq j} C_{j}^{k} g^{j}(1-g)^{k-j}\mu(k|l) \quad \forall j,l. \tag{1.23}$$

(1.22) says that a random subordinate is guilty if she is either criminal (with probability s), or innocent and corrupt, where $\mathbb{X}\{j \geq \underline{j}^{l}\}$ is the indicator that a subordinate with l supervisors and j guilty subordinates is corrupt. To understand (1.23), notice that if an agent has k subordinates, $j \leq k$ of them are guilty with probability $C_{j}^{k} g^{j}(1-g)^{k-j}$.

Since any component of the monitoring network is finite $\big($Assumption 1.1$(i)\big)$, equations (1.22-1.23) pin down the unique solutions for g and $\mu_{g}(\cdot|l)$ for each l.

Hence, the unconditional distribution of the number of guilty subordinates an agent has μ_{g} is derived from:

$$\mu_{g}(j) = \sum_{k\geq j} C_{j}^{k} g^{j}(1-g)^{k-j}\mu(k) \quad \forall j.$$

Risk Functions P, R

The upward risk function is expressed as:

$$\begin{aligned} R\big(n,(\underline{j}^{l})_{l}\big) = \sum_{l}\sum_{j\geq 1}\Big\{ & s\Big(1-\frac{n\wedge j}{j}\Big) + (1-s)\cdot \\ & \Big(1-\mathbb{X}\{j\geq \underline{j}^{l}\}\cdot\frac{n\wedge j}{j}\Big)\Big\}\hat{\mu}_{g}(j)\lambda_{g}(l|j), \\ \text{where} \quad \hat{\mu}_{g}(j) \equiv \frac{j\cdot\mu_{g}(j)}{\mathbb{E}_{\mu_{g}}(\tilde{j})} \quad & \text{and} \quad \lambda_{g}(l|j) \equiv \frac{\mu_{g}(j|l)\cdot\lambda(l)}{\mu_{g}(j)} \quad \forall l,j. \end{aligned}$$

By definition, $\hat{\mu}_{g}$ is the distribution of the number of guilty subordinates any supervisor of a guilty agent has, and $\lambda_{g}(\cdot|j)$ is an agent's in-degree distribution conditional on her having j guilty subordinates. While a criminal supervisor of a guilty agent who has degrees (l,j) reports the agent with probability $\big(1-(n\wedge j)/j\big)$, an innocent one reports her with probability $\big(1-\mathbb{X}\{j\geq\underline{j}^{l}\}\cdot(n\wedge j)/j\big)$.

The downward risk function $P\big(n,(\underline{j}^l)_l\big)$ is recursively defined by:

$$p=\frac{\eta}{g}\sum_j\sum_{l\geq 1}\Bigg\{\Big[s+(1-s)\cdot\mathbb{X}\{j\geq\underline{j}^l\}\Big]\Big[1-(1-q)\Big(1-R\big(n,(\underline{j}^l)_l\big)\Big)^{l-1}$$
$$\cdot(1-p)^{\lfloor n\wedge j\rfloor}\big(1-(n\wedge j-\lfloor n\wedge j\rfloor)p\big)\Big]\Bigg\}\hat{\lambda}(l)\mu_g(j|l).$$

It measures the probability a subordinate of an agent is caught and their bribe transaction is consequently detected (with probability η) conditional on the subordinate's being guilty (with probability g) and the agent's having accepted her bribe. Specifically, for a subordinate with degrees (l,j), $\big[s+(1-s)\cdot\mathbb{X}\{j\geq\underline{j}^l\}\big]$ is the probability she is guilty, and $\big[1-(1-q)\big(1-R(n,(\underline{j}^l)_l)\big)^{l-1}(1-p)^{\lfloor n\wedge j\rfloor}\big(1-(n\wedge j-\lfloor n\wedge j\rfloor)p\big)\big]$ is the probability she is caught conditional on her being guilty and the agent's having accepted her bribe.

Just like in the baseline model, the downward risk function $P\big(n,(\underline{j}^l)_l\big)$ is well-defined due to the finiteness of any component in the monitoring network $\big($Assumption 1.1$(i)\big)$.

Corruption Level κ

The corruption level κ – the average number of bribes an agent accepts – is given by:

$$\kappa=\mathbb{E}_{\pi_g}\Big[\Big(s+(1-s)\cdot X\{j\geq\underline{j}^l\}\Big)(n\wedge j)\Big]$$
$$\text{where}\quad\pi_g(l,j)=\lambda(l)\mu_g(j|l)\quad\forall l,j.$$

π_g is the joint distribution of the numbers of supervisors and guilty subordinates an agent has. While a criminal agent with degrees (l,j) accepts $(n\wedge j)$ bribes, an innocent one accepts $(n\wedge j)$ bribes only if her number of guilty subordinates reaches the cutoff policy: $j\geq\underline{j}^l$; otherwise, she accepts 0 bribe.

CHAPTER 2

Cooperative Games with Parochial Fairness Concerns

2.1 Introduction

There is strong evidence that people compare wages with their coworkers,[1] and that it has direct impact on workers' job separation choices. Using survey data, [CMM12] find that employees for the University of California respond higher intention to search for new jobs after being informed about their coworkers' wages. [DGL19] find that rises in the average peer wages increase job separations, and that this effect is triggered solely by fairness concerns.

Wage comparison affects separation choices of not only individual workers, but also groups. Numerous partnership dissolutions are driven by unfair profit distribution. Besides, financial dispute is oftentimes a main factor that propels a state to seek independence from a country. A highly relevant case is the separation of Andersen Consulting from its former parent company Arthur Andersen in 2000. Before that, Andersen Consulting had been sharing its much higher profits with its founding firm under contract [Leo00].

In this paper, we propose a novel, tractable theoretical framework to study group deviation behaviors driven by wage comparison. Our model answers important questions such as what profit sharing rules are the most robust to group deviation, and how group members redistribute incomes after the separation. We show that at the center of our results are the differences in group members' fairness concerns before and after the separation.

Our model is naturally built upon the standard transferable-utility (TU) cooperative game [NM04]. We extend the utility space to allow players' payoffs to be dependent on their

[1] See, for example, [CP22], etc.

coalition members' earnings. We then generalize the definitions of profitable coalitional deviations and the core accordingly.[2]

To model players' fairness concerns, we borrow the specification proposed by [FS99].[3] In their model, a player's utility depends on her own income as well as the differences between her income and other players'. She is averse to disadvantageous inequality and derives negative payoffs from comparing incomes with the richer, but may derive positive (if she is *competitive*, namely, likes advantageous inequality) or negative (if she is *benevolent*, namely, dislikes advantageous inequality) payoffs from comparing incomes with the poorer. We select the Fehr-Schmidt model to account for players' fairness concerns because it is an affine function, thus easily adaptable to the cooperative game structure; and because it is rich enough to distinguish between different types of fairness concerns, generating nontrivial results and allowing for interesting comparative statics.

Two underlying assumptions are essential in driving our main results. First, we assume that players' fairness concerns are *parochial*, that is, they only compare incomes with those in the same coalition. Parochialism finds its strong empirical support in evolutionary theories. In a field experiment on indigenous groups in Papua New Guinea, [BFF06] find that punishers tend to exert more serious punishment on outgroup social norm violators than ingroup ones. More relevantly, another paper, [FBR08], directly links fairness concerns with parochialism. They observe that an overwhelming majority of children at age 7-8 favor egalitarian allocations over unequal ones, whether they hold advantageous or disadvantageous stakes in the latter. Moreover, a child is roughly 20% more likely to choose the egalitarian allocation if their peer is an ingroup member.

As parochialism is broadly observed in human behaviors, it is only reasonable to claim its existence in economics, and specifically, in wage comparison. Indeed, in [DGL19], in contrast to the finding that wage comparison with higher-paid colleagues pushes workers to resign and look for new jobs, market wage comparison exerts influence statistically indistin-

[2]We rely on the original notion of the core developed by [Gil59] and [SS53].

[3]There are several other representative models for fairness concerns. See [BO00], [CFG07], etc.

guishable from zero. In a similar spirit, we define coalitional deviations based on parochial income comparison: in evaluating if a potential deviation is profitable, a player only compares incomes with the other members of the deviating coalition, in contrast to her income comparison prior to the deviation, which she does with everyone else in the grand coalition (GC).

Our second assumption is that fairness concerns are invariant with the coalition size, and depend only on the income distribution. The Fehr-Schmidt specification that we adopt in the model already takes care of this aspect through normalizing a player's disutility from income comparison with the group size, so that proportionally altering the numbers of the others with whom a player compares incomes does not change her utility level. This is not a peculiar feature, as another preference model for fairness concerns proposed by [BO00] also controls for the group size. They do so by summarizing income comparison into one variable that enters a player's utility function – the group average income.

In fact, only after controlling for the coalition size can a player reasonably compare income distributions drawn from groups with different populations, which is commonly encountered in real-world scenarios.[4] In the baseline model, we assume away heterogeneity among players and coalitions,[5] and study how changes in the coalition size after coalitional deviations interact with players' parochial fairness concerns to govern the results of income redistribution for deviating coalitions and the structures of core allocations.

In the first result, we discuss how income redistribution after a coalition deviates is affected by the three different types of fairness concerns: aversion to disadvantageous inequality, inclination to advantageous inequality (when players are competitive), and aversion to advantageous inequality (when players are benevolent). For this purpose, for each GC allocation and coalition, we characterize the unique associated *shadow allocation* – the allocation for the coalition that makes each member indifferent between blocking the GC allocation

[4]We can think about the situation where a worker job-hops from a large company to a small startup, or mergers and acquisitions cases.

[5]Specifically, we assume that all players share the same preference parameters, and all the coalitions of the same size produce the same worth.

and not. We then analyze the coalition's income redistribution outcomes by comparing the shadow allocation with the initial GC allocation.

We find that while aversion to inequality, be it disadvantageous or advantageous, enlarges the income gaps among players after a coalition deviates, inclination to advantageous inequality narrows them. This is because a deviating coalition consists of the poorest players at the initial GC allocation. In terms of disadvantageous inequality, it implies that by deviating, the coalition members are freed from comparing incomes with the richer, and so can tolerate larger income gaps among themselves. As for advantageous inequality, it means any coalition member finds all those poorer than her also recruited by the coalition. Suppose players are benevolent. Since now one is in a smaller coalition and so pays more attention to comparing incomes with the poorer, thus suffering from higher disutility from advantageous inequality aversion, she demands higher income as compensation.[6] This in turn widens the income gaps. The intuition is flipped if players are otherwise competitive.

Our second result characterizes the GC allocations that are the most robust against coalitional deviations under different magnitudes of fairness concerns. Though most cases generate the same outcome as the TU game (where players are completely self-interested) does – the equal allocation; curiously, if players are moderately benevolent, we arrive at the "tyranny-of-the-majority" allocation that features a unique poorest player, a group of advantageous equal income earners, and an infinitely large gap between their income levels.[7] This drastically unequal allocation is robust despite the absence of heterogeneity among players' preferences and the fact that the equal allocation maximizes any deviating coalition's total income.

The intuition for the robustness of the "tyranny-of-the-majority" allocation is twofold

[6]We inherit the assumption made by [FS99] that a player is not so benevolent as to being willing to discard money to feel better. In other words, a rise in one's income brings her more utility from self-interest than disutility from advantageous inequality aversion.

[7]Upper constraints on the degree of advantageous inequality aversion are necessary in achieving this result. If players are so benevolent that they are willing to donate to the poorer, the only Pareto optimal allocation for a coalition becomes the equal allocation, and so the most robust GC allocation is again the equal one.

and concerns both disadvantageous and advantageous inequality aversion. On one hand, disadvantageous inequality aversion diminishes players' incomes after a coalition deviates, thereby making a shadow allocation cheaper, thus more likely to be feasible. This is because now that the members of the deviating coalition have gotten rid of the richer players in the GC, they can tolerate lower income levels. Nevertheless, only at the "tyranny-of-the-majority" allocation is this force completely wiped out, as the poorest player, the only one suffering from disadvantageous inequality aversion, does not fare better from the deviation – she faces the same peer income distribution before and after the move. In other words, the "tyranny-of-the-majority" allocation blocks the channel conducive to coalitional deviations.

On the other hand, advantageous inequality aversion elevates players' income levels after coalitional deviations, for in a smaller coalition, players pay more attention to their poorer peers, thus deriving higher disutility from comparing incomes with them. Hence, they desire higher incomes as compensation, making a shadow allocation more expensive and so less likely to be feasible. The "tyranny-of-the-majority" allocation takes this effect to extremes by indefinitely widening the income gap between the rich and the poor, so as to elevate the cost of the associated shadow allocation to infinity to maximally prevent the coalition from deviating.

The paper is organized as follows. Section 2.2 posits the general cooperative game with other-regarding preferences, and defines the stability concept. Section 2.3 analyzes parochial fairness concerns in this framework: 2.3.1 introduces the Fehr-Schmidt preference model; 2.3.2 derives the core, characterizing shadow allocations along the way; 2.3.3 presents the main results. Section 2.4 concludes.

2.2 Model

We define the game as follows:

Definition 2.1. *A **cooperative game with other-regarding preferences** consists of*

i a finite set of players N;

ii a worth function v that assigns to every coalition $C \subseteq N$ a total worth $v(C) \in \mathbb{R}_+$;

iii for each player i in coalition $C \subseteq N$, a utility function $u_{i,C}$ that for every allocation $\mathbf{x} \in \mathbb{R}^C$ returns player i's utility $u_{i,C}(x_i; \mathbf{x}_{-i})$.

Notice that the standard TU game is a special case of this game: let $u_{i,C}(x_i; \mathbf{x}_{-i}) = x_i$ for any coalition $C \subseteq N$ and player $i \in C$.

For each coalition $C \subseteq N$, we denote by $V(C)$ the set of its *feasible allocations* $\left\{\mathbf{x} \in \mathbb{R}^C \middle| \sum_{i \in C} x_i \leq v(C)\right\}$. We call an allocation $\mathbf{x} \in \mathbb{R}^N$ for the grand coalition N an *income profile*.

Coalitional deviations are formalized below.

Definition 2.2. *A coalition $C \subseteq N$ **blocks** an income profile $\mathbf{x} \in V(N)$ if there exists an allocation $\mathbf{y} \in V(C)$ at which all players in C are strictly better off: $u_{i,C}(y_i; \mathbf{y}_{-i}) > u_{i,N}(x_i; \mathbf{x}_{-i})$ for all $i \in C$.*

Notice that we allow the grand coalition N to block an income profile too. In the TU game, the grand coalition N never blocks an income profile that uses up the total worth $v(N)$, for it is Pareto optimal; in general, though, this is not always the case.

Define the core of the game as follows.

Definition 2.3. *The **core** is the set of income profiles $\mathbf{x} \in V(N)$ that are not blocked by any coalition $C \subseteq N$.*

2.3 Parochial Fairness Concerns

In this section, we let agents' utilities assume a specific functional form to study how parochial fairness concerns influence income redistribution outcomes after coalitional deviations, as well as how they shape the characteristics of core allocations. We discuss the chosen preference

model in 2.3.1, propose an efficient way of solving for the core in 2.3.2, and present the main results in 2.3.3.

2.3.1 Fehr-Schmidt Preference for Fairness

We adopt the preference model for fairness proposed by [FS99], and adapt it to our framework: for each coalition $C \subseteq N$ and allocation $\mathbf{x} \in \mathbb{R}^C$, the utility for player $i \in C$ is given by:

$$\begin{aligned} u_{i,C}(x_i; \mathbf{x}_{-i}) = x_i - \frac{1}{|C|-1} \sum_{\substack{j \in C, \\ j \neq i}} \Big[& \alpha_i \cdot (x_j - x_i, 0)^+ \\ & + \beta_i \cdot (x_i - x_j, 0)^+ \Big] \end{aligned} \qquad \text{if } |C| \geq 2, \tag{2.1}$$

$$u_{i,C}(x_i) = x_i \qquad \text{if } |C| = 1.$$

Player i derives linear payoffs from her own income x_i, as well as disutility from comparing incomes with the other players in coalition C. Each unit of income difference with a richer player gives i disutility $\alpha_i \geq 0$, that with a poorer player giving her disutility β_i. If $\beta_i \geq 0$, i is *benevolent* and dislikes comparing incomes with the poorer; otherwise, if $\beta_i < 0$, she enjoys the presence of poorer members and thus is considered *competitive*. The total disutility from income comparison is normalized by the size of the coalition net of i herself – divided by $\big(|C| - 1\big)$.

To better match with reality, [FS99] impose two assumptions on a player's preference parameters (α_i, β_i). We carry them over to this paper: (1) a player is more averse to disadvantageous income inequality than advantageous one, i.e., $\alpha_i \geq \beta_i$; (2) the highest income earner(s) in a coalition is (are) not so benevolent as to being willing to throw money away, i.e., $\beta_i < 1$.[8]

[8]To see that, notice that from equation (2.1), a richest player i's utility can be rearranged as $(1-\beta_i)x_i + \beta_i/(|C|-1)\sum_{j \in C, j \neq i} x_j$. If $\beta_i \geq 1$, it weakly decreases in i's income x_i. Hence, if i is the only richest player, she prefers to throw some money away as long as she remains the richest after doing that. Likewise, if there are multiple richest players, they are better off throwing away some equal amount of money provided they still earn the highest income after it.

