Bad apples in a barrel. Field evidence about corruption in organizations

Romain Ferrali

January 10, 2019

Abstract

Organizations matter for corruption because they structure social relationships that may either deter corruption through increased monitoring or foster it by giving access to additional accomplices. I introduce a model that shows that which of these two functions prevails has important implications for organizational design, for it determines whether maximally or minimally connected organizations are optimal. Answering this question requires field measurements of social relationships and fraudulent behavior of individual members. I use a fine-grained, original dataset on the daily operations of a large company that allows measuring, over time, dishonest behavior at the individual level and interactions among employees. I identify agents that are likely to be dishonest and show that dishonesty has important costs for the organization. A series of natural experiments show that dishonest contaminate their honest peers, and reinforce their own behavior. This suggests it would be sensible to redesign government agencies to increase isolation among bureaucrats.

---

*Social Science Division, New York University Abu Dhabi, rferrali@nyu.edu* I thank Matias Iaryczower, Kosuke Imai, Jenn Larson, John Londregan, Kinga Makovi, Rebecca Morton, Malte Reichelt, as well as seminar participants at the Princeton Political Economy seminar, the Imai Research Group, the WF Networks and IMEBESS workshops for their helpful comments. I gratefully acknowledge support from the Mamdouha S. Bobst Center for Peace and Justice.
Corruption, “the abuse of entrusted power for private gain” (Transparency International) occurs within organizations: corrupt bureaucrats and politicians are all affiliated to government agencies, or other political organizations. Yet, we know very little about how organizations affect corruption. A long line of theoretical research has highlighted that organizations endow their members with colleagues, who are a mixed blessing. An honest individual can monitor her colleagues, but may also turn into their accomplices (Tirole, 1986; Laffont, 1990; Laffont and Tirole, 1991; McAfee and McMillan, 1995; Melumad, Mookherjee and Reichelstein, 1995; Gambetta, 1996; Ting, 2008; Vannucci and Della Porta, 2013; Ferrali, 2018). The specific layout of these relationships matters, because it tells us which agents are more at risk by their position within the organization. As a consequence, these relationships also determine which organizations are structurally better able to mitigate corruption. Pinning down these dynamics has important consequences for our understanding of bureaucracies, for they may explain why corruption sometimes involves isolated individuals, and sometimes conspiracies of many accomplices. Similarly, they may explain why corruption varies within a bureaucracy, where institutional features are held constant, but organizational layout varies widely across agencies. In particular, they may give us a better grasp at the impact of policies that are frequently observed in practice, where bureaucracies combat corruption through major reforms of their organizational chart (e.g. Bennet, 2012; Friedman, 2012; Hausman, 2011).

Yet, evaluating how interpersonal relationships affect corruption empirically has proven difficult, for this requires measuring both individual-level corruption and social interactions within an organization. Available evidence often comes from the lab, which typically evaluates one side of the relationship; that is, either how honest agents may deter dishonest ones (Schickora, 2011; Berninghaus et al., 2013), or how dishonest agents may corrupt honest ones (Gino and Pierce, 2009; Gino, Ayal and Ariely, 2009). Available field studies are incomplete, in that they either consider one organization and focus on a few known corrupt agents but lack systematic evidence on the rest of their peers (Aven, 2015; Gambetta, 1996; Vannucci and Della Porta, 2013), or consider many organizations, but have little information about the structure of relationships within them, and are therefore only able to quantify the effect of organization-level features on corrupt behavior (Pierce and Snyder, 2008, 2015; Khanna, Kim and Liu, 2015).

This paper answers the following question: within an organization, how much do “good apples” (honest agents) deter “bad apples” (dishonest agents) and how much do bad apples corrupt good apples? To do so, it exploits a unique dataset on the daily operations of a call-center company based in Morocco where clerks allocate callers’ claims to external service providers who generate revenue by serving callers. The allocation process leaves some discretion to clerks and opens up opportunities for fraud: instead of allocating claims to the best service provider, clerks may allocate them to their chosen provider, in return for a kickback. The data records more than a year of clerks’ interactions with the internal company software used to allocate claims to service providers. This allows the detection of suspicious behavior during claim allocation, revealing suspicious deviations from company policy and potential instances of preferential allocation of claims to specific service providers. The data also measures rich networks of interactions among clerks, enabling an investigation of how interpersonal relationships affect corrupt behavior. Finally, management style provides an advantageous natural
experiment that allows coupling these two measures to derive causal estimates of influence: interactions among colleagues are determined by management, who assigns clerks various tasks independently of their type, addressing potential concerns over homophily or contextual factors biasing the results \cite{Manski, 1993; Fowler et al., 2011; VanderWeele and An, 2013}.