2.3.2 Solving for the Core

This section proposes a convenient tool for finding core allocations, which also paves the way for understanding income redistribution after coalitional deviations (to be discussed in Section 2.3.3).

To simplify analysis, we make the following symmetry assumptions on the worth function v and the utility functions $u_{i,C}$.

Assumption 2.1 (Symmetry). *i Any two coalitions C, T with the same size $\big(|C| = |T|\big)$ produce the same worth: $v(C) = v(T)$.*

ii All agents share the same preference parameters: $(\alpha_i, \beta_i) = (\alpha, \beta)$ $\forall i \in N$.

(ii) implies a utility function $u_{i,C}(x_i; \mathbf{x}_{-i})$ is player-independent. It is also coalition-independent, as an allocation $\mathbf{x} \in \mathbb{R}^C$ contains information on the coalition size $|C|$. We thus drop the subscripts and write $u(x_i; \mathbf{x}_{-i})$.

We now derive the conditions for examining if an income profile $\mathbf{x} \in V(N)$ is in the core. First, notice that Assumption 2.1(ii), coupled with the parameter restriction $\beta < 1$ (which ensures the richest players never want to throw money away), suggests that a richer player derives higher utility: given coalition C and allocation $\mathbf{x} \in \mathbb{R}^C$, if the incomes for some two players $i, j \in C$ satisfy $x_i > x_j$, then their utilities satisfy $u(x_i; \mathbf{x}_{-i}) > u(x_j; \mathbf{x}_{-j})$ (see Lemma 2.1, Appendix 2.5.1). Also, Assumptions 2.1$(i$-$ii)$ imply any two coalitions with the same size induce the same set of achievable utility vectors. Hence, at any income profile $\mathbf{x} \in V(N)$, out of all the coalitions with the same size $t \in \big\{1, ..., |N|\big\}$, the one consisting of the t poorest players (who derive the lowest utilities) is the most likely to block $\mathbf{x}$. This property greatly reduces the computational burden for deriving core allocations.

We simplify the notations prior to presenting coalitions' deviation constraints. For any coalition C and allocation $\mathbf{x} \in \mathbb{R}^C$, w.l.o.g., we index the players' incomes in ascending order: $x_1 \leq ... \leq x_{|C|}$. Denote the grand coalition size by $n \equiv |N| \geq 3$. Let v_t be the total worth produced by a coalition with size $t \in \{1, ..., n\}$, and $C_t \equiv \{1, ..., t\}$ be the coalition formed by the t poorest players (applicable to any income profile $\mathbf{x} \in \mathbb{R}^n$). Thus, to check if an

income profile $\mathbf{x} \in V(N)$ is in the core, it suffices to check if coalition C_t blocks $\mathbf{x}$ for each $t \in \{1, ..., n\}$, adding up to n constraints.

Quite straightforwardly, the singleton C_1 – the poorest player – does not block an income profile $\mathbf{x} \in V(N)$ if and only if she finds it not worthwhile to produce alone: $u(x_1; \mathbf{x}_{-1}) \geq v_1$. Group deviations generate more complexity. To examine if a multiplayer coalition C_t ($t \geq 2$) is willing to block the income profile $\mathbf{x}$, we discuss two cases: when players are benevolent enough – when the parameter β is positive and large, and when they are not.

Suppose players are sufficiently benevolent, in particular, suppose $\beta \geq (t-1)/t$. We consider the set of Pareto optimal allocations for coalition C_t, as they are the ones C_t is most willing to deviate to. First, notice that equation (2.1) suggests any allocation not using up the entire worth v_t is not Pareto optimal, for we can give each player some equal amount to make them strictly better off – it makes everyone richer without altering the income differences. Now pick any allocation $\mathbf{y} \in V(C_t)$ that does not waste resources: $\sum_{i=1}^{t} y_i = v_t$. Then so long as income inequality persists in coalition C_t, i.e., $y_t > y_1$, the richest players can do better by donating to poorer members while remaining the richest. Hence, the only (weakly) Pareto optimal allocation – and so the only allocation coalition C_t is most likely to deviate to – is the equal allocation where each player gets income as well as utility v_t/t (see Lemma 2.2, Appendix 2.5.2), thus C_t does not block an income profile $\mathbf{x} \in V(N)$ if and only if the highest income earner belonging to it – player t – is reluctant to deviate to the equal allocation: $u(x_t; \mathbf{x}_{-t}) \geq v_t/t$.[9]

The more interesting case is when players are either moderately benevolent or competitive: $\beta < (t-1)/t$. Notice that now any allocation that uses up the total worth v_t is Pareto optimal, for no one – not even the richest players – is willing to make a donation (see Lemma 2.2, Appendix 2.5.2). Hence, to check if an income profile $\mathbf{x} \in V(N)$ is blocked by coalition C_t, we employ a new technique. In particular, we link $\mathbf{x}$ to an allocation $\mathbf{y} \in \mathbb{R}^t$ for coalition C_t and examine the latter's feasibility. The specifics are formalized in the following

[9]It implies that if $\beta > (n-1)/n$, the only income profile likely to be in the core is the equal income profile where each player gets v_n/n, as the others are not (strongly) Pareto optimal, and are thus blocked by the grand coalition N.

proposition.

Proposition 2.1. *Given coalition C_t with $t \in \{2, ..., n\}$, if $\beta < (t-1)/t$, then*

*i any income profile $\mathbf{x} \in \mathbb{R}^n$ is associated with a unique **shadow allocation** $\mathbf{y} \in \mathbb{R}^t$ that satisfies $u(y_i; \mathbf{y}_{-i}) = u(x_i; \mathbf{x}_{-i})$ for all players $i = 1, ..., t$;*[10]

ii an income profile $\mathbf{x} \in V(N)$ is not blocked by coalition C_t iff the associated shadow allocation $\mathbf{y}(\mathbf{x}, C_t) \in \mathbb{R}^t$ satisfies $\sum_{i=1}^{t} y_i(\mathbf{x}, C_t) \geq v_t$.

Proof: See Appendix 2.5.3.

In other words, if players are not too benevolent: $\beta < (t-1)/t$, then coalition C_t does not block an income profile $\mathbf{x} \in V(N)$ if and only if it cannot afford the unique shadow allocation $\mathbf{y}(\mathbf{x}, C_t) \in \mathbb{R}^t$ that gives each member the same payoff as at $\mathbf{x}$.[11][12]

2.3.3 Comparative Statics

This section presents the main results. We discuss how parochial fairness concerns shape income redistribution outcomes after a coalition deviates, and characterize and compare across the income profiles most robust to coalitional deviations under different degrees of parochial fairness concerns.

The following proposition shows that there exists a simple, explicit relationship between an income profile and its shadow allocations.

---

[10] Since the ordering of players' preferences in a coalition is always the same as that of their income levels, players' identity indices do not change after coalitional deviations. See the proof for details.

[11] Notice that in a TU game where players ignore income differences – the special case where the preference parameters $(\alpha, \beta) = \vec{\mathbf{0}}$, given any income profile $\mathbf{x} \in \mathbb{R}^n$ and coalition C_t, the associated shadow allocation $\mathbf{y}(\mathbf{x}, C_t) = (x_i)_{i=1}^t$. Hence, to ensure that an income profile $\mathbf{x} \in V(N)$ is not blocked by a coalition C_t, we are back at the familiar constraint $\sum_{i=1}^{t} x_i \geq v_t$.

[12] Clearly, if $\beta < \big(n-1\big)/n$, then at any income profile $\mathbf{x} \in \mathbb{R}^n$, the shadow allocation for the grand coalition N – $\mathbf{y}(\mathbf{x}, N)$ – is just $\mathbf{x}$. Hence, Proposition 2.1(ii) suggests that the grand coalition N does not block an income profile $\mathbf{x} \in V(N)$ if and only if $\sum_{i=1}^{n} x_i = v_n$, i.e., no resource is wasted at $\mathbf{x}$. It is consistent with the fact that when $\beta < \big(n-1\big)/n$, the Pareto optimal income profiles are those that use up the total worth v_n (Lemma 2.2, Appendix 2.5.2).

Proposition 2.2. *Given income profile* $\mathbf{x} \in \mathbb{R}^n$ *and coalition* C_t *with* $t \in \{2, ..., n\}$*, if* $\beta < (t-1)/t$*, then the associated shadow allocation* $\mathbf{y} \in \mathbb{R}^t$ *satisfies:*

$$y_{j+1} - y_j = \theta_j^t \cdot (x_{j+1} - x_j) \quad \forall j = 1, ..., t-1$$
$$\textit{where} \quad \theta_j^t \equiv \frac{1 + \alpha \cdot (n-j)/(n-1) - \beta \cdot j/(n-1)}{1 + \alpha \cdot (t-j)/(t-1) - \beta \cdot j/(t-1)} > 0.$$

Proof: See Appendix 2.5.3.

Proposition 2.2 has implications on how players' parochial fairness concerns govern income redistribution outcomes after a coalition deviates. We measure income redistribution with the parameter $\theta_j^t = (y_{j+1} - y_j)/(x_{j+1} - x_j)$ – changes in the income gap between two adjacent players after they deviate to a smaller coalition, and separately discuss the three different parochial fairness concerns: disadvantageous inequality aversion, advantageous inequality inclination, and advantageous inequality aversion.

First, suppose $\alpha > 0$ and $\beta = 0$, i.e., a player dislikes comparing incomes with the richer, and ignores the poorer. In this scenario, $\theta_j^t > 1$, indicating that income gaps enlarge after players deviate to a smaller coalition. Intuitively, deviating to coalition C_t means getting rid of the other $(n-t)$ richer players in the grand coalition N. Hence, a player derives less disutility from comparing incomes with the richer, and so to maintain the same payoff as before, she can tolerate larger income gaps.

Now assume $\alpha = 0$ and $\beta < 0$: players disregard richer members, and enjoy comparing incomes with the poorer, i.e., they are competitive. This time we arrive at $\theta_j^t < 1$, that is, income gaps shrink after a coalitional deviation. To understand that, notice that when a player deviates, so do all those poorer than her. As they are now in a smaller coalition, the player pays more attention to her poorer members, thus obtaining higher payoff from comparing incomes with them (remember, in the Fehr-Schmidt preference specification, the payoff from income comparison is normalized by the coalition size). Therefore, to reach the same utility as before, she desires smaller income gaps.

Although fairness concerns towards the poorer drive opposite outcomes from those towards the richer when players are competitive; if they are benevolent, the two incentives align. To see that, suppose $\alpha \geq \beta > 0$, so that players dislike comparing incomes with anyone. In this case, income gaps enlarge after a coalition deviates: $\theta_j^t > 1$, and is even larger compared to when $\beta = 0$. This is because now not only can a player tolerate larger income gaps with the richer (as discussed before), but she also demands larger income gaps with the poorer. Intuitively, the deviating coalition enlists all those poorer than her. As reduction in the coalition size makes the presence of these poorer members more conspicuous and so the income comparison with them more unpleasant, the player demands higher income as compensation (remember, $\beta < 1$, so that a unit increase in one's income brings more joy from self-interest than unhappiness from benevolence). This increases her income gaps with the poorer.

We now characterize the income profiles most robust to coalitional deivations, which have direct implications on core nonemptiness. Define them as follows:

Definition 2.4. $\mathbf{x} \in V(N)$ *is a* ***critical income profile*** *for coalition* C_t $(t = 1, ..., n-1)$ *if* C_t *blocks* $\mathbf{x}$ *implies it blocks any income profile in* $V(N)$.[13]

In other words, a coalition is most reluctant to block its critical income profiles.

To understand the structure of a critical income profile, we first work through a few simple cases. The critical income profile for the singleton C_1 is obviously the equal income profile, as it guarantees the poorest player the highest income and utility v_n/n. Likewise, for any multiplayer coalition C_t, if players are so benevolent $\big(\beta \geq (t-1)/t\big)$ that the only Pareto optimal allocation they can deviate to is the equal one that gives each player income v_t/t, the critical income profile is again the equal income profile, for it ensures that player t – the richest and thus the most reluctant to deviate – gets the highest income and utility v_n/n.

[13]This definition only makes sense for strict subcoalitions of the grand coalition $C_t \subset N$, as the grand coalition N never blocks its (strongly) Pareto optimal income profiles, and always blocks the others.

Nevertheless, the following theorem shows that in some cases the equal income profile fails to be a critical one.

Theorem 2.1. *Given any coalition C_t with $t \in \{2, ..., n-1\}$, there exists a cutoff $\beta^t \in \big(0, (t-1)/t\big)$ such that if $\beta \neq \beta^t$, the unique critical income profile $\mathbf{x} \in V(N)$ for C_t satisfies $x_1 \to -\infty$, $x_2 = ... = x_n$, and $\sum_{i=1}^n x_i = v_n$ when $\beta^t < \beta < (t-1)/t$, and is the equal income profile where each player gets v_n/n otherwise; if $\beta = \beta^t$, any income profile $\mathbf{x} \in V(N)$ satisfying $x_1 \leq x_2 = ... = x_n$ and $\sum_{i=1}^n x_i = v_n$ is a critical income profile.*

Proof: See Appendix 2.5.4.

To put into words, for any multiplayer coalition C_t, if players are sufficiently benevolent ($\beta > \beta^t > 0$), but not so much as to being willing to donate to their poorer members after deviation $\big(\beta < (t-1)/t\big)$, then the "tyranny-of-the-majority" income profile supersedes the equal one as the critical income profile, namely, there is a single poorest player who earns indefinitely small amount, and the others split the rest of the worth evenly.

This is surprising for two reasons. First, naivete suggests that benevolence should induce fair outcomes, while in fact it can lead to considerably more unequal results than pure self-interest does (just consider the TU game with $\alpha, \beta = 0$, where the critical income profile is the equal income profile). Second, it seems natural to claim that a coalition is the most reluctant to block the equal income profile. Indeed, for any multiplayer coalition C_t, the equal income profile where everyone gets v_n/n not only maximizes the coalition members' total income, but also cancels their disutility from income comparison completely.

Below we explain this puzzle. In particular, we show how both disadvantageous and advantageous inequality aversion work together to induce the "tyranny-of-the-majority" income profile as the critical income profile.

Notice that when players are not too benevolent, a coalition does not block an income profile if and only if the associated shadow allocation is not feasible $\big($Proposition 2.1$(ii)\big)$. It implies that the income profiles the coalition is the most reluctant to block – the critical income profiles – must be those associated with the most expensive shadow allocations.

Consider first players' disadvantageous inequality aversion (governed by the parameter $\alpha \geq 0$). We argue that under this force, a critical income profile assumes the particular shape where except for the poorest player, everyone else earns equal amount.

First, notice that income comparison with the richer drives down the total value of a shadow allocation. To see that, remember that a deviating coalition $C_t \subset N$ only enlists the t poorest players, while ousting the other richer ones. Hence, players in C_t now derive less disutility from comparing incomes with the richer, and thus can tolerate lower incomes.[14]

However, if everyone except for the poorest player earns the same amount, this effect is completely shut off, as the poorest player – the only player suffering from income comparison with the richer – does not find her condition better after deviating to a smaller coalition. More precisely, consider any income profile $\mathbf{x} \in V(N)$ with $x_1 \leq x_2 = ... = x_n$. If coalition C_t spins off from the grand coalition N, and each member of C_t earns the same as before, then player 1's utility does not change: $u(x_1; x_2, ..., x_t) = u(x_1; x_2, ..., x_n)$, leaving no space for depreciating the shadow allocation $\mathbf{y}(\mathbf{x}, C_t)$.

Now look at players' advantageous inequality aversion (governed by the parameter β). We argue that if players are benevolent enough ($\beta > \beta^t > 0$), this force widens the income gap between the poorest player and the others to infinity, leading us to the "tyranny-of-the-majority" income profile as the critical income profile.

First, we show that when players are benevolent: $\beta > 0$, then at any income profile $\mathbf{x} \in V(N)$ with $x_1 < x_2 = ... = x_n$, i.e., where the poorest player earns strictly less than the others do, the shadow allocation for a deviating coalition assigns to each member higher income than before. Specifically, suppose coalition C_t $(1 < t < n)$ deviates. Then its richer members – players 2 through t – suffer more from comparing incomes with player 1, for her presence in a smaller coalition is more conspicuous. They thus desire higher incomes

[14]To gain more insights into it, we work through a simple example. In a three-player game ($n = 3$), suppose players only compare incomes with the richer ($\alpha > 0$, $\beta = 0$). Consider any income profile $\mathbf{x} \in V(N)$. After the two-player coalition C_2 deviates, player 2 no longer compares incomes with the richest player – player 3, and so to achieve the same utility as before, she demands lower income: $y_2 \leq x_2$. It in turn shortens the income gap between 1 and 2. Since 1 now stops comparing incomes with 3 too, she also demands lower income: $y_1 \leq x_1$.

as compensation: $y_2 = ... = y_t > x_2 = ... = x_t$. Since it in turn enlarges the income gap between player 1 and the others, elevating 1's disutility from comparing incomes with the richer, 1 also demands higher income as compensation: $y_1 \geq x_1$.

In a word, at an unequal income profile $\mathbf{x} \in V(N)$ with $x_1 < x_2 = ... = x_n$, benevolence boosts up the total value of the shadow allocation for a coalition C_t. Nevertheless, since the equal income profile where everyone gets v_n/n maximizes the total income for C_t members (at tv_n/n), only when players are benevolent enough ($\beta > \beta^t > 0$) does the first force dominate to make coalition C_t's shadow allocation at the unequal income profile $\mathbf{x}$ more expensive than that at the equal income profile. In this case, the larger the initial income gap between player 1 and the others ($x_2 - x_1$) is, the more expensive the shadow allocation becomes. Thus, to obtain the critical income profile featuring the most expensive shadow allocation, we increase the income gap to infinity.

2.4 Conclusion

In a cooperative game setting, this paper studies how parochial fairness concerns shape income redistribution outcomes after coalitional deviations and the structures of core allocations. We generalize the TU game through extending the utility space to allow for other-regarding incentives, and apply in the Fehr-Schmidt utility specification to analyze parochial fairness concerns. Given any income profile, we characterize the associated shadow allocation for each coalition that makes the coalition members indifferent between blocking the income profile and not.

One preliminary result shows that shadow allocations have implications on how parochial fairness concerns affect income redistribution results following coalitional deviations: while aversion to both disadvantageous and advantageous inequality exaggerates income inequality after a coalition deviates, advantageous inequality inclination reduces it. Built upon its intuition, our second result states that when players are sufficiently and moderately benevolent, the critical income profile most robust to coalitional deviations is the "tyranny-of-the-

majority" income profile where a single poorest player earns indefinitely small amount, and all the others share the same income.

We plan to carry this research further by exploring how the structure of the worth function and parochial fairness concerns jointly determine the characteristics of core allocations. We will also push beyond the symmetry assumption, asking how heterogeneity in players' fairness concerns and coalitions' productivities may alter the main results.

2.5 Appendix

2.5.1 Ordering of Players' Preferences

Lemma 2.1. *Given any coalition C with $|C| \geq 2$ and allocation $\mathbf{x} \in \mathbb{R}^C$, the income levels for some two players $i, j \in C$ satisfy $x_i > x_j$ if and only if their utilities satisfy $u(x_i; \mathbf{x}_{-i}) > u(x_j; \mathbf{x}_{-j})$.*

Proof: Given any multiplayer coalition C and two players $i, j \in C$ with $x_i > x_j$, we show that the difference in i and j's utilities $\Delta u_{ij} \equiv u(x_i; \mathbf{x}_{-i}) - u(x_j; \mathbf{x}_{-j}) > 0$. Notice that it is equal to their positive income difference $(x_i - x_j > 0)$ plus the difference in their disutilities from income comparison (denoted by $\Delta disu_{ij}$).

If i and j are the only two players in coalition C, then $\Delta disu_{ij} = (-\beta + \alpha)(x_i - x_j) \geq 0$, such that the utility difference $\Delta u_{ij} > 0$. The inequality holds because agents are assumed to be more averse to disadvantageous inequality than advantageous one: $\alpha \geq \beta$.

Now suppose coalition C includes players other than i and j. Pick any player $k \in C/\{i, j\}$, she could earn more than i does, less than j does, or something in between i and j's income levels. The difference in i and j's disutilities from comparing incomes with k (denoted by $\Delta disu_{ij}^k$) is, respectively, $\alpha(x_i - x_j)/\big(|C| - 1\big)$, $-\beta(x_i - x_j)/\big(|C| - 1\big)$, and $\big[-\beta(x_i - x_k) + \alpha(x_k - x_j)\big]/\big(|C| - 1\big)$. If players are competitive: $\beta \leq 0$, then $\Delta disu_{ij}^k \geq 0$ in all three cases. Hence, the total difference in i and j's disutility from income comparison $\Delta disu_{ij} \geq 0$, indicating that their utility difference $\Delta u_{ij} > 0$. If players are benevolent:

$\beta > 0$, then $\Delta disu_{ij}^k$ is minimized at $-\beta(x_i - x_j)/\big(|C| - 1\big)$ when k earns lower than both i and j do. It indicates that the total difference in i and j's disutility from income comparison $\Delta disu_{ij}$ is minimized if i, j are the two richest players in coalition C. In this case, their utility difference Δu_{ij} is still positive, as shown below:

$$\begin{aligned}\Delta u_{ij} &= (x_i - x_j) + \frac{1}{|C| - 1} \cdot (-\beta + \alpha)(x_i - x_j) - \frac{|C| - 2}{|C| - 1} \cdot \beta(x_i - x_j) \\ &= \Big(1 - \beta + \frac{\alpha}{|C| - 1}\Big)(x_i - x_j) \\ &> 0,\end{aligned}$$

where we obtain the inequality from the "no-discarding" assumption: $\beta < 1$. Hence, $\Delta u_{ij} > 0$ always holds.

Q.E.D.

2.5.2 Pareto Optimality

Lemma 2.2. *Given any coalition C_t with $t \in \{2, ..., n\}$, if $\beta \geq (t-1)/t$, the set of weakly Pareto optimal allocations contains only the equal allocation where each player gets v_t/t; otherwise, it is $\big\{\mathbf{y} \in V(C_t) \big| \sum_{i=1}^t y_i = v_t\big\}$.*

Proof: Given any multiplayer coalition $C_t \subseteq N$, we first prove that if players are benevolent enough: $\beta \geq (t-1)/t$, then the only weakly Pareto optimal allocation is the equal allocation where each player gets income v_t/t. It is easy to see that the equal allocation is weakly Pareto optimal, as moving to any other feasible allocation results in some players – the ones who end up to be the poorest – doing strictly worse. We now show any other feasible allocation is not weakly Pareto optimal.

It is clear that a wasteful allocation is not weakly Pareto optimal, for we can give each player some equal amount to make all strictly better off. Pick any non-wasteful allocation $\mathbf{y} \in V(C_t)$ with income inequality: $\sum_{i=1}^t y_i = v_t$ and $y_t > y_1$, we show that there exists another allocation $\mathbf{y}' \in V(C_t)$ at which all players are weakly better off than at $\mathbf{y}$, and some

are strictly better off.

Denote by m the number of richest players at allocation $\mathbf{y}$. Then $y_{t-m} < y_{t-m+1} = ... = y_t$. Pick any positive scalar $\Delta \leq (y_{t-m+1} - y_{t-m}) \cdot (t-m)/t$, and construct an allocation $\mathbf{y}' \in \mathbb{R}^t$ as follows:

$$y_i' = \begin{cases} y_i - \Delta & \text{if } i > t-m, \\ y_i + \Delta \cdot m/(t-m) & \text{if } i \leq t-m. \end{cases}$$

It is easy to verify that allocation $\mathbf{y}'$ is feasible, and that the ordering of players' income levels follows that at allocation $\mathbf{y}$: $y_1' \leq ... \leq y_t'$.

We now show all players are better off at allocation $\mathbf{y}'$ than at allocation $\mathbf{y}$. It is obvious that any player $i \leq t-m$ is strictly better off at $\mathbf{y}'$, for she now earns strictly more than before. Besides, her income difference with anyone poorer than her does not change, and that with anyone richer than her either remains the same, or drops.

The utility for any player $i > t-m$ also weakly increases after moving from allocation $\mathbf{y}$ to allocation $\mathbf{y}'$:

$$\begin{aligned} u(y_i'; \mathbf{y}_{-i}') - u(y_i; \mathbf{y}_{-i}) &= -\Delta + \frac{\beta}{t-1} \cdot (t-m)\Big(1 + \frac{m}{t-m}\Big)\Delta \\ &= -\Delta + \frac{\beta}{t-1} \cdot t\Delta \\ &\geq 0, \end{aligned}$$

where the inequality results from the fact that $\beta \geq (t-1)/t$. Hence, any non-wasteful, unequal allocation $\mathbf{y} \in V(C_t)$ is not weakly Pareto optimal.

Next, suppose players are not too benevolent: $\beta < (t-1)/t$. We prove that the weakly Pareto optimal allocations are the non-wasteful ones. Specifically, take any non-wasteful allocation $\mathbf{y} \in V(C_t)$, and perform a random perturbation on players' incomes: pick any non-zero vector $\mathbf{z} \in \mathbb{R}^t$ for which $\sum_{i=1}^t z_i = 0$, and consider the new non-wasteful allocation $\mathbf{y}' = \mathbf{y} + \mathbf{z}$.[15] We show that at least one player does strictly worse at $\mathbf{y}'$ than at $\mathbf{y}$.

[15]Notice that at allocation $\mathbf{y}'$ we may encounter abuse of notation: the indexing may not be consistent

If players are benevolent: $\beta > 0$, let j be the highest index for which $z_j < 0$, i.e., the highest income earner at allocation $\mathbf{y}$ who earns strictly less after the perturbation. We show that j is strictly worse off at $\mathbf{y}'$ than at $\mathbf{y}$. Let Δ be the total decrease in j's income $|z_j|$, then the total increase in other players' incomes $\sum_{i \neq j} z_i = \Delta$. Notice that since j is benevolent, the best perturbation outcome for her is when all this total increase in other players' incomes Δ is applied to those who are initially poorer, and that they are still poorer than j after the perturbation, for this arrangement minimizes the increase in j's total income differences with the richer at $(t-j)\Delta$ (coming solely from the drop in j's income level), and maximizes the decrease in her total income differences with the poorer at $\Delta + (j-1)\Delta$ (where the first and the second parts come from the total increase in the poorer's incomes and the drop in j's earnings, respectively). However, even in this most optimistic scenario, j is still strictly worse off after the perturbation, as the total changes in her utility

$$\begin{aligned} &-\Delta - \frac{\alpha}{t-1} \cdot (t-j)\Delta + \frac{\beta}{t-1} \cdot [\Delta + (j-1)\Delta] \\ &\quad \leq -\Delta + \frac{\beta}{t-1} \cdot t\Delta \\ &\quad < 0, \end{aligned}$$

where the first inequality is obtained by letting $j = t$, and the second by invoking $\beta < (t-1)/t$.

If players are competitive: $\beta \leq 0$, pick a player whose income decreases the most: $k \in \arg\min_i z_i$. We show that k is strictly worse off after the perturbation. It is easy to see that k now derives higher disutility from comparing incomes with anyone else, as her income difference with anyone initially weakly richer than her increases, and that with anyone initially strictly poorer than her decreases (if their positions are not interchanged, which is even worse for k). Since k also derives strictly less payoff from self-interest, she is strictly worse off.

Lastly, at any wasteful allocation, we can distribute the wasted worth equally to the

with the ordering of players' income levels after the perturbation. But this is innocuous.

players to make all strictly better off. As the resulting non-wasteful allocation, if not allocation $\mathbf{y}$, features at least one player being strictly worse off than at $\mathbf{y}$, so does the wasteful allocation we start with. Hence, $\mathbf{y}$ is weakly Pareto optimal.

Q.E.D.

2.5.3 Proof of Proposition 2.1 & 2.2

Suppose $\beta < (t-1)/t$. Given any income profile $\mathbf{x} \in \mathbb{R}^n$ and multiplayer coalition C_t, we solve for the associated shadow allocation $\mathbf{y} \in \mathbb{R}^t$, which furnishes the proof for its existence and for Proposition 2.2.

By definition, at a shadow allocation $\mathbf{y} \in \mathbb{R}^t$ for income profile $\mathbf{x} \in \mathbb{R}^n$, members of coalition C_t derive the same utilities as at $\mathbf{x}$. Thus Lemma 2.1 suggests the ordering of their income levels also remains the same as at $\mathbf{x}$, and so do their identity indices. Take any two adjacent players $j, (j+1) \in C_t$. Their equivalence conditions are:

$$u(x_j; \mathbf{x}_{-j}) = u(y_j; \mathbf{y}_{-j}), \tag{2.2}$$

$$u\big(x_{j+1}; \mathbf{x}_{-(j+1)}\big) = u\big(y_{j+1}; \mathbf{y}_{-(j+1)}\big). \tag{2.3}$$

Subtracting (2.3) with (2.2), we obtain:

$$\begin{aligned} &(x_{j+1} - x_j) + \frac{\alpha}{n-1} \cdot (n-j)(x_{j+1} - x_j) - \frac{\beta}{n-1} \cdot j(x_{j+1} - x_j) = \\ &(y_{j+1} - y_j) + \frac{\alpha}{t-1} \cdot (t-j)(y_{j+1} - y_j) - \frac{\beta}{t-1} \cdot j(y_{j+1} - y_j) \end{aligned}$$

$$\begin{aligned} \Rightarrow \quad y_{j+1} - y_j &= \frac{1 + \alpha \cdot (n-j)/(n-1) - \beta \cdot j/(n-1)}{1 + \alpha \cdot (t-j)/(t-1) - \beta \cdot j/(t-1)} \cdot (x_{j+1} - x_j) \\ &= \theta_j^t \cdot (x_{j+1} - x_j). \end{aligned} \tag{2.4}$$

Notice that the condition $\beta < (t-1)/t$ guarantees $\theta_j^t > 0$, so that indeed we have $y_{j+1} \geq y_j$.

Next, we compute the explicit expressions for the shadow allocation $\mathbf{y}$. Notice that the

income for any player $k = 2, ..., t$ can be expressed as:

$$\begin{aligned} y_k &= y_1 + \sum_{i=1}^{k-1}(y_{i+1} - y_i) \\ &= y_1 + \sum_{i=1}^{k-1}\theta_i^t(x_{i+1} - x_i), \end{aligned} \tag{2.5}$$

where the second inequality is obtained from equation (2.4). Now we derive player 1's income y_1 from her equivalence condition:

$$\begin{aligned} u(y_1; \mathbf{y}_{-1}) = u(x_1; \mathbf{x}_{-1}) \quad &\Rightarrow \\ y_1 - \frac{\alpha}{t-1}\sum_{k=2}^{t}(y_k - y_1) = x_1 - \frac{\alpha}{n-1}\sum_{k=2}^{n}(x_k - x_1) \quad &\Rightarrow \\ y_1 - \frac{\alpha}{t-1}\sum_{k=2}^{t}\sum_{i=1}^{k-1}\theta_i^t(x_{i+1} - x_i) = x_1 - \frac{\alpha}{n-1}\sum_{k=2}^{n}\sum_{i=1}^{k-1}(x_{i+1} - x_i) \quad &\Rightarrow \\ y_1 = x_1 - \frac{\alpha}{n-1}\sum_{k=2}^{n}\sum_{i=1}^{k-1}(x_{i+1} - x_i) + \frac{\alpha}{t-1}\sum_{k=2}^{t}\sum_{i=1}^{k-1}\theta_i^t(x_{i+1} - x_i), \end{aligned} \tag{2.6}$$

where the second step is derived from equation (2.5). Plugging (2.6) back into (2.5), we obtain the income levels for the other players $y_2, ..., y_t$.

The fact that $\mathbf{y}$ is the unique shadow allocation for income profile $\mathbf{x}$ is directly implied by Lemma 2.2. Specifically, let coalition C_t's total worth $v_t = \sum_{i=1}^{t} y_i$, then Lemma 2.2 suggests $\mathbf{y}$ is a weakly Pareto optimal allocation, such that any other allocation for coalition C_t with total income weakly below $\sum_{i=1}^{t} y_i$ features at least one player being strictly worse off than at allocation $\mathbf{y}$, and thus cannot be a shadow allocation. Now, consider any allocation with total income strictly higher than that at allocation $\mathbf{y}$: $\mathbf{z} \in \mathbb{R}^t$ with $\sum_{i=1}^{t} z_i > \sum_{i=1}^{t} y_i$. Let coalition C_t's total worth $v_t = \sum_{i=1}^{t} z_i$, then Lemma 2.2 suggests $\mathbf{z}$ is weakly Pareto optimal. Hence, at least one player is strictly better off at allocation $\mathbf{z}$ than at allocation $\mathbf{y}$, indicating that $\mathbf{z}$ is not a shadow allocation.