The approach uses a large private organization as a proxy for a government agency. Analogies are a common approach in studies of the bureaucracy. The various organizational units that make up the bureaucracy are very heterogeneous, which makes comparing across them difficult. As a result, since the proverbial forest ranger \cite{Kaufman, 1960}, researchers often rely on single-agency case studies to extend conclusions to an entire bureaucracy \cite[e.g.][]{Carpenter, Gordon, 2001; Haeder and Yackee, 2015}. By considering a single organization, this paper places itself within this tradition. However, because obtaining this kind of rich data for a bureaucracy proved unfeasible, I consider instead a private organization whose features readily map to bureaucracies. Broadly speaking, this organization is similar to a bureaucracy in that it is a large organization of relatively educated, white-collar workers. Equally as important, states also operate call-centers that can fall prey to corruption. Police dispatch operators have been caught dissimulating critical information, leaking it to accomplices, or dispatching fake radio calls to cover up for illegal operations.

I first extend Jackson and Rogers' \cite{2007} model of diffusion on a network to show that different patterns of influence may call for dramatically different forms of organizational design. To give a trivial example, if good apples deter bad apples, but bad apples have no influence on good ones, then the optimal organization maximizes contact between employees. In the reverse case, the optimal organization isolates employees, to prevent bad apples from infecting good ones. The model generalizes this reasoning to effects of different magnitudes. Theoretically, the influence that one type exerts on the other may go two ways, triggering either contagion or differentiation from the other type \cite{Gino, Ayal and Ariely, 2009}. In other words, a good apple may either be contaminated by surrounding bad apples and turn into a bad apple itself, or differentiate from them by reinforcing its good behavior. I show that when effects go in opposite directions, then there is a unique optimal organizational design; the empty network when good apples are contaminated and bad ones differentiate, and the complete network in the reverse scenario. When effects go in the same direction, then the optimal design depends on the current state of the organization: when both types contaminate each other, the complete network is optimal if there is sufficiently few bad apples, while the empty network is optimal if there are too many.

I then turn to the data and first construct a measure of fraud that reveals that a small number of clerks show statistical evidence for suspicious behavior, with large costs for the organization. Of course, it is difficult to simply observe an organization and find corruption. The setting combines features that allow detecting statistically likely instances of dishonest behavior among clerks. Clerks perform a standardized task where fraud is well-defined: the only kind of dishonest behavior that may occur in this environment is the undue allocation of claims to specific service providers, in return for a kickback. This behavior has observable implications: dishonest clerks should allocate large amounts of revenue to their chosen service provider, both in absolute terms and compared to their honest colleagues. Furthermore, claims
can be allocated through several procedures, some of which are suspicious in that they can be exploited by dishonest clerks to award claims to their favorite provider. Dishonest clerks should have more such suspicious transactions than their honest counterparts. I compare across clerks to identify statistically significant deviations in these directions, thereby identifying the clerks whose behavior is so different from their colleagues’ that it becomes highly suspect. Since the measure is only a proxy for dishonest behavior, it is important to validate it. I show that the measure conforms to expectations about fraudulent behavior: most clerks behave honestly with all the providers they interact with. The few clerks that behave dishonestly only do so with a handful of providers, and tend to engage in such behavior over extended periods of time. I finally use the measure to quantify the amount of misallocated revenue, and show that it amounts to about 5 percent of the company’s total revenue, which is about the revenue generated by its second largest market.