Lastly, we prove Proposition 2.1(ii). That $\sum_{i=1}^{t} y_i < v_t$ implies income profile $\mathbf{x}$ is blocked by coalition C_t is obvious: take any positive scalar $\Delta < \big(v_t - \sum_{i=1}^{t} y_i\big)/t$, then C_t

can profitably deviate from income profile $\mathbf{x}$ to the feasible allocation where each player i gets $y_i + \Delta$. We now prove the other direction: if $\sum_{i=1}^{t} y_i \geq v_t$, we show that it is not profitable for coalition C_t to deviate from income profile $\mathbf{x}$ to any allocation in $V(C_t)$. Let $\Delta = \big(\sum_{i=1}^{t} y_i - v_t\big)/t \geq 0$, and define allocation $\mathbf{y}' \in V(C_t)$ by letting each player i's income $y_i' = y_i - \Delta$. Apparently, all players are weakly worse off at $\mathbf{y}'$ than at $\mathbf{y}$. Since they are indifferent between allocation $\mathbf{y}$ and income profile $\mathbf{x}$, deviating from $\mathbf{x}$ to $\mathbf{y}'$ is not profitable. Since $\mathbf{y}'$ is a non-wasteful allocation, Lemma 2.2 implies that it is weakly Pareto optimal. Hence, any other allocation in $V(C_t)$ features at least one player doing strictly worse than at allocation $\mathbf{y}'$, and so strictly worse than at income profile $\mathbf{x}$. As a result, coalition C_t does not find it profitable to deviate from income profile $\mathbf{x}$ to any allocation in $V(C_t)$.

Q.E.D.

2.5.4 Proof of Theorem 2.1

Given coalition C_t with $t = 2, ..., n-1$, we first show that if $\beta \geq (t-1)/t$, the unique critical income profile for C_t is the equal income profile where each player gets v_n/n. Since the only weakly Pareto optimal allocation for coalition C_t is the equal one where each member gets v_t/t (Lemma 2.2), C_t blocks an income profile $\mathbf{x} \in V(N)$ if and only if at $\mathbf{x}$, its richest player – player t – is willing to deviate to the equal allocation: $u(x_t, \mathbf{x}_{-t}) < v_t/t$. Now, notice that the equal income profile where each player gets v_n/n is the unique one that maximizes player t's utility $u(x_t, \mathbf{x}_{-t})$: it removes inequality (remember, players are benevolent, so any inequality makes them worse off), and is the only income profile that maximizes player t's income. Therefore, on one hand, if coalition C_t blocks the equal income profile, it also blocks any other income profile: $u(x_t, \mathbf{x}_{-t}) \leq v_n/n < v_t/t$ for all $\mathbf{x} \in V(N)$, indicating that the equal income profile is critical; on the other hand, given any unequal income profile $\mathbf{x} \in V(N)$, if coalition C_t's total worth v_t satisfies $u(x_t, \mathbf{x}_{-t}) < v_t/t \leq v_n/n$, then C_t blocks $\mathbf{x}$, but not the equal income profile, suggesting that $\mathbf{x}$ is not critical.

Next, we solve for the critical income profiles for the case where $\beta \geq (t-1)/t$. We first establish two lemmas that greatly shrink the set of income profiles we need to consider. See

below.

Lemma 2.3. *If $\beta < (t-1)/t$, a critical income profile $\mathbf{x} \in V(N)$ for coalition C_t $(t = 2, ..., n-1)$ satisfies $x_t = x_i \ \forall i = t+1, ..., n$ and $\sum_{i=1}^{t} x_i = v_t$.*

Proof: First, we show that any non-wasteful income profile $\mathbf{x} \in V(N)$ for which $x_j < x_{j+1}$ for some $j \geq t$ is not critical, as there exists another income profile $\mathbf{x}' \in V(N)$ that coalition C_t is strictly more reluctant to block. Specifically, take any positive scalar $\Delta < (x_{j+1} - x_j)/(j+1)$, and let

$$x'_i = \begin{cases} x_i + \Delta & \text{if } i \leq j, \\ x_i - j\Delta & \text{if } i = j+1, \\ x_i & \text{if } i > j+1. \end{cases}$$

Since $j \geq t$, the new income profile $\mathbf{x}'$ makes all members of coalition C_t strictly happier – each has strictly higher income, smaller total income differences with the richer, and the same total income differences with the poorer. Hence, its shadow allocation must be strictly more expensive: $\sum_{i=1}^{t} y(\mathbf{x}', C_t) > \sum_{i=1}^{t} y(\mathbf{x}, C_t)$.[16] If coalition C_t's total worth v_t satisfies $\sum_{i=1}^{t} y(\mathbf{x}, C_t) < v_t \leq \sum_{i=1}^{t} y(\mathbf{x}', C_t)$, then C_t does not block $\mathbf{x}'$, but blocks $\mathbf{x}$ $\big($Proposition 2.1$(ii)\big)$. Thus $\mathbf{x}$ cannot be a critical income profile.

It is easy to see that any wasteful income profile is not critical, as we can split the wasted worth equally among the players, and the resulting non-wasteful income profile makes everyone strictly better off. Hence, coalition C_t is strictly more reluctant to block the latter. This prevents the former to be a critical income profile.

Q.E.D.

Lemma 2.4. *At a critical income profile $\mathbf{x} \in V(N)$ for coalition C_t $(t = 2, ..., n-1)$,*

i if $\beta < -\alpha$, then $x_1 = x_i \ \forall i = 2, ..., t-1$;

[16] Otherwise, set coalition C_t's total worth $v_t = \sum_{i=1}^{t} y(\mathbf{x}, C_t)$, then Lemma 2.2 suggests $\mathbf{y}(\mathbf{x}, C_t)$ is weakly Pareto optimal, contradicting the fact that $\mathbf{y}(\mathbf{x}', C_t) \in V(C_t)$ and makes all players strictly better off.

ii if $-\alpha < \beta < (t-1)/t$, *then* $x_i = x_t \ \forall i = 2, ..., t-1$.

Proof for (i): We show that if $\beta < -\alpha$, then given any income profile $\mathbf{x} \in V(N)$ not satisfying $x_1 = x_i \ \forall i = 2, ..., t-1$, there exists another income profile $\mathbf{x}' \in V(N)$ that coalition C_t is strictly more reluctant to block.

Denote by $j < t-1$ the poorest player who earns strictly less than the adjacent player above: $x_1 = x_j < x_{j+1}$. Take any positive scalar $\Delta \leq (x_{j+1} - x_j) \cdot (t-1-j)/(t-1)$, and construct a new income profile $\mathbf{x}' \in V(N)$ as follows:

$$x'_i = \begin{cases} x_i + \Delta & i \leq j \\ x_i - j/(t-1-j) \cdot \Delta & i = j+1, ..., t-1 \\ x_i & i \geq t \end{cases} \tag{2.7}$$

It is easy to verify that $\sum_{i=1}^n x_i = \sum_{i=1}^n x'_i$. Our goal is to show that the shadow allocation for $\mathbf{x}'$ – $\mathbf{y}(\mathbf{x}', C_t)$ – is strictly more expensive than that for $\mathbf{x}$ – $\mathbf{y}(\mathbf{x}, C_t)$.

Let $\mathbf{y}' \equiv \mathbf{y}(\mathbf{x}', C_t)$, $\mathbf{y} \equiv \mathbf{y}(\mathbf{x}, C_t)$. In the next few steps, we express a player's income at the new shadow allocation y'_i as a function of her income at the old one y_i. By Proposition 2.2:

$$\begin{aligned} y'_t - y'_{t-1} &= \theta^t_{t-1}(x'_t - x'_{t-1}) \\ &= \theta^t_{t-1}\Big(x_t - x_{t-1} + \frac{j}{t-1-j} \cdot \Delta\Big) = y_t - y_{t-1} + \frac{j}{t-1-j} \cdot \theta^t_{t-1}\Delta, \end{aligned} \tag{2.8}$$

$$\begin{aligned} y'_{j+1} - y'_j &= \theta^t_j(x'_{j+1} - x'_j) \\ &= \theta^t_j\Big(x_{j+1} - x_j - \frac{t-1}{t-1-j} \cdot \Delta\Big) = y_{j+1} - y_j - \frac{t-1}{t-1-j} \cdot \theta^t_j\Delta, \end{aligned} \tag{2.9}$$

$$y'_{i+1} - y'_i = \theta^t_i(x'_{i+1} - x'_i) = \theta^t_i(x_{i+1} - x_i) = y_{i+1} - y_i \qquad \forall i \neq t-1, j. \tag{2.10}$$

Notice that player t is indifferent between income profiles $\mathbf{x}$ and $\mathbf{x}'$: she earns the same, and faces the same total income differences with the richer and the poorer. Therefore, she is also

indifferent between the two associated shadow allocations:

$$
\begin{aligned}
& u(y_t; \mathbf{y}_{-t}) = u(y'_t; \mathbf{y}'_{-t}) \\
\Rightarrow \quad & y_t - \frac{\beta}{t-1}\sum_{i=1}^{t-1}(y_t - y_i) = y'_t - \frac{\beta}{t-1}\sum_{i=1}^{t-1}(y'_t - y'_i) \\
& = y'_t - \frac{\beta}{t-1}\sum_{i=j+1}^{t-1}\Big(y_t - y_i + \frac{j}{t-1-j}\cdot\theta^t_{t-1}\Delta\Big) \\
& -\frac{\beta}{t-1}\sum_{i=1}^{j}\Big(y_t - y_i + \frac{j}{t-1-j}\cdot\theta^t_{t-1}\Delta - \frac{t-1}{t-1-j}\cdot\theta^t_j\Delta\Big) \\
& \Rightarrow \quad y'_t = y_t + \frac{j}{t-1-j}\cdot\beta(\theta^t_{t-1} - \theta^t_j)\Delta, \qquad (2.11)
\end{aligned}
$$

where the first step follows from equations (2.8-2.10), and the second is a direct algebraic rearrangement. Hence, for any player $i = j+1, ..., t-1$, her income at the new shadow allocation

$$
\begin{aligned}
y'_i &= y'_t - \sum_{k=i}^{t-1}(y'_{k+1} - y'_k) \\
&= y'_t - \sum_{k=i}^{t-1}(y_{k+1} - y_k) - \frac{j}{t-1-j}\cdot\theta^t_{t-1}\Delta \\
&= y_i + \frac{j}{t-1-j}\cdot\beta(\theta^t_{t-1} - \theta^t_j)\Delta - \frac{j}{t-1-j}\cdot\theta^t_{t-1}\Delta, \qquad (2.12)
\end{aligned}
$$

where the second equality follows from equations (2.8) and (2.10), the third obtained by plugging in equation (2.11). Likewise, we compute the new income for any player $i = 1, ..., j$:

$$
\begin{aligned}
y'_i &= y'_t - \sum_{k=i}^{t-1}(y'_{k+1} - y'_k) \\
&= y'_t - \sum_{k=i}^{t-1}(y_{k+1} - y_k) - \frac{j}{t-1-j}\cdot\theta^t_{t-1}\Delta + \frac{t-1}{t-1-j}\cdot\theta^t_j\Delta \\
&= y_i + \frac{j}{t-1-j}\cdot\beta(\theta^t_{t-1} - \theta^t_j)\Delta - \frac{j}{t-1-j}\cdot\theta^t_{t-1}\Delta + \frac{t-1}{t-1-j}\cdot\theta^t_j\Delta. \qquad (2.13)
\end{aligned}
$$

Now equation (2.9) is also used in deriving the second equality. Lastly, adding up equations

(2.11-2.13), we arrive at the total income for the new shadow allocation:

$$\sum_{i=1}^{t} y'_i = \sum_{i=1}^{t} y_i + \frac{j}{t-1-j} \cdot \big[\beta t - (t-1)\big](\theta^t_{t-1} - \theta^t_j)\Delta. \tag{2.14}$$

Notice that $\beta < -\alpha$ implies $\theta^t_{t-1} < \theta^t_j$ (derived from a straightforward derivative computation), thus $\sum_{i=1}^{t} y'_i > \sum_{i=1}^{t} y_i$, indicating that $\mathbf{x}$ is not a critical income profile – any total worth v_t satisfying $\sum_{i=1}^{t} y_i < v_t \leq \sum_{i=1}^{t} y'_i$ renders coalition C_t block $\mathbf{x}$ but not $\mathbf{x}'$.

Proof for (ii): The proof logic is similar to that for (i): we show that if $\beta > -\alpha$, then given any income profile $\mathbf{x} \in V(N)$ not satisfying $x_i = x_t \; \forall i = 2, ..., t-1$, there exists another income profile $\mathbf{x}' \in V(N)$ that coalition C_t is strictly more reluctant to block.

Denote by $j < t$ the richest player who earns strictly less than the adjacent player above: $x_j < x_{j+1} = x_t$. Take any positive scalar $\Delta \leq (x_{j+1} - x_j)$, and construct a new income profile $\mathbf{x}' \in V(N)$ as follows:

$$x'_i = \begin{cases} x_i - (j-1)\Delta & i = 1 \\ x_i + \Delta & i = 2, ..., j \\ x_i & i \geq j+1 \end{cases}$$

It is easy to verify that $\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} x'_i$. Our goal is to show that the shadow allocation for $\mathbf{x}'$ – $\mathbf{y}(\mathbf{x}', C_t)$ – is strictly more expensive than that for $\mathbf{x}$ – $\mathbf{y}(\mathbf{x}, C_t)$.

Let $\mathbf{y}' \equiv \mathbf{y}(\mathbf{x}', C_t)$, $\mathbf{y} \equiv \mathbf{y}(\mathbf{x}, C_t)$. In the next few steps, we express a player's income at the new shadow allocation y'_i as a function of her income at the old one y_i. By Proposition 2.2:

$$y'_{j+1} - y'_j = \theta^t_j(x'_{j+1} - x'_j) = \theta^t_j(x_{j+1} - x_j - \Delta) = y_{j+1} - y_j - \theta^t_j\Delta, \tag{2.15}$$

$$y'_2 - y'_1 = \theta^t_1(x'_2 - x'_1) = \theta^t_1(x_2 - x_1 + j\Delta) = y_2 - y_1 + j\theta^t_1\Delta, \tag{2.16}$$

$$y'_{i+1} - y'_i = \theta^t_i(x'_{i+1} - x'_i) = \theta^t_i(x_{i+1} - x_i) = y_{i+1} - y_i \qquad \forall i \neq 1, j. \tag{2.17}$$

Since $x_{j+1} = ... = x_t$, equation (2.17) suggests $y_{j+1} = ... = y_t$ and $y'_{j+1} = ... = y'_t$. We now compute player $(j+1)$'s income at the new shadow allocation y'_{j+1}. Notice that $(j+1)$ is indifferent between income profiles $\mathbf{x}$ and $\mathbf{x}'$: she earns the same, and faces the same total income differences with the richer and the poorer. Therefore, she is also indifferent between the two associated shadow allocations:

$$
\begin{aligned}
& u\big(y_{j+1}; \mathbf{y}_{-(j+1)}\big) = u\big(y'_{j+1}; \mathbf{y}'_{-(j+1)}\big) \\
\Rightarrow \quad & y_{j+1} - \frac{\beta}{t-1}\sum_{i=1}^{j}(y_{j+1} - y_i) = y'_{j+1} - \frac{\beta}{t-1}\sum_{i=1}^{j}(y'_{j+1} - y'_i) \\
\Rightarrow \quad & y_{j+1} - \frac{\beta}{t-1}\sum_{i=1}^{j}(y_{j+1} - y_i) = y'_{j+1} - \frac{\beta}{t-1}\sum_{i=2}^{j}(y_{j+1} - y_i - \theta_j^t\Delta) \\
& \qquad\qquad - \frac{\beta}{t-1}(y_{j+1} - y_1 - \theta_j^t\Delta + j\theta_1^t\Delta) \\
& \Rightarrow \quad y'_{j+1} = y_{j+1} - \frac{j}{t-1}\cdot\beta(\theta_j^t - \theta_1^t)\Delta \\
& \Rightarrow \quad y'_i = y_i - \frac{j}{t-1}\cdot\beta(\theta_j^t - \theta_1^t)\Delta \quad \forall i = j+1, ..., t
\end{aligned} \tag{2.18}
$$

where the first and the last steps hold because $y_{j+1} = ... = y_t$ and $y'_{j+1} = ... = y'_t$, the second follows from equations (2.15-2.17), and the third is a direct algebraic rearrangement. Hence, for any player $i = 2, ..., j$, her income at the new shadow allocation

$$
\begin{aligned}
y'_i &= y'_{j+1} - \sum_{k=i}^{j}(y'_{k+1} - y'_k) \\
&= y'_{j+1} - \sum_{k=i}^{j}(y_{k+1} - y_k) + \theta_j^t\Delta \\
&= y_i - \frac{j}{t-1}\cdot\beta(\theta_j^t - \theta_1^t)\Delta + \theta_j^t\Delta,
\end{aligned} \tag{2.19}
$$

where the second equality follows from equations (2.15) and (2.17), the third obtained by substituting y'_{j+1} out with equation (2.18). Likewise, we compute the new income for player

1:

$$\begin{aligned} y'_1 &= y'_{j+1} - \sum_{k=1}^{j}(y'_{k+1} - y'_k) \\ &= y'_{j+1} - \sum_{k=1}^{j}(y_{k+1} - y_k) + \theta_j^t\Delta - j\theta_1^t\Delta \\ &= y_1 - \frac{j}{t-1}\cdot\beta(\theta_j^t - \theta_1^t)\Delta + \theta_j^t\Delta - j\theta_1^t\Delta. \end{aligned} \tag{2.20}$$

Now equation (2.16) is also used in deriving the second equality. Lastly, adding up equations (2.18-2.20), we arrive at the total income for the new shadow allocation:

$$\sum_{i=1}^{t} y'_i = \sum_{i=1}^{t} y_i + j\Big(1 - \frac{t\beta}{t-1}\Big)(\theta_j^t - \theta_1^t)\Delta.$$

Notice that $\beta > -\alpha$ implies $\theta_j^t > \theta_1^t$ (derived from a straightforward derivative computation). Since also $\beta < (t-1)/t$, we have $\sum_{i=1}^{t} y'_i > \sum_{i=1}^{t} y_i$, indicating that $\mathbf{x}$ is not a critical income profile – any total worth v_t satisfying $\sum_{i=1}^{t} y_i < v_t \leq \sum_{i=1}^{t} y'_i$ renders coalition C_t block $\mathbf{x}$ but not $\mathbf{x}'$.

Q.E.D.

Lemma 2.3 and Lemma 2.4(i) suggest that when $\beta < -\alpha$, a critical income profile $\mathbf{x} \in V(N)$ for coalition C_t $(1 < t < n)$ satisfies $x_1 = ... = x_{t-1} \leq x_t = ... = x_n$ and $\sum_{i=1}^{n} x_i = v_n$. We now show that out of all the income profiles satisfying the conditions stated above, the equal one where each player gets v_n/n is the only critical income profile. This part is the easiest to be proved by graph. Suppose $\alpha > 0$, then figure 2.1 completely captures the proof. Below we offer a verbal explanation on the graphical outcome. Notice that any candidate income profile $\mathbf{x} \in V(N)$ can be pinpointed in a two-dimensional graph: plot x_t on the horizontal axis, and x_{t-1} on the vertical axis. Since Proposition 2.2 suggests

that its shadow allocation $\mathbf{y} \equiv \mathbf{y}(\mathbf{x}, C_t)$ satisfies

$$\begin{aligned} y_t - y_{t-1} &= \theta_{t-1}^t (x_t - x_{t-1}) \geq 0, \\ y_{i+1} - y_i &= \theta_i^t (x_{i+1} - x_i) = 0 \quad \forall i < t-1, \end{aligned}$$

such that $y_1 = ... = y_{t-1} \leq y_t$, we can locate $\mathbf{y}$ in the same graph: plot y_t on the horizontal axis, and y_{t-1} on the vertical axis. The black line represents all the non-wasteful, thus candidate, income profiles. We randomly pick one unequal income profile $\hat{\mathbf{x}}$ on it $\big($represented by point $(\hat{x}_t, \hat{x}_{t-1})\big)$, and show that its shadow allocation $\hat{\mathbf{y}} \equiv \mathbf{y}(\hat{\mathbf{x}}, C_t)$ $\big($represented by the point $(\hat{y}_t, \hat{y}_{t-1})\big)$ is strictly cheaper than the shadow allocation for the equal income profile $\big($represented by point $(v_n/n, v_n/n)\big)$. Since each player earns v_n/n at the latter, it translates into showing that the point $(\hat{y}_t, \hat{y}_{t-1})$ is strictly below the green line. Now, before proceeding with the proof, we make the following definition: for any coalition C_t, player $i \leq t$, and income profile $\mathbf{x} \in \mathbb{R}^n$, we define the associated *indifference set* $IC_i(\mathbf{x}, C_t)$ as the set of allocations $\mathbf{y} \in \mathbb{R}^t$ that makes i indifferent between $\mathbf{x}$ and $\mathbf{y}$: $u(y_i; \mathbf{y}_{-i}) = u(x_i; \mathbf{x}_{-i})$. Applied to the proof, we can plot player $(t-1)$'s indifference sets in the grand coalition N and coalition C_t for income profile $\hat{\mathbf{x}}$: $IC_{t-1}(\hat{\mathbf{x}}, N)$ and $IC_{t-1}(\hat{\mathbf{x}}, C_t)$.[17] Denote $(t-1)$'s utility at $\hat{\mathbf{x}}$ – $u\big(\hat{x}_{t-1}; \hat{\mathbf{x}}_{-(t-1)}\big)$ – as u_{t-1}, then $IC_{t-1}(\hat{\mathbf{x}}, N)$ is derived as follows:

$$\begin{aligned} & u\big(x_{t-1}; \mathbf{x}_{-(t-1)}\big) = u_{t-1} \\ \Rightarrow \quad x_{t-1} - \frac{\alpha}{n-1} \cdot (n-t+1)(x_t - x_{t-1}) &= u_{t-1} \\ \Rightarrow \quad x_{t-1} &= \frac{\alpha_n}{1+\alpha_n} \cdot x_t + \frac{u_{t-1}}{1+\alpha_n} \\ \text{where } \alpha_n &\equiv \frac{(n-t+1)\alpha}{n-1} \end{aligned}$$

In the first step, we utilize the fact that $x_1 = ... = x_{t-1} \leq x_t = ... = x_n$, the second step only involves algebraic rearrangement. Similarly, using $u\big(y_{t-1}; \mathbf{y}_{-(t-1)}\big) = u_{t-1}$, we arrive at

[17] More precisely, due to the restriction to two dimensions, we can only plot the intersection of $IC_{t-1}(\hat{\mathbf{x}}, N)$ and $\{\mathbf{x} \in \mathbb{R}^n | x_1 = ... = x_{t-1} \leq x_t = ... = x_n\}$, as well as the intersection of $IC_{t-1}(\hat{\mathbf{x}}, C_t)$ and $\{\mathbf{y} \in \mathbb{R}^t | y_1 = ... = y_{t-1} \leq y_t\}$.

$IC_{t-1}(\hat{\mathbf{x}}, C_t)$:

$$y_{t-1} = \frac{\alpha_t}{1+\alpha_t} \cdot y_t + \frac{u_{t-1}}{1+\alpha_t} \quad \text{where } \alpha_t \equiv \frac{\alpha}{t-1}$$

Since $0 < \alpha_t < \alpha_n$, both $IC_{t-1}(\hat{\mathbf{x}}, C_t)$ and $IC_{t-1}(\hat{\mathbf{x}}, N)$ have positive slopes, and the former is flatter than the latter. Since they intersect the 45 degree line at the same point – at an equal income profile in $\mathbb{R}^n$, if we drop players $(t+1), ..., n$, the utility for player t remains to be her income level, graphically, $IC_{t-1}(\hat{\mathbf{x}}, C_t)$ rotates downward from $IC_{t-1}(\hat{\mathbf{x}}, N)$ (see the two red lines). Notice also that $IC_{t-1}(\hat{\mathbf{x}}, N)$ passes the point $(\hat{x}_t, \hat{x}_{t-1})$, as $\hat{\mathbf{x}} \in IC_{t-1}(\hat{\mathbf{x}}, N)$. Now, following the same principle, we derive player t's indifference set for income profile $\hat{\mathbf{x}}$ in the grand coalition $IC_t(\hat{\mathbf{x}}, N)$:

$$\begin{aligned} & u(x_t; \mathbf{x}_{-t}) = u_t \equiv u(\hat{x}_t; \hat{\mathbf{x}}_{-t}) \\ \Rightarrow \quad & x_t - \frac{\beta}{n-1} \cdot (t-1)(x_t - x_{t-1}) = u_t \\ \Rightarrow \quad & x_{t-1} = -\frac{1-\beta_n}{\beta_n} \cdot x_t + \frac{u_t}{\beta_n} \quad \text{where } \beta_n \equiv \frac{(t-1)\beta}{n-1} \end{aligned}$$

and that in coalition C_t – $IC_t(\hat{\mathbf{x}}, C_t)$:

$$y_{t-1} = -\frac{1-\beta}{\beta} \cdot y_t + \frac{u_t}{\beta}$$

Since $\beta < \beta_n < -\alpha < 0$, both $IC_t(\hat{\mathbf{x}}, C_t)$ and $IC_t(\hat{\mathbf{x}}, N)$ have positive slopes, and the former is flatter than the latter. Since they intersect the 45 degree line at the same point, $IC_t(\hat{\mathbf{x}}, C_t)$ rotates inward from $IC_t(\hat{\mathbf{x}}, N)$ (see the two blue lines). Now, notice that in the grand coalition N, player t's indifference set $IC_t(\hat{\mathbf{x}}, N)$ intersects player $(t-1)$'s $IC_{t-1}(\hat{\mathbf{x}}, N)$ at (x_t, x_{t-1}), for by definition, income profile $\hat{\mathbf{x}} \in IC_t(\hat{\mathbf{x}}, N) \cap IC_{t-1}(\hat{\mathbf{x}}, N)$. Similarly, in coalition C_t, t and $(t-1)$'s indifference sets – $IC_t(\hat{\mathbf{x}}, C_t)$ and $IC_{t-1}(\hat{\mathbf{x}}, C_t)$ – both include income profile $\hat{\mathbf{x}}$'s shadow allocation $\hat{\mathbf{y}}$; thus graphically, $IC_t(\hat{\mathbf{x}}, C_t)$ and $IC_{t-1}(\hat{\mathbf{x}}, C_t)$ intersect at $(\hat{y}_t, \hat{y}_{t-1})$.[18] Since $IC_{t-1}(\hat{\mathbf{x}}, C_t)$ rotates downward from $IC_{t-1}(\hat{\mathbf{x}}, N)$, and $IC_t(\hat{\mathbf{x}}, C_t)$ rotates

[18]It is clear that they only intersect once, so their unique intersecting point must be the shadow allocation $\hat{\mathbf{y}}$.

inward from $IC_t(\hat{\mathbf{x}}, N)$, we have $(\hat{y}_t, \hat{y}_{t-1}) << (\hat{x}_t, \hat{x}_{t-1})$. Since $(\hat{x}_t, \hat{x}_{t-1})$ is strictly below the green line, so is $(\hat{y}_t, \hat{y}_{t-1})$, that is, income profile $\hat{\mathbf{x}}$'s shadow allocation $\hat{\mathbf{y}}$ is strictly cheaper than that for the equal income profile. Hence, the randomly chosen unequal income profile $\hat{\mathbf{x}}$ is not critical, and so the only critical income profile is the equal one. Lastly, notice that if $\alpha = 0$, our result still holds – in the graph, we only need to make the two indifference sets for player $(t-1)$ – $IC_{t-1}(\hat{\mathbf{x}}, N)$ and $IC_{t-1}(\hat{\mathbf{x}}, C_t)$ – completely flat, all else follows.

Lemma 2.3 and Lemma 2.4(ii) suggest that when $-\alpha < \beta < (t-1)/t$, a critical income profile $\mathbf{x} \in V(N)$ for any coalition C_t $(1 < t < n)$ satisfies $x_1 \leq x_2 = ... = x_n$ and $\sum_{i=1}^{n} x_i = v_n$. We now characterize the range of β in which out of all the income profiles satisfying the conditions stated above, only the equal one where each player gets v_n/n is critical. Denote the equal income profile by $\mathbf{x}^e$. Notice that its shadow allocation, denoted as $\mathbf{y}^e$, satisfies $y_i^e = v_n/n \ \forall i$, and that any other candidate income profile $\mathbf{x} \in V(N)$ can be expressed as follows:

$$x_i = \begin{cases} x_i^e - (n-1)\Delta & \text{if } i = 1, \\ x_i^e + \Delta & \text{if } i \geq 2, \end{cases} \tag{2.21}$$

where Δ is any positive scalar. Denote the two associated shadow allocations $\mathbf{y}(\mathbf{x}^e, C_t)$ and $\mathbf{y}(\mathbf{x}, C_t)$ by $\mathbf{y}^e$ and $\mathbf{y}$, respectively. We want to find the range of β in which $\mathbf{y}^e$ is more expensive than $\mathbf{y}$. From Proposition 2.2, we obtain:

$$y_2 - y_1 = \theta_1^t(x_2 - x_1) = n\theta_1^t\Delta, \tag{2.22}$$

$$y_{i+1} - y_i = \theta_i^t(x_{i+1} - x_i) = 0 \quad \forall i \geq 2. \tag{2.23}$$

Then, we derive player 1's income y_1 from her indifference condition:

$$\begin{aligned} & u(y_1; \mathbf{y}_{-1}) = u(x_1; \mathbf{x}_{-1}) \\ \Rightarrow \quad & y_1 - \frac{\alpha}{t-1} \sum_{i=2}^{t} (y_i - y_1) = x_1 - \frac{\alpha}{n-1} \sum_{i=2}^{n} (x_i - x_1) \\ \Rightarrow \quad & y_1 - \frac{\alpha}{t-1} \cdot (t-1) n \theta_1^t \Delta = x_1 - \frac{\alpha}{n-1} \cdot (n-1) n \Delta \\ & \Rightarrow \quad y_1 = x_1 - \alpha n \Delta + \alpha n \theta_1^t \Delta, \end{aligned}$$

where the second step is derived from conditions (2.21-2.23). Since (2.22-2.23) also imply that $y_2 = ... = y_t = y_1 + n\theta_1^t \Delta$, we can express players' total income at allocation $\mathbf{y}$ as:

$$\begin{aligned} \sum_{i=1}^{t} y_i &= t\big(x_1 - \alpha n \Delta + \alpha n \theta_1^t \Delta\big) + (t-1) n \theta_1^t \Delta \\ &= t\big(x_1^e - (n-1)\Delta - \alpha n \Delta + \alpha n \theta_1^t \Delta\big) + (t-1) n \theta_1^t \Delta \\ &= \sum_{i=1}^{t} y_i^e - (n-1) t \Delta - \alpha n t \Delta + \alpha n t \theta_1^t \Delta + (t-1) n \theta_1^t \Delta, \end{aligned} \tag{2.24}$$

where the first equality is obtained from condition (2.21), and the second follows from the fact that $y_1^e = ... = y_t^e = x_1^e$. Hence, $\sum_{i=1}^{t} y_i^e > \sum_{i=1}^{t} y_i$ if and only if

$$(n-1)t + \alpha n t - \alpha n t \theta_1^t - (t-1) n \theta_1^t > 0,$$

if and only if

$$\beta < \frac{(1+\alpha) \cdot \big(1/t - 1/n\big)}{\big(1 + \alpha - 1/n\big)/(t-1) - \big(1 + \alpha - 1/t\big)/(n-1)} \equiv \beta^t \in \Big(0, \frac{t-1}{t}\Big)^{19}$$

which is obtained by plugging in the expression for θ_1^t exhibited in Proposition 2.2. In other words, only when $-\alpha < \beta < \beta^t$ is the equal income profile $\mathbf{x}^e$ the unique critical one for coalition C_t. If $\beta = \beta^t$, then $\sum_{i=1}^{t} y_i = \sum_{i=1}^{t} y_i^e$, and so any non-wasteful income

[19] A direct derivative computation tells us that β^t strictly increases in n. Also, $\lim_{n \to \infty} \beta^t = (t-1)/t$.

profile $\mathbf{z} \in V(N)$ for which $z_1 \leq z_2 = ... = z_n$ is critical. If $\beta^t < \beta < (t-1)/t$, then $\sum_{i=1}^{t} y_i > \sum_{i=1}^{t} y_i^e$. By letting $\Delta \to \infty$, we can make income profile $\mathbf{x}$'s shadow allocation $\mathbf{y}$ arbitrarily expensive $\big($see equation (2.24)$\big)$, thus $\mathbf{x}$ becomes the critical income profile. It satisfies $x_1 \to -\infty$, $x_2 = ... = x_n$ and $\sum_{i=1}^{n} x_i = v_n$.

Lastly, we show that when $\beta = -\alpha$, the unique critical income profile for any coalition C_t $(1 < t < n)$ is the equal one. Notice that Lemma 2.3 allows us to restrict attention to the income profiles $\mathbf{x} \in V(N)$ satisfying $x_t = ... = x_n$ and $\sum_{i=1}^{n} x_i = v_n$ (denote this set by T_1). Define a subset of T_1: $T_2 \equiv \big\{\mathbf{x} \in V(N) \big| x_1 = ... = x_{t-1} \leq x_t = ... = x_n,\ \sum_{i=1}^{n} x_i = v_n\big\}$. We first show that every income profile in $T_1 \backslash T_2$ is linked to an income profile in T_2 with equally expensive shadow allocation. Take any income profile $\mathbf{x} \in T_1 \backslash T_2$. Denote by $j < t-1$ the poorest player who earns strictly less than the adjacent player above: $x_1 = x_j < x_{j+1}$. Let $\Delta = (x_{j+1} - x_j) \cdot (t-1-j)/(t-1)$, and construct a new income profile $\mathbf{x}' \in T_1$ according to condition (2.7). Notice that $x'_j = x'_{j+1}$. Since $\beta = -\alpha$, from Proposition 2.2,

$$\theta_j^t = \frac{1 + \alpha \cdot n/(n-1)}{1 + \alpha \cdot t/(t-1)} \quad \forall j < t,$$

thus equation (2.14) suggests $\sum_{i=1}^{t} y'_i = \sum_{i=1}^{t} y_i$ $\big($the derivation of (2.14) follows exactly the same procedure as before$\big)$. If the new income profile $\mathbf{x}' \notin T_2$, we repeat the process above on $\mathbf{x}'$, and keep iterating until reaching an income profile in T_2. Hence, any income profile $\mathbf{x} \in T_1 \backslash T_2$ shares equally expensive shadow allocation with an income profile in T_2. From figure 2.1, the equal income profile where each player gets v_n/n is the unique one in T_2 that maximizes the associated shadow allocation's total income. Hence, it has strictly more expensive shadow allocation than any other income profile $\mathbf{x} \in T_1$ does.[20] In other words, the equal income profile is the unique critical one.