Finally, I couple this measure of fraudulent behavior with social network data on interactions within the workplace to examine patterns of influence between good and bad apples. I show that good apples are contaminated by their honest peers. Bad apples, on the other hand, tend to self-reinforce when surrounded by more bad apples. Theory suggests that in this setting, decreasing interpersonal interactions within the organization would reduce dishonest behavior. Showing evidence of peer effects using observational data is notoriously difficult. The main concern in this setting is that clerks exhibit homophily by type; in other words, that good apples tend to cluster together, and bad apples tend to cluster together. Management style provides a natural experiment, because it allocates clerks to their tasks but largely ignores their types, which implies that whether two clerks interact should be independent of their types. I check the validity of this identifying assumption two ways. First, I directly examine its validity by conducting a series of placebo tests. Second, I use a recently developed permutation test to assess potential concerns over bias more formally.

By estimating the relative magnitude of peer effects among good and bad apples in organizations, this paper makes two contributions. First, it contributes to a literature that studies the behavioral foundations of corruption in organizations (Gambetta 1996; Vannucci and Della Porta 2013; Gino and Pierce 2009; Aven 2015). It considers a setting that allows examining relationships between honest and dishonest agents in a setting that combines the ecological validity of a field study with the advantages in measurement and causal inference that are typically afforded by the lab, allowing to credibly compare the influence of good apples over bad ones and that of bad apples over good ones. Second, this paper gives micro-foundations to a literature in political science that has examined the impact of organizational structure on corruption from a macro perspective (Evans 1995; Rauch and Evans 2000; Charron et al. 2017; Carpenter and Moss 2013). Although this literature points at features of the organization—such as meritocracy—that should reduce corruption, it is unable to make specific claims as to which organizational structures should reduce corruption because it lacks micro-foundations. This paper quantifies the interpersonal dynamics of influence that make particular organizational structures more or less prone to corruption and derives from the estimates the resulting optimal organizational structure.

The remainder of this paper proceeds as follows: I first introduce a theoretical model of
diffusion on a network that shows how different patterns of influence may prompt for dramatically different organizations (section 1). I then provide additional details about the data and the context, highlighting how features of this particular case map onto actual bureaucracies (section 2). I then derive and validate a measure of fraudulent behavior (section 3). Finally, I quantify peer effects between good and bad apples, and conduct a variety of robustness checks (section 4). I conclude by discussing the empirical results in light of the theory and outlining avenues for future research (section 5).

1 Theory

1.1 Literature review

Corruption is a major issue in developing and developed countries alike: bribery alone is estimated to cost about 2 percent of global GDP (International Monetary Fund 2016). Corruption hinders economic growth (Mauro 1995) and undermines the legitimacy and capacity of government (Rothstein 2011; Rose-Ackerman and Palifka 2016). Tackling the issue of corruption has generated a lot of research, with a heavy emphasis on norms and institutions. As a result, we have a good understanding of how various policy instruments such as wages or monitoring technologies affect the extent of corruption (e.g. Besley and McLaren 1993; Banerjee, Hanna and Mullainathan 2013).

Yet, while corruption occurs in organizations, we have little understanding of how organizations affect corruption. Corrupt bureaucrats and politicians are embedded in organizations: bureaucratic agencies and political bodies (Granovetter 1985; Zukin and DiMaggio 1990). Despite an early interest in how the organizational structure of the bureaucracy affects political outcomes (Weber 1948; Crozier 1964), we have little to say about simple questions on the relationship between organizational structure and corruption: which bureaucrats are, by their position, most likely to engage in corruption? When do corrupt bureaucrats “infect” their honest colleagues, and when do they get deterred by them? Which organizational structures best mitigate corruption?

In political science, extant work on organizations and corruption has generated valuable insights on the impact of macro-level organizational features on corruption. Evans (1995) considers ties between the agency and the rest of society. Rauch and Evans (2000) and Charron et al. (2017) analyze the impact of an array of organizational features that define the “Weberian-ness” of a bureaucracy, such as meritocratic recruitment procedures. Carpenter and Moss (2013) examine how the organization’s relationship with other actors, such as politicians and the industry may lead to agency capture.