Q.E.D.

[20]Condition (2.7) suggests that any income profile $\mathbf{x} \in T_1 \backslash T_2$ is linked to a strictly unequal income profile in T_2.

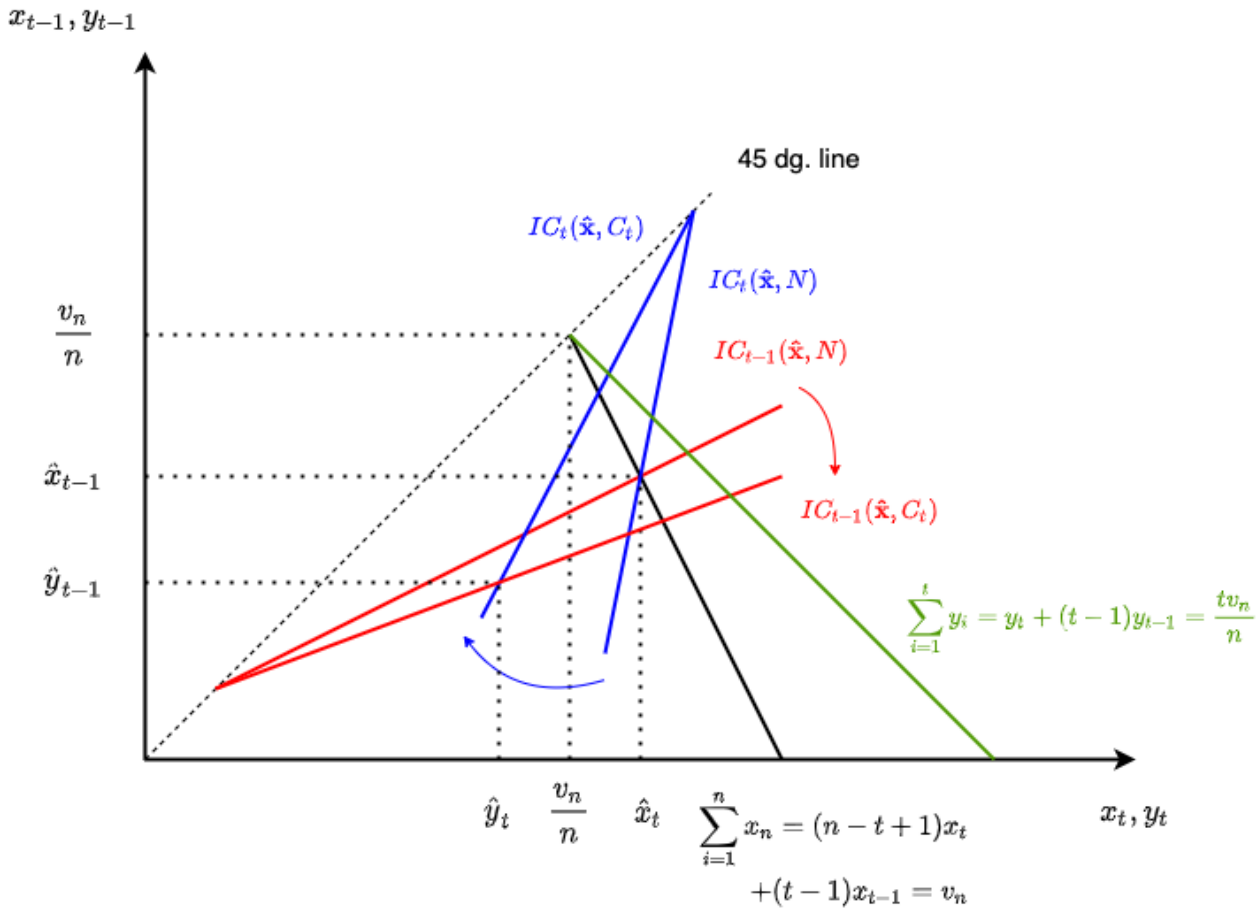


Figure 2.1: Graphic Proof for Theorem 2.1 when $\beta \leq -\alpha$

CHAPTER 3

Reward Pricing in Crowdfunding

3.1 Introduction

As a low-barrier fundraising method, *rewards-based crowdfunding* has become more and more popular, especially among independent artists and innovators eager to market their artworks, designs, inventions and ideas. A crowdfunding project is usually run through an interactive online platform. When launching a project, the *creator* – the person in need of fund to support their initiative – creates a webpage on the designated platform and posts on it the total amount of fund they aim to raise through the project (the *funding goal*), the project starting and ending time, a detailed marketing campaign clarifying the purpose, design, budgeting and expected future progress of their work, as well as a set of related *rewards* with detailed descriptions and associated minimum prices – the least amount of money a donor needs to pay to the creator to obtain the corresponding reward.

There are several different rewards-based crowdfunding platforms, but most crowdfunding activities are done on two websites: *Kickstarter* and *Indiegogo*. While the market share of Kickstarter is considerably larger than that of Indiegogo, they have very similar crowdfunding policies and website design.

The duration of a crowdfunding project ranges from 1 day to 3 months, but is usually around 1 month. Prior to launching a project, the creator can choose from two funding modes. In one mode, they can claim the fund raised through the platform at the end of the project only if the total amount of the fund has exceeded the funding goal by then; otherwise, all the raised fund would be returned to their original donors, leaving nothing to the creator. In another mode, the creator can claim however much is raised during the

crowdfunding period regardless of the set funding goal. Most creators chose the first mode, perhaps because it implies that the donors are charged only if the project is guaranteed sufficient funding, making them more willing to donate. The crowdfunding success rate is averaged at 31% on Kickstarter.

In the language of crowdfunding, we call a project's potential donors *backers*, and sometimes term donating *pledging* or *backing*. A backer browses the crowdfunding platform, examines the project they are interested in, and determines whether they would like to support it through choosing a reward from the list and pledging its minimum price.

Depending on the nature of a project, a reward can either be intangible – say, an appreciative hug from the creator – or involve a substantial product related to the project, such as a newly-released music album. To incentivize backers with different tastes to pledge, a creator usually prepares assorted rewards with minimum prices ranked in accordance with their values.

In this paper, we build a structural model to understand a creator's optimal pricing decision. For a given project, its value and the values of its associated rewards are constant, but each backer has an idiosyncratic *preference level* that leads to dispersion in reward choices. Knowing the distribution of backers' preferences and the project and reward values, the creator chooses a set of (minimum) reward prices to maximize the total amount of fund raised. Given data on reward prices and the numbers of backers buying each reward, we estimate the preference distribution and the project and reward values, and use this information to analyze the creator's optimal pricing strategies.

Since the rewards for a project are partial substitutes, the creator's optimal pricing problem is similar to that of a monopolist selling differentiated products. Using theoretical modeling, [MR78] first study how, in face of consumers with different tastes, a monopolist optimally prices a line of products similar in nature but different in quality. [MR84] study a similar problem with discrete consumer types. Other relevant papers include [MS80] and [OSW84].

This paper is also related to the literature on demand estimation for multiple products

(see [HLZ94], [BLP95], [Pet02], [Nev03] and [Dub05], etc.). Specifically, We build a structural model based on [Bre87] that estimates the demands for the American automobile market with horizontally differentiated consumer types.

Also relevant is the literature on rewards-based crowdfunding. [KBE17] find that the reward supplies are positively correlated with the project success rate, providing favorable evidence for product differentiation in successfully designing a crowdfunding project. [Ste17] uses survey data to identify two motivations for backing a project: the "altruistic motive" and the "purchasing motive," which we take into consideration when building the model. In a behavioral study, [SWT17] identify the "middle-option bias," that is, given a list of rewards, backers tend to choose the middle option. Our model exhibits the potential to test this empirical finding.

We arrange the paper as follows. Section 3.2 introduces the model and defines an equilibrium. Section 3.3 describes the dataset. Section 3.4 proposes the estimation method and presents the estimation results. Section 3.5 discusses several robustness concerns. Section 3.6 concludes.

3.2 Model

We focus on a single crowdfunding project. There are two player types: the *backers* (she) and the *creator* (he). Given the project and its associated rewards, the creator chooses a set of reward prices, and a backer chooses whether to pledge to the project and if so, which reward to buy. Information is perfect. Time variation is abstracted away.

Backers

A backer's motivation for backing a crowdfunding project can be separated into two parts: her support to the project and the utility she derives from consuming the reward. The first part is somewhat altruistic, whereas the second part comes solely from self-interest. For ease of estimation, we assume they are additively separable.

Let $v > 0$ be the value pertaining to the project. Notice that v is not the intrinsic value of the project, but rather, the value a backer derives from donating to the project she deems meaningful. If the project is financed by a significant number of backers, then v is much smaller than – in fact, negligible compared to – the intrinsic project value. Let K be the total number of rewards for the project. Each reward $k \in \{1, 2, ..., K\}$ has value $w_k > 0$. We assume that each reward is valued distinctly and index the rewards by value in ascending order: $w_1 < w_2 < ... < w_K$. Both v and w_k's are constants.

Each backer i has an idiosyncratic *preference level* $\alpha^i > 0$. It is distributed i.i.d. with support on $[\underline{\alpha}, \bar{\alpha}]$, where $\bar{\alpha} > \underline{\alpha} \geq 0$, and according to the c.d.f. $F : [\underline{\alpha}, \bar{\alpha}] \to [0, 1]$.

We assume that each backer buys at most one reward. Let p_k be the price of reward k. If backer i backs the project and buys reward k, she derives utility $\alpha^i(v + w_k) - p_k$.

Let $\hat{\alpha}_k$ be the preference level at which a backer is indifferent between choosing reward $(k-1)$ and reward k, then:

$$\begin{aligned} \hat{\alpha}_k(v + w_k) - p_k = \hat{\alpha}_k(v + w_{k-1}) - p_{k-1} \quad &\Rightarrow \\ \hat{\alpha}_k = \frac{p_k - p_{k-1}}{w_k - w_{k-1}} \quad &\text{for } k = 2, ..., K. \end{aligned} \tag{3.1}$$

Kickstarter – the source of our data – has the policy that each participating backer has to pledge at least \$1. Backers who pledge less than p_1 – the price of the cheapest reward – receive nothing. They back the project merely to support the creator's initiative. Using this information, we can derive two additional indifference conditions:

$$\hat{\alpha}_1(v + w_1) - p_1 = \hat{\alpha}_1 v - p_0, \tag{3.2}$$

$$\hat{\alpha}_0 v - p_0 = 0. \tag{3.3}$$

(3.2) characterizes the preference level $\hat{\alpha}_1$ that makes a backer indifferent between buying the cheapest reward and backing the project with no reward. (3.3) – the participation constraint – characterizes the lowest preference level $\hat{\alpha}_0$ at which a backer is willing to back the project.

For the cutoff preference levels $\hat{\alpha}_k$'s to be positive and hence properly defined, it is necessary to assume that $1 = p_0 < p_1 < ... < p_k$, i.e., a more valuable reward is priced higher. This assumption is largely consistent with reality. A stronger assumption is required to ensure the existence of a separating equilibrium where a more valuable reward attracts backers with higher preference levels. We will discuss the details in the creator's problem below.

Notice that, by defining a backer to be utility-maximizing, we have implicitly assumed that all backers pay the posted minimum reward prices. In reality, the amount of the pledge a backer pays sometimes exceeds the minimum price of her chosen reward. We can think of the "residual pledge" beyond the minimum price as an individual-specific donation coming into play only after the backer has made the rational choice, thus irrelevant to our study.

Creator

The aim of the creator is to solicit as much fund as possible to finance his work. In reality, he chooses three elements when launching a crowdfunding project: the funding goal, the project duration and the reward prices. In this model, we focus on the reward prices and abstract away from the others.

Assume that, prior to launching the project, the creator has conducted thorough market research and so knows the distribution of backers' preference levels $F(\alpha)$. In addition, he has obtained K different rewards, each with sufficient supply to meet the market demand. Knowing the project value v and the reward values $(w_k)_{k=1}^K$, the creator chooses a set of reward prices $(p_k)_{k=1}^K$ to maximize the total profit:

$$\max_{(p_k)_{k=1}^K} \sum_{k=0}^{K-1} \left[F(\hat{\alpha}_{k+1}) - F(\hat{\alpha}_k)\right] \cdot (p_k - w_k) + \left[1 - F(\hat{\alpha}_K)\right] \cdot (p_K - w_K) \tag{3.4}$$

subject to the constraints $p_0 = 1$, $w_0 = 0$ and $0 \leq \hat{\alpha}_1 \leq \hat{\alpha}_2 \leq ... \leq \hat{\alpha}_K$, where each $\hat{\alpha}_k$ is defined by equations (3.1-3.3).

$\left[F(\hat{\alpha}_{k+1}) - F(\hat{\alpha}_k)\right]$ proportion of backers choose reward k for $k = 0, ..., K-1$. $\left[1 - F(\hat{\alpha}_K)\right]$

proportion of backers choose reward K. Per reward k sold, the creator derives profit $(p_k - w_k)$. One way to think about it is to consider w_k as reward k's market value. The creator charges a price p_k beyond the market value w_k and receives the premium $(p_k - w_k)$ as the donation.

The last constraint of the profit maximization problem ensures that the solution to (3.4) is a properly-defined separating equilibrium where a backer with higher preference level chooses a reward with higher value. In other words, the price differences $(p_k - p_{k-1})$'s have to be large enough to select the backers into different rewards in accordance with their preference levels.

Equilibrium

In order to solve for a unique separating equilibrium, the following assumptions regarding the preference distribution $F(\alpha)$ are imposed:

Assumption 3.1. *i* F *is everywhere twice differentiable on* $(\underline{\alpha}, \bar{\alpha})$ *with p.d.f.* f*. Moreover,* $2f(\alpha) + (\alpha - 1)f'(\alpha) > 0$ *for all* $\alpha \in (\underline{\alpha}, \bar{\alpha})$*.*

ii $1 - F(\alpha) = (\alpha - 1)f(\alpha)$ *has* K *distinct positive solutions. In addition, its smallest solution is no less than* $1/v$*.*

Assumption 3.1(i) ensures the creator's problem (3.4) is concave and can be solved by the first order conditions:

$$1 - F(\hat{\alpha}_k) = (\hat{\alpha}_k - 1)f(\hat{\alpha}_k) \quad \text{for } k = 1, ..., K. \tag{3.5}$$

The first part of Assumption 3.1(ii) ensures that equations (3.5) are solvable. Coupled with the constraint $0 \leq \hat{\alpha}_1 \leq \hat{\alpha}_2 \leq ... \leq \hat{\alpha}_K$, the cutoff preference levels $(\hat{\alpha}_k)_{k=1}^K$ can be uniquely identified. The second part of 3.1(ii) ensures that $\hat{\alpha}_1$ is no less than $\hat{\alpha}_0 = 1/v$ given by the participation constraint (3.3).

Once the equilibrium cutoff preference levels $(\hat{\alpha}_k)_{k=0}^K$ are known, the equilibrium prices $(p_k)_{k=1}^K$ can be directly backed out from the indifference conditions (3.1-3.2).

We formally define the equilibrium below.

Definition 3.1 (***Equilibrium***). *Let* $(\alpha^{(k)})_{k=1}^{K}$ *be the* K *positive solutions of* $1 - F(\alpha) = (\alpha - 1)f(\alpha)$ *such that* $\alpha^{(1)} < \alpha^{(2)} < ... < \alpha^{(K)}$. *Then in equilibrium,*

i a backer backs the project if and only if her preference level is at least $1/v$. *Backers with preference levels in* $[1/v, \alpha^{(1)})$ *back the project without reward; those with preference levels in* $[\alpha^{(k)}, \alpha^{(k+1)})$ *choose reward* k *for* $k = 1, ..., K-1$*; and those with preference levels no less than* $\alpha^{(K)}$ *choose reward* K*;*

ii the creator chooses prices $(p_k)_{k=1}^{K}$ *satisfying the following set of formulas:* $p_k = \alpha^{(k)}(w_k - w_{k-1}) + p_{k-1}$ *for* $k = 1, ..., K$, *where* $p_0 = 1$ *is given.*

The assumption that $1 - F(\alpha) = f(\alpha)(\alpha - 1)$ has exactly K positive solutions is rather strict. However, if the creator internalizes the number of rewards K as one of his optimal choices, then this is nothing but a natural maximizing outcome. In this case, we need only to assume that at least one positive solution exists.