Yet, this approach is largely silent on its micro-foundations, making it difficult to determine which organizational structures may best mitigate corruption. In most of these accounts, it is unclear (1) how corrupt agents interact with each other, and (2) how they interact with honest agents. As a result, it is unclear which specific organizational structures underpin the macro-level features highlighted by this literature.

Other work in economics and psychology have considered the micro-foundations of corruption in organization, usually theorizing interpersonal relationships as a double-edged sword that
may either facilitate or hinder corruption. Various mechanisms have been put forward. Peers may affect the probability of detection or the size of the reward, by acting as additional monitors or as accomplices. Peers’ behavior may also define a moral standard that becomes more compelling as one identifies more strongly with the group. 

Unfortunately, existing work on the micro-foundations of corruption in organizations does not have much to say about how these dyadic patterns of interaction play out in larger organizational structures. Some work in economics and political science has tackled this question. Yet, these models usually rely on a principal-agent approach, which makes them well-suited to study how corruption occurs in environments that the principal finds optimal, but is unable to speak of how corruption occurs in any organization that we might encounter. Furthermore, those models often consider very stark organizational structures, either perfect hierarchies or perfectly flat organizations, making it hard to extend their conclusions to more realistic organizations.

As a consequence, the next subsection develops a model that (1) encompasses any form of interactions between good and bad apples, (2) may apply to any kind of organization we might encounter, and (3) can be calibrated empirically using data from the firm. I model the organization as a network where nodes are bureaucrats, and ties represent relationships of professional collaboration. On this network, individuals may either be good or bad apples. At each time period, they switch type with some probability. That probability is governed by the types of one’s neighbors. I analyze the equilibria of this system, and how they are affected by network structure. The model extends Jackson and Rogers (2007) by considering more general functional forms for the transition probabilities. Like Jackson and Rogers (2007) and several other models of this class, I use a mean-field approximation to make the problem tractable.

1.2 Model

1.2.1 Setting

Consider a population of $n$ agents connected by the undirected graph $g = (G, N)$ where $N$ is a set of nodes and $G$ a set of ties. This network represents an organization where nodes are bureaucrats and ties represent relationships of professional collaboration. Let $N_i \equiv \{ j : ij \in G \}$ be the set of $i$’s neighbors and $d_i \equiv |N_i|$ be $i$’s degree; that is, the number of bureaucrats she collaborates with.

At each time period $t \geq 0$, agent $i$ can be honest or dishonest. We write $y_{it} = 0$ if $i$ is honest at time $t$ and $y_{it} = 1$ otherwise. Types are randomly drawn, with the following transition

---

1Some work in political science has also considered the micro-foundations of corruption in organization, but focuses largely on how bad apples cooperate. Because the focus is on interactions between good and bad apples, I omit this body of work from the discussion.
probabilities:
\[
\Pr(y_{i,t+1} = 0|y_{i,t} = 1, y_{N_i,t}) = p(\theta_{it}, d_i) = \alpha_p + \beta_p(1 - \theta_{it})d_i \tag{1}
\]
\[
\Pr(y_{i,t+1} = 1|y_{i,t} = 0, y_{N_i,t}) = q(\theta_{it}, d_i) = \alpha_q + \beta_q \theta_{it}d_i, \tag{2}
\]
with \(\theta_{it} \equiv \sum_{j \in N_i} y_{jt}d_i\) agent \(i\)'s infection rate at time \(t\); that is, the share of dishonest agents among \(i\)'s neighbors at time \(t\). In this specification, transition probabilities depend on the number of neighbors from the other type, \(\theta d\) and \((1 - \theta)d\) respectively. The parameters \(\alpha \geq 0\) capture an intrinsic propensity to switch types, while the parameters \(\beta\) capture how much switching types depends upon one’s peers. If \(\beta = 0\), then one’s type does not depend on her neighbors. Cases where \(\beta > 0\) capture a notion of influence: having more neighbors from the opposite type fosters switching type. Conversely, cases where \(\beta < 0\) capture a notion of differentiation: more neighbors from the opposite type reinforces adherence to one’s type. The linear specification follows [Jackson (2008)], but extends it to cases where \(\beta < 0\). It has the additional advantage of directly mapping onto the empirical specification that estimates an average peer effect (equation 5).

Since the graph \(g\) represents an organization, I assume that it is connected. I also assume that \(g\) is a random graph generated according to the configuration model ([Jackson and Rogers, 2007]). This model generates a graph according to a fixed degree sequence and has the property that the degrees of neighbors in the network are independent of one another. The sequence \(d_1, ..., d_n\) is the network’s degree sequence. We denote by \(P(d) \equiv \sum \frac{1}{n}d_i = d\) the degree distribution of the network, and \(\bar{d}\) the average degree. A generative process for this network could be as follows: suppose that each node has \(d_i\) “stubs;” that is, half-ties emanating from the node that are not connected to any other node. If there are \(m\) ties, then there are \(2m\) stubs. To form the network, pick two stubs from two different nodes at random, and connect them. Repeat the process with the remaining \(2m - 2\) stubs. Continue until no stubs are left. The resulting network has the desired degree sequence. An important feature of the configuration model is that it is easy to derive the excess degree distribution; that is, the degree distribution of a randomly chosen neighbor of a randomly chosen node. [Jackson and Rogers (2007)] show that the excess degree distribution writes

\[
\tilde{P}(d) \equiv \frac{P(d)d}{d}
\]

### 1.2.2 Steady states under the mean field approximation

To study the properties of this model, we make a standard simplification, and consider a degree-based mean field approximation. In other words, we solve the model assuming that all agents have the same infection rate at time \(t\) equal to mean infection rate in society: \(\theta_{it} = \theta \equiv \mathbb{E}[\theta_{it}]\) for any \(i \in N\). Let \(\rho(d) \equiv \Pr(y_{it} = 1|d_i = d)\) be the mean infection rate among agents of degree \(d\). Note that \(\rho\), the mean infection rate in society is \(\rho = \sum_d P(d)\rho(d)\) and that

\[
\theta = \sum_d \tilde{P}(d)\rho(d) \tag{3}
\]
Using the mean field approximation allows analyzing steady states of this dynamic process; that is, points at which the infection rate remains constant. At each time period $t$, a fraction $q(\theta, d)$ of non-infected agents of degree $d$ becomes infected, and a fraction $p(\theta, d)$ of infected agents of degree $d$ recovers. This pins down the law of motion of $\rho(d)$:

$$\frac{\partial \rho(d)}{\partial t} = [1 - \rho(d)]q(\theta, d) - \rho(d)p(\theta, d)$$

At the steady state, $\rho(d)$ remains constant. In other words, its law of motion satisfies

$$\frac{\partial \rho(d)}{\partial t} = 0 \iff \rho(d) = \frac{q(\theta, d)}{p(\theta, d) + q(\theta, d)} \quad (4)$$

Substituting this expression for $\rho(d)$ at the steady state into 3 gives $\theta = \sum_d \tilde{P}(d)\frac{q(\theta, d)}{p(\theta, d) + q(\theta, d)}$. Define $H(\theta) = \sum_d \tilde{P}(d)\frac{q(\theta, d)}{p(\theta, d) + q(\theta, d)}$. At the steady state, it must be that $H(\theta) = \theta$. In other words, steady states correspond to fixed points of $H$. Let $\mathcal{F} \equiv \{ \theta : H(\theta) = \theta \}$ be the set of such fixed points. Since $H : [0, 1] \to [0, 1]$ is continuous, $\mathcal{F}$ is non-empty. Define $\bar{\theta} \equiv \max \theta \in \mathcal{F}$ and $\underline{\theta} \equiv \min \theta \in \mathcal{F}$ be the steady states that generate most and least infection, respectively.

Generically, there may be many steady states. However, some specific parameter values give us more traction. For instance, if there is no natural propensity to turn dishonest, then a situation where everyone is honest is stable. In other words, if $\alpha_q = 0$, then there is a steady state at 0. The following proposition generalizes the intuition:

**Proposition 1** (Corner solutions). $\theta_g = 0 \iff \alpha_q = 0$ and $\theta_g = 1 \iff \alpha_p = 0$.

Because situations in which all members of society have the same type seem unrealistic, I focus on cases where $\alpha_p, \alpha_q \neq 0$. Additionally, note that when $\beta_p$ and $\beta_q$ have the same sign, our linear setup implies that $H$ is monotonic, and its derivative is either concave or convex. In this case, there is a unique steady state. Formally:

**Proposition 2** (Uniqueness). If $\alpha_p, \alpha_q \neq 0$ and $\beta_p$ and $\beta_q$ have the same sign, then $H_g(\theta)$ has a unique fixed point $\theta^* = \theta = \bar{\theta}$.