3.3 Data

The database [Li19] we use is generously provided by *CrowdBerkeley* – a leading research initiative on rewards-based crowdfunding at the University of California, Berkeley.

This is a structured relational database collected on January 30, 2019. It records a total of 408,637 crowdfunding projects launched on Kickstarter between April 21, 2009 and January 30, 2019.

The main table *project* records the main statistics for each project, including the funding goal, the starting and ending time, the total amount of fund raised, the total number of backers, the project status, creator information, country and location information, and marketing information.

The table most important to our study is *reward*. This table provides information on the rewards per project recorded in the main table. It gives us data on reward descriptions,

minimum prices, number limits, backer counts and shipping information. We are primarily interested in minimum prices and backer counts.

Besides, the *category* table summarizes all the project categories, the *location* table records creators' countries and locations, and the *creator* table contains detailed information on each project creator. All these tables reveal useful information for our study.

3.4 Estimation

Identification

Per project, the data provide us with its reward prices $(p_k)_{k=1}^K$, the total number of backers M^{in} pledging to the project, and the numbers of backers choosing each reward $(M_k)_{k=1}^K$. The number of backers pledging less than p_1 and receiving no reward is easily backed out through $M_0 = M^{in} - \sum_{k=1}^K M_k$. The unknowns we need to estimate are the project and reward values $\big(v, (w_k)_{k=1}^K\big)$, the cutoff preference levels $(\hat{\alpha}_k)_{k=0}^K$ and the two preference boundaries $(\underline{\alpha}, \bar{\alpha})$, as well as the number of backers choosing not to back the project M^{out} and the *market size* $M = M^{in} + M^{out}$.

An obvious estimation approach is to utilize the creator's first order conditions (3.5) to compute the cutoff preference levels $(\hat{\alpha}_k)_{k=0}^K$, and plug them into backers' indifference conditions (3.1-3.3) to compute the project and reward values $\big(v, (w_k)_{k=1}^K\big)$. Though the c.d.f. values of the cutoff preference levels $\big(F(\hat{\alpha}_k)\big)_{k=0}^K$ are not known due to the lack of data on the market size M, their ratios can be identified with the numbers of backers choosing each reward $(M_k)_{k=1}^K$. The main difficulty lies in identifying the corresponding p.d.f. values $\big(f(\hat{\alpha}_k)\big)_{k=0}^K$. A simple "local uniform" assumption on the preference distribution $F(\alpha)$ facilitates the recovering of these values:

Assumption 3.2. *A backer's preference level α is distributed uniformly on each interval $[\hat{\alpha}_k, \hat{\alpha}_{k+1})$ for $k = 1, ...K-1$, with slight perturbations around the cutoffs $\hat{\alpha}_k$'s such that Assumption 3.1(i) is satisfied.*

Under Assumption 3.2, the p.d.f. value of each cutoff preference level $f(\hat{\alpha}_k)$ equals $\big[F(\hat{\alpha}_{k+1}) - F(\hat{\alpha}_k)\big]/(\hat{\alpha}_{k+1} - \hat{\alpha}_k)$ for $k = 1, ..., K-1$. Hence, we can rewrite the first order conditions (3.5) as:

$$\frac{\hat{\alpha}_{k+1} - \hat{\alpha}_k}{\hat{\alpha}_k - 1} = \frac{F(\hat{\alpha}_{k+1}) - F(\hat{\alpha}_k)}{1 - F(\hat{\alpha}_k)} \equiv d_k \quad \Rightarrow$$
$$\hat{\alpha}_{k+1} = d_k(\hat{\alpha}_k - 1) + \hat{\alpha}_k \quad \text{for } k = 1, ...K-1. \tag{3.6}$$

Equations (3.6) give us the relationship between each pair of neighboring cutoff preferences, where each $d_k = M_k/\big(\sum_{s=k}^{K} M_s\big)$ can be directly computed from data. However, the cutoff preferences $\hat{\alpha}_k$'s can still not be identified without a locational assumption. Since a crowdfunding project is typically created by an individual or entity lacking wide recognition and dedicated to a specific cause, it is reasonable to consider its backers as a cult following, the members of which value the project and its associated rewards beyond their market values – the values an agent with preference level $\alpha = 1$ would obtain. Therefore, we assume the lower boundary of the preference distribution $\underline{\alpha}$ is slightly above 1. This locational assumption is formalized below.

Assumption 3.3. *The lower preference boundary $\underline{\alpha} = 1+\epsilon$, where $\epsilon > 0$ is arbitrarily small; moreover, the preference level α is distributed uniformly on $[\underline{\alpha}, \hat{\alpha}_2)$ and $[\hat{\alpha}_K, \bar{\alpha}]$, respectively, with slight perturbations around the cutoffs such that Assumption 3.1(i) is satisfied.*

Assumption 3.3 also ensures that the cutoffs $\hat{\alpha}_k > 1$ for all $k \geq 1$ – a necessary condition for the first order conditions (3.5) to be well-defined. In order to add an effective boundary condition to facilitate model estimation, $\underline{\alpha}$ is chosen to be slightly larger than 1 instead of equal to 1, and the "local uniform" assumption is extended to the lower and upper boundaries. For example, we can set $\epsilon = 0.05$, so that $\underline{\alpha} = 1.05$.

With Assumption 3.3, the cutoff $\hat{\alpha}_1$ can be pinned down through:

$$\frac{\hat{\alpha}_2 - \hat{\alpha}_1}{M_1} = \frac{\hat{\alpha}_1 - \underline{\alpha}}{M^{out} + M_0} \tag{3.7}$$

where $\hat{\alpha}_2 = d_1(\hat{\alpha}_1 - 1) + \hat{\alpha}_1$.

We are left with computing the outside share of the market M^{out}. We would have information on it if we could observe the number of visits to the crowdfunding project webpage. However, such data is not available. Given the cult nature of the target market and the low pledging barrier at \$1, the outside share M^{out} should be small. We set $M^{out} = 0.1M$, i.e., the outside share is 10%.

Once M^{out} is known, we can compute the cutoff $\hat{\alpha}_1$ through equation (3.7) and, consequently, the other cutoffs $\left(\hat{\alpha}_0, (\hat{\alpha}_k)_{k=2}^K\right)$ using the first order conditions (3.6). The project and reward values $\left(v, (w_k)_{k=1}^K\right)$ are then backed out through the indifference conditions (3.1-3.3).

Example

To accurately estimate the preference distribution $F(\alpha)$, we need abundant observations of backers. Thus, we restrict attention to those projects with at least 10,000 backers and funding goal greater than \$5,000. Besides, we focus on the projects launched in or after 2014. We only consider successful projects, as it is an indicator that the project creator had made wise pricing decisions.

We only preserve the projects with at least 6 rewards to ensure sufficient preference differentiation. Since many rewards had limited copies and were sold out before the funding period ended, in order not to distort the estimation outcome, we only consider those projects for which the rewards were either unlimitedly offered or not sold out during the crowdfunding period.

Though the dataset records more than 400,000 project observations in total, the restrictions above narrow them down to only 37 projects categorized in figure 3.1.

Kickstarter groups its projects into 15 categories. But those satisfying our requirements come from only 6 categories, with design, games, and technology having predominant shares. More than half of these projects are for gaming products.

We select one project from the list to perform the estimation. It is titled "Tak: A

parent_slug	Freq.	Percent	Cum.
design	9	24.32	24.32
film & video	1	2.70	27.03
food	1	2.70	29.73
games	20	54.05	83.78
music	1	2.70	86.49
technology	5	13.51	100.00
Total	37	100.00	

Figure 3.1: Categories of the Target Projects

Beautiful Game." This crowdfunding project was launched by *Cheapass Games* in April 19, 2016 and lasted for 31 days. During the funding period, 12,187 backers pledged to the project, helping it to raise fund totaling $1,351,142, far exceeding its goal of $50,000.

Cheapass Games is a gaming company funded and run by James Ernest since 1996, based in Seattle, Washington. Headed by an experienced professional game designer, the company has designed and released dozens of games of different types, including board games, card games and white box games [Wik].

Cheapass Games launched its first crowdfunding project on Kickstarter in 2012, and since then has run 12 projects on this platform, all successfully funded. The funding goals of their projects were averaged at $27,909.09, with the lowest at $10,000. An average fund of $232,070 was raised per project. The least funded project solicited $43,714.5. The data speak to the company's successful crowdfunding management. Undoubtedly, it has a stable target market on Kickstarter.

"Tak: A Beautiful Game" was the 9th project Cheapass Games launched, which among all its projects collected the largest amount of fund from the largest number of backers. What the company wanted to sell through the project was *Tak* – a two-player abstract strategy game invented by James Ernest and Patrick Rothfuss, based on Rothfuss' fantasy novel *The Wise Man's Fear*. During the funding period, the company was still actively designing Tak

and posted several updates. Since backers backed this project both to support the company's future game design and to obtain their products in desire, this project can be suitably applied to our model [Gam].

The project offered several products related to the game, each accompanied by detailed description, such as the Companion Book delineating the game rule, different types of game boards/sets, and collector coins. There were 11 rewards in total. We rank them by price in ascending order. Some rewards were product bundles, which were charged simply at the total price of the included products. For instance, reward 2 was a companion book worth $15, reward 3 was a Tavern Set for $25, and reward 4 offered a combination of both products at $40. A complete list of the rewards' rankings, contents, minimum prices and backer counts (from left to right) is presented in figure 3.2.

reward~k	reward	pk	Mk
0	none	1	95
1	access to Pledge Manager	5	621
2	Companion Book	15	614
3	Tavern Set	25	594
4	Tavern Set and Companion Book	40	1463
5	Classic Set	55	2056
6	Classic Set and Companion Book	65	2878
7	Classic, Tavern, Book and Coin	90	2004
8	Tinker's Pack plus Archanist's Board	155	982
9	Tinker's Pack plus Devi's Board	250	328
10	Devi's Box, a handcrafted hardwood set, Book	475	170
11	Tinker's Pack plus Devi's Box	550	382

Figure 3.2: List of the Rewards' Contents, Minimum Prices and Backers Counts

The first row shows that 95 backers pledged less than $5 and received no reward. With the exception of reward 1 that granted access to the Pledge Manager, all the rewards offered some physical products or product bundles. The price gap between each two neighboring rewards was either $10 or $15 in the lower and middle tiers and increased tremendously in the upper tier – a common pricing strategy in crowdfunding.

We now estimate our model using the project data. The estimation results are presented

in figure 3.3. Starting from the left, each column represents, respectively, the reward ranks k, the cutoff preference levels $\hat{\alpha}_k$, the cutoff c.d.f. values $F(\hat{\alpha}_k)$, and the reward values w_k. The two boundaries of the preference distribution $\underline{\alpha}$, $\bar{\alpha}$ are shown in the first and last rows.

reward~k	a	cdf	wk
.	1.05	0	.
0	1.056361	.0999926	0
1	1.056807	.1070083	3.784985
2	1.059725	.1528691	13.2214
3	1.062922	.1982128	22.62943
4	1.066364	.2420796	36.69592
5	1.075824	.3501219	50.63872
6	1.09354	.501957	59.78334
7	1.133458	.7144967	81.83974
8	1.202637	.8624917	135.8876
9	1.309506	.9350122	208.4341
10	1.424868	.9592349	366.3435
11	1.555714	.9717894	414.5528
.	2.111429	1	.

Figure 3.3: Model Estimation Results

All estimated reward values w_k's are smaller than their corresponding minimum prices p_k's. To better understand the creator's pricing strategy, let us take a close look at the project's preference distribution $F(\alpha)$.

Figure 3.4 plots the c.d.f. of the preference distribution $F(\alpha)$. The curve is very steep in the lower range but much smoother above around $\alpha = 1.2$. It indicates that the market for this project consisted primarily of backers with comparatively low preference levels. Though

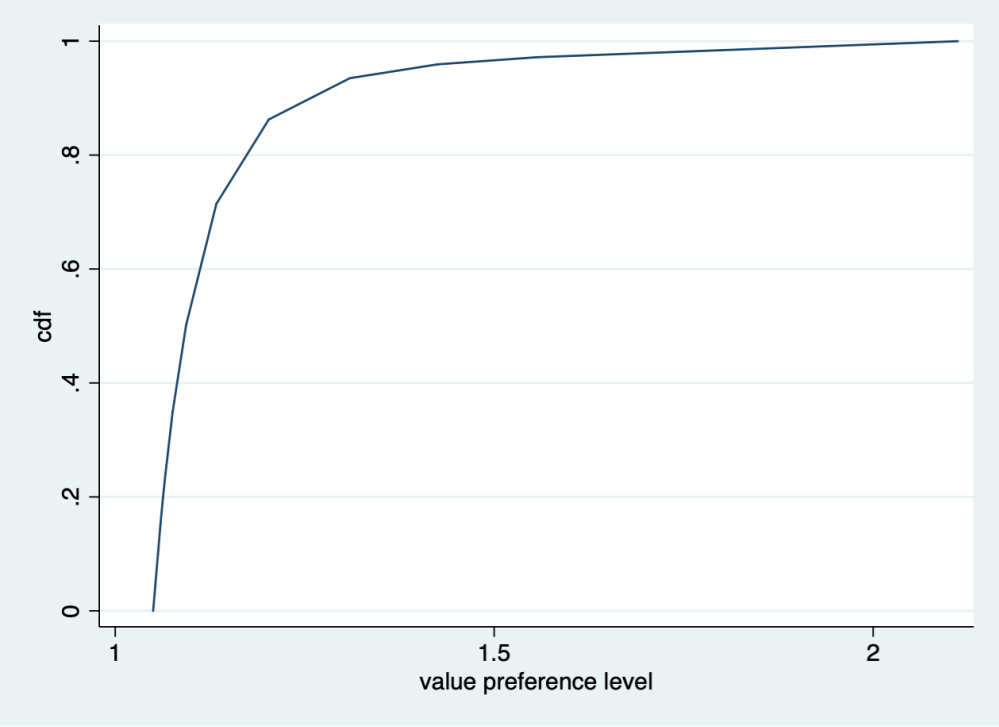


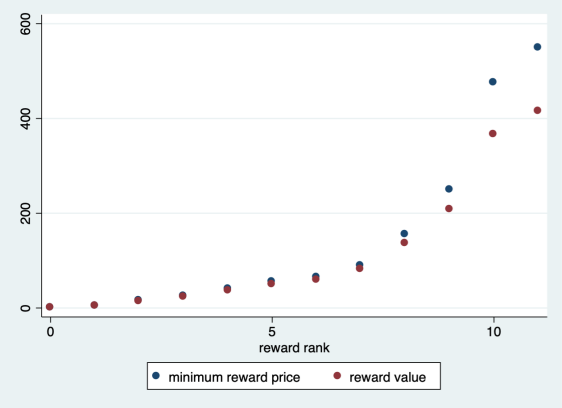
Figure 3.4: C.D.F. of the Preference Distribution $F(\alpha)$

there existed backers with exceptionally high preferences reaching up to 2.1, around 85% of the backers had preference levels lower than 1.2 – much closer to the lower boundary.

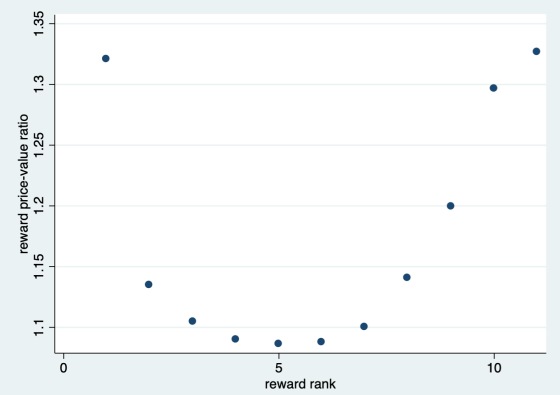
One would expect the creator to have prepared rewards with large value and price gaps to attract backers with different preference levels. Indeed, figures 3.2 shows that the price gap between each two neighboring rewards was notably higher starting at reward 7, and figure 3.3 shows that 86% of the backers either chose rewards with prices lower than that of reward 8 or did not pledge at all. The pricing strategy effectively differentiated the market.

Figure 3.5 compares the reward price p_k and value w_k along the reward rank k. 3.5a confirms our previous finding that the price gaps on the upper tier are steeper. Such trends also apply to the reward values w_k. Besides, the price-value differences are much larger for higher-ranked rewards, indicating that the creator extracted considerably more surplus from backers with higher preferences – compared to the majority of the backers who had low preference levels, they resembled donors more than consumers.