### 1.2.3 Comparative statics

Having a better sense of the properties of steady states, we can now analyze how they change across organizations. Let $o \equiv (g, \mu)$ be an organization, with $g = (G, N)$ its associated network, and $\mu = \{\alpha_p, \alpha_q, \beta_p, \beta_q\}$ a vector with its associated parameters. We compare organizations $o$ and $o'$ that differ either with respect to their graph $g$ or their parameter values $\mu$. In what follows, notations use the $'$ symbol to refer to $o'$.

In order to make such comparisons, we must be able to compare across organizations that may each have multiple steady states. We say that an organization is more honest than another if both its maximal and minimal steady states sustain less dishonesty than the other. Formally:

**Definition 1** (Honest organizations). We say that organization $o'$ is more honest than organization $o$ and write $o' \succeq o$ when $\theta' \leq \theta$ and $\theta' \leq \theta$. 

8
A simple lemma gives a lot of traction. As illustrated by figure 1 if the \( H \) function of organization \( o' \) is above that of organization \( o \), then \( o' \) is less honest than \( o \). Formally:

**Lemma 1** (Comparing organizations). If \( H_{o'}(\theta) \geq H_o(\theta) \) for any \( \theta \in [0, 1] \), then \( o' \preceq o \).

Together, definition and lemma allow comparing organizations that differ only by one of their parameters \( \phi \in \mu \). Results are immediate: organizations in which honesty spreads better – either through higher intrinsic propensity to become honest \( \alpha_p \) or better contagion \( \beta_p \) – are more honest. Conversely, organizations in which dishonesty spreads better are less honest. The following proposition articulates the reasoning.

**Proposition 3.** Suppose organizations \( o \) and \( o' \) only differ by the value of one of their parameters \( \phi \in \mu \). We have:

\[
\begin{align*}
\alpha'_p &\geq \alpha_p & \iff & \ o' \geq o \\
\beta'_p &\geq \beta_p & \iff & \ o' \geq o \\
\alpha'_q &\geq \alpha_q & \iff & \ o' \preceq o \\
\beta'_q &\geq \beta_q & \iff & \ o' \preceq o
\end{align*}
\]

Finally, let’s compare organizations that have the same parameter values \( \mu \), but differ in their network \( g \). Two networks can differ in many ways. I consider one specific kind of variation, namely a shift in the degree distribution such that the degree distribution on \( o' \) first-order stochastically dominates (FOSD) the degree distribution on \( o \), implying that the mean degree on \( o' \) is higher than on \( o \).

Instances where both patterns of influence go in opposite directions are relatively straightforward. If bad apples are influenced by good apples (\( \beta_p \geq 0 \)) and good apples differentiate from bad apples (\( \beta_q < 0 \)), then increasing the mean degree helps, for additional ties make either effect stronger. Conversely, if good apples are influenced by bad apples (\( \beta_q \geq 0 \)) and bad apples differentiate from good apples (\( \beta_p < 0 \)), then increasing the mean degree hurts. Formally:

\[2\text{That is, } o' \text{ has a degree distribution } P'(d) \text{ such that } \sum_d P'(d)u(d) \geq \sum_d P(d)u(d) \text{ for any non-decreasing function } u.\]
**Proposition 4** (Increasing mean degree (1)). Consider organizations $o$ and $o'$ with degree distributions $P$ and $P'$ respectively such that $P' \text{ FOSD } P$. If $\beta_p < 0$ and $\beta_q \geq 0$, then $o' \preceq o$. If $\beta_p \geq 0$ and $\beta_q < 0$, then $o' \succeq o$.