3.5b plots the rewards' price-value ratios p_k/w_k against their ranks k. The convex shape



(a) p_k, w_k on k



(b) p_k/w_k on k

Figure 3.5: Comparison between p_k and w_k along k

of the curve implies that the price-value ratios approach 1 for the middle-ranked rewards and are much higher for those with low and high ranks. Interestingly, the creator funded his project predominantly from the "generous donors" placing exceptionally high values on his work and the "indifferent multitude" with only lukewarm interest in supporting his work and reluctant even to pay for the products with medium values, while charging little or no premiums from the backers standing in the middle.

Figure 3.6 demonstrates the near-linear relationship between the reward values w_k and the reward prices p_k, with slope estimated at 1.32.

The project value v is estimated at 0.95, lower even than the value of the cheapest reward. It suggests the importance of providing rewards to incentivize donation. For backers, the motivation for funding the project came from their relatively high preference levels, which were well above 1 and effectually boosted up the subjective project and reward values. The contribution of the objective project value v was trivial. The multiplicative nature of backers' preferences α justifies the offering of rewards, as the creator could significantly raise backers' willingness to pay by providing rewards much more valuable than the project itself.

3.5 Robustness

This section enumerates several robustness concerns for model specification and estimation.

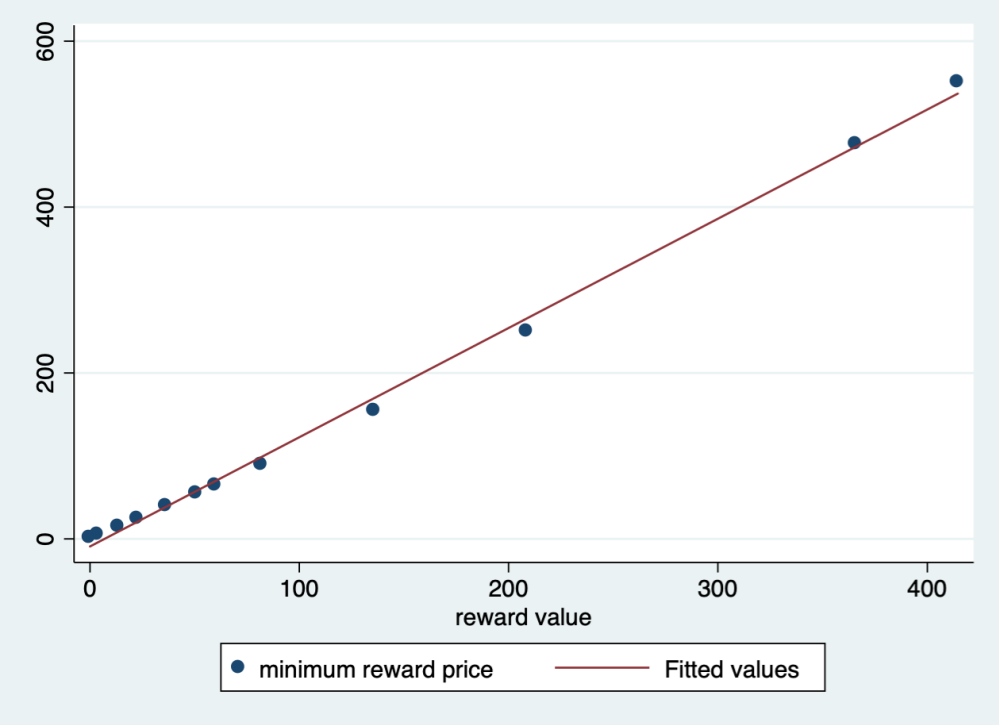


Figure 3.6: p_k over w_k

In the model, the project and reward values v, w_k's are multiplied by the same preference level α. While facilitating calibration, this assumption is not necessarily consistent with reality and imposes restriction on model estimation. However, without it, it is impossible to identify and compare the two forces given limited data. An alternative specification is to allow each backer to have two different, possibly correlated preference levels for the project value v and reward values w_k's, respectively. Ideally, a richer dataset (perhaps one that allows us to vary v within a project) would enable us to identify both preference distributions, pin down their correlation patterns and discern which effect is dominant.

The "local uniform" assumption for the preference distribution, while serving as an expedient for estimation, is quite arbitrary. A slight modification to the truncation of intervals could drastically change the estimation results. For example, if the preference level α is distributed uniformly on each $(\hat{\alpha}_k, \hat{\alpha}_{k+1}]$, and not $[\hat{\alpha}_k, \hat{\alpha}_{k+1})$ as assumed, then all the cutoff p.d.f. values $f(\hat{\alpha}_t)$'s need to be recalculated, which would consequently change all our estimation results. The "local uniform" assumption would be more reliable if the price gaps are

small. While some projects set their price gaps within \$1, this is certainly not the case for the example we use. Perhaps a more reasonable estimator for the cutoff p.d.f. values is the average of the p.d.f.s for the uniform distributions over two adjacent intervals.

Also dubious is the locational assumption. The cutoff preference levels are pinned down through guessing the outside market share M^{out} and the lower boundary for the preference distribution $\underline{\alpha}$. We choose them as the fixed points because it is comparatively easier to justify their guessed values. However, the guesses are not corroborated with data evidence. Hopefully, with the help of richer data, we could find a better approach to fix the location in the future.

In the model, we assume that each backer pledges at most once per project. This is not necessarily the case, and the equilibrium would be characterized differently should we allow a backer to purchase more than one reward for each project. We cannot incorporate in this possibility because we do not know backers' identities, which are confidential information withheld by Kickstarter. Nevertheless, we think it a reasonable assumption that most backers pledged only once to a project. So, the effect of not counting in the repetitive purchases is minor.

Since our data were collected at one point of time, we are not able to identify time variations in the project and reward values as well as the preference distribution, which may contain interesting information.

In the creator's profit maximization problem, the profit per reward sold is represented by the reward's minimum price minus its value, deviating from the usual setting where the per unit profit equals the price minus the cost. This setting would be ideal to adopt should we have data on cost information. As explained above, our specification can be justified by treating the reward values as the market values of rewards – the equilibrium reward prices in a normal market. The creator tries to solicit as much fund as possible from his devoted backers by charging them premiums beyond the normal prices.

3.6 Conclusion

This paper studies a creator's optimal pricing strategies in the rewards-based crowdfunding industry. We construct a structural model, propose an estimation strategy for backers' preference distribution and the project and reward values, and demonstrate the estimation results with a selected example. In the example, the preference distribution weighs towards the lower side, and the creator follows a bow-shaped pricing strategy where the price-value ratio is close to one at the medium-ranked rewards and much higher on the two ends.

While we have only examined one example, the model demonstrates the potential to analyze multiple examples, identify different patterns of preference distributions, and compare various optimal pricing strategies. We predict that projects in the same category would exhibit similar patterns.

We have ignored some other information in the dataset possibly correlated with the project and reward values, such as whether the project was spotlighted. Incorporating in this information may help generate more accurate results.

REFERENCES

[AL12] James E Alt and David Dreyer Lassen. "Enforcement and public corruption: Evidence from the American states." *The Journal of Law, Economics, and Organization*, **30**(2):306–338, 2012.

[AMO16] Daron Acemoglu, Azarakhsh Malekian, and Asu Ozdaglar. "Network security and contagion." *Journal of Economic Theory*, **166**:536–585, 2016.

[BB08] Mariagiovanna Baccara and Heski Bar-Isaac. "How to organize crime." *The Review of Economic Studies*, **75**(4):1039–1067, 2008.

[BBK21] Oriana Bandiera, Michael Carlos Best, Adnan Qadir Khan, and Andrea Prat. "The allocation of authority in organizations: A field experiment with bureaucrats." *The Quarterly Journal of Economics*, **136**(4):2195–2242, 2021.

[BFF06] Helen Bernhard, Urs Fischbacher, and Ernst Fehr. "Parochial altruism in humans." *Nature*, **442**(912-915), 2006.

[BLP95] Steven Berry, James Levinsohn, and Ariel Pakes. "Automobile prices in market equilibrium." *Econometrica*, **63**(4):841–890, 1995.

[BM93] Timothy Besley and John McLaren. "Taxes and bribery: The role of wage incentives." *The Economic Journal*, **103**(416):119–141, 1993.

[BO00] Gary E Bolton and Axel Ockenfels. "ERC: A theory of equity, reciprocity, and competition." *American Economic Review*, **90**(1):166–193, 2000.

[Bre87] Timothy F Bresnahan. "Competition and collusion in the American automobile industry: The 1955 price war." *The Journal of Industrial Economics*, **35**(4):457–482, 1987.

[BV21] Simon Board and Moritz Meyer ter Vehn. "Learning dynamics in social networks." *Econometrica*, **89**(6):2601–2635, 2021.

[CFG07] James C Cox, Daniel Friedman, and Steven Gjerstad. "A tractable model of reciprocity and fairness." *Games and Economic Behavior*, **59**(1):17–45, 2007.

[CMM12] David Card, Alexandre Mas, Enrico Moretti, and Emmanuel Saez. "Inequality at work: The effect of peer salaries on job satisfaction." *American Economic Review*, **102**(6):2981–3003, 2012.

[CP22] Zoë Cullen and Ricardo Perez-Truglia. "How much does your boss make? The effects of salary comparisons." *Journal of Political Economy*, **130**(3):766–822, 2022.

[CW92] Parkash Chander and Louis Wilde. "Corruption in tax administration." *Journal of Public Economics*, **49**(3):333–349, 1992.

[DGL19] Arindrajit Dube, Laura Giuliano, and Jonathan Leonard. "Fairness and frictions: The impact of unequal raises on quit behavior." *American Economic Review*, **109**(2):620–663, 2019.

[DGP13] Esther Duflo, Michael Greenstone, Rohini Pande, and Nicholas Ryan. "Truth-telling by third-party auditors and the response of polluting firms: Experimental evidence from India." *The Quarterly Journal of Economics*, **128**(4):1499–1545, 2013.

[DT13] Bin Dong and Benno Torgler. "Causes of corruption: Evidence from China." *China Economic Review*, **26**(152-169), 2013.

[Dub05] Jean-Pierre Dubé. "Product differentiation and mergers in the carbonated soft drink industry." *Journal of Economics & Management Strategy*, **14**(4):879–904, 2005.

[FBR08] Ernst Fehr, Helen Bernhard, and Bettina Rockenbach. "Egalitarianism in young children." *Nature*, **454**:1079–1083, 2008.

[FLM03] Antoine Faure-Grimaud, Jean-Jacques Laffont, and David Martimort. "Collusion, delegation, supervision and soft information." *The Review of Economic Studies*, **70**(2):253–279, 2003.

[FS99] Ernst Fehr and Klaus M Schmidt. "A theory of fairness, competition, and cooperation." *The Quarterly Journal of Economics*, **114**(3):817–868, 1999.

[Gam] Cheapass Games. "Tak: A beautiful game." Kickstarter, `https://www.kickstarter.com/projects/cheapassgames/tak-a-beautiful-game`. Accessed: September 2019.

[Gil59] Donald B Gillies. "Solutions to general non-zero-sum games." *Contributions to the Theory of Games*, **4**(40):47–85, 1959.

[GN98] Rajeev K Goel and Michael A Nelson. "Corruption and government size: A disaggregated analysis." *Public Choice*, **97**:107–120, 1998.

[GN11] Rajeev K Goel and Michael A Nelson. "Measures of corruption and determinants of US corruption." *Economics of Governance*, **12**:155–176, 2011.

[GR89] Rajeev K Goel and Daniel P Rich. "On the economic incentives for taking bribes." *Public Choice*, **61**:269–275, 1989.

[GS06] Edward L Glaeser and Raven E Saks. "Corruption in America." *Journal of Public Economics*, **90**(6-7):1053–1072, 2006.

[HLZ94] Jerry Hausman, Gregory Leonard, and J Douglas Zona. "Competitive analysis with differenciated products." *Annales d'Économie et de Statistique*, **34**:159–180, 1994.

[Jac08] Matthew O Jackson. *Social and economic networks*. Princeton University Press, 2008.

[KBE17] Michael Marcin Kunz, Ulrich Bretschneider, Max Erler, and Jan Marco Leimeister. "An empirical investigation of signaling in reward-based crowdfunding." *Electronic Commerce Research*, **17**:425–461, 2017.

[KL93] Fred Kofman and Jacques Lawarrée. "Collusion in hierarchical agency." *Econometrica*, **61**(3):629–656, 1993.

[KRS06] Gökhan R Karahan, Laura Razzolini, and William F Shughart. "No pretense to honesty: County government corruption in Mississippi." *Economics of Governance*, **7**:211–227, 2006.

[KS14] Atsushi Kato and Takahiro Sato. "The effect of corruption on the manufacturing sector in India." *Economics of Governance*, **15**:155–178, 2014.

[KW16] Xiaoqi Kong and Yuqian Wang. "Jiazhong zaojie qianchu tanfu Shanxijiaomei yuandongshizhang Bai Peizhong huoxing shisannianban." Caixin, `https://china.caixin.com/2016-12-21/101029499.html`, December 2016.

[Leo00] David Leonhardt. "Andersen split into two firms by arbitrator." The New York Times, `https://www.nytimes.com/2000/08/08/business/andersen-split-into-two-firms-by-arbitrator.html`, August 2000.

[Li19] Guan-Cheng Li. "Kickstarter structured relational database." Harvard Dataverse, V2, `https://doi.org/10.7910/DVN/EOYBXM`, January 2019.

[MH92] Kenneth J Meier and Thomas M Holbrook. ""I seen my opportunities and I took'em:" Political corruption in the American states." *The Journal of Politics*, **54**(1):135–155, 1992.

[MO19] Sauro Mocetti and Tommaso Orlando. "Corruption, workforce selection and mismatch in the public sector." *European Journal of Political Economy*, **60**:101809, 2019.

[MP95] Dilip Mookherjee and Ivan Paak-Liang Png. "Corruptible law enforcers: How should they be compensated?" *The Economic Journal*, **105**(428):145–159, 1995.

[MR78] Michael Mussa and Sherwin Rosen. "Monopoly and product quality." *Journal of Economic Theory*, **18**:301–317, 1978.

[MR84] Eric Maskin and John Riley. "Monopoly with incomplete information." *The RAND Journal of Economics*, **15**(2):171–196, 1984.

[MS80] Leonard J Mirman and David Sibley. "Optimal nonlinear prices for multiproduct monopolies." *The Bell Journal of Economics*, **11**(2):659–670, 1980.

[Nev03] Aviv Nevo. "Measuring market power in the ready-to-eat cereal industry." *Econometrica*, **69**(2):307–342, 2003.

[NM04] John von Neumann and Oskar Morgenstern. *Theory of games and economic behavior*. Princeton University Press, 60th anniversary commemorative edition, 2004.

[OC18] Juan Ortner and Sylvain Chassang. "Making corruption harder: Asymmetric information, collusion, and crime." *Journal of Political Economy*, **126**(5):2108–2133, 2018.

[OSW84] Shmuel Oren, Stephen Smith, and Robert Wilson. "Pricing a product line." *The Journal of Business*, **57**(1):73–99, 1984.

[Pet02] Amil Petrin. "Quantifying the benefits of new products: The case of the minivan." *Journal of Political Economy*, **110**(4):705–729, 2002.

[Pun00] Maurice Punch. "Police corruption and its prevention." *European Journal on Criminal Policy and Research*, **8**:301–324, 2000.

[Pun09] Maurice Punch. *Police corruption: Deviance, accountability and reform in policing*. Routledge, 2009.

[Sad20] Evan Sadler. "Diffusion games." *American Economic Review*, **110**(1):225–270, 2020.

[SS53] Lloyd S Shapley and Martin Shubik. "Solutions of N-person games with ordinal utilities." *Econometrica*, **21**(2):348–349, 1053.

[SSZ16] Günther G Schulze, Bambang Suharnoko Sjahrir, and Nikita Zakharov. "Corruption in Russia." *The Journal of Law and Economics*, **59**(1):135–171, 2016.

[Ste17] Norbert Steigenberger. "Why supporters contribute to reward-based crowdfunding." *International Journal of Entrepreneurial Behavior & Research*, **23**(2):336–353, 2017.

[SWT17] Alexander Simons, Markus Weinmann, Matthias Tietz, and Jan vom Brocke. "Which reward should I choose? Preliminary evidence for the middle-option bias in reward-based crowdfunding." In *Proceedings of the 50th Hawaii International Conference on System Sciences*, pp. 4344–4353. HICSS, 2017.

[Tir86] Jean Tirole. "Hierarchies and bureaucracies: On the role of collusion in organizations." *The Journal of Law, Economics, and Organization*, **2**(2):181–214, 1986.

[Ver99] Arvind Verma. "Cultural roots of police corruption in India." *Policing: An International Journal of Police Strategies & Management*, **22**(3):264–279, 1999.

[Wik] Wikipedia. "Cheapass games." `https://en.wikipedia.org/wiki/Cheapass_Games`. Accessed: Sepetember 2019.

[Zak19] Nikita Zakharov. "Does corruption hinder investment? Evidence from Russian regions." *European Journal of Political Economy*, **56**:39–61, 2019.