Instances where both patterns of influence go in the same direction are less straightforward. In this case, since $\beta_p$ and $\beta_q$ have the same sign, they may cancel each other out. Note that in this case, proposition [2] tells us that there is a unique steady state. Whether increasing mean degree helps or hurts depends on the position of the initial steady state. Suppose that good apples influence bad apples and vice versa; that is, suppose that $\beta_p, \beta_q \geq 0$. If there are sufficiently many honest types in the original steady state, then additional ties will connect few bad apples to many good apples, turning the former into good apples. Conversely, if there are too few honest types, then additional ties will connect many bad apples to few good apples, turning the latter into bad apples. Formally:

**Proposition 5** (Increasing mean degree (2)). If $\beta_p \geq 0$ and $\beta_q \geq 0$, or $\beta_p < 0$ and $\beta_q < 0$, then there is a threshold $\hat{\theta} \in (0, 1)$ such that

\[
\begin{align*}
\text{if } \beta_p, \beta_q \geq 0 \text{ and } \theta^* < \hat{\theta} & \quad o' \succeq o, \\
\text{if } \beta_p, \beta_q < 0 \text{ and } \theta^* \geq \hat{\theta} & \quad o' \preceq o, \\
\text{if } \beta_p, \beta_q \geq 0 \text{ and } \theta^* \geq \hat{\theta} & \quad o' \preceq o, \\
\text{if } \beta_p, \beta_q < 0 \text{ and } \theta^* < \hat{\theta} & \quad o' \succeq o,
\end{align*}
\]

Together, propositions [4] and [5] show that the optimal organization differs widely depending on parameter values (see figure 2 for a graphical summary). A maximally connected organization is optimal if (1) $\beta_p \geq 0$ and $\beta_q < 0$, or (2) $\beta_p \geq 0$, $\beta_q \geq 0$, and there is sufficiently few bad apples at the (unique) steady state, or (3) $\beta_p < 0$, $\beta_q \geq 0$, and there is many bad apples at the steady state.

Figure 2: **Illustration of propositions [4] and [5]**. Consider $o$ and $o'$ such that $P'(d)$ FOSD $P(d)$. $+$ signs denote $o' \geq o$, $-$ signs denote $o' \preceq o$. Increasing mean degree may help or hurt, depending on the values of $\beta_p, \beta_q$ and the initial steady state $\theta^*$. 


Table 1: **Sample descriptive statistics.** Mean chain length is the average number of draws required to award claims that were not forced. Clerks’ salary amounts to 1.6 times private sector minimum wage. The subset column considers markets pertaining to the top 20% in terms of revenue, in which the firm rolled out a new system for claim allocation, and where at least 2 firms operate on any month. Markets from the subset account for 42 percent of the company’s total revenue over the period.

### Context and data

The data describes the daily operations of a call-center company based in Casablanca, Morocco, between Novembers 1st, 2016 and August 31st, 2018. This call-center operates 24-7 to satisfy subscribing customers. When calling, customers file claims to call-center clerks employed by the company, who dispatch a partnering service provider to satisfy the customer’s demand. The data is a backup from the company’s internal software. It contains all 704,800 claims awarded during the period, alongside with characteristics of such claims (see Table 1 for descriptive statistics). These claims are typically urgent, and need to be serviced within the shortest amount of time – 5 minutes on average for urgent services and 30 minutes on average for non-urgent services. To service these claims, the company maintains a network of 971 partnering service providers,
whom the company compensates for providing those services using revenue from customers’ subscriptions. In the sample, average monthly compensation is $1,485. A key concern for the company is to allocate revenue fairly among service providers, allocating more claims to better-performing providers, but maintaining some level of activity for other providers, in order to make sure that some provider will be available at any time to satisfy potential claims.

In this environment, markets are a natural unit of analysis. Markets are defined as a type of services in a particular city. As we will see, they structure clerks’ interactions with the company’s software, and define the relevant set of social interactions for dishonest behavior. Yet, most markets are small: the top 20% markets in terms of revenue account for 82% of the company’s total transfers to service providers. The analysis focuses on a subset of these markets, where the company rolled out a new system to allocate claims to providers (see below), and at least two service providers operate on any given month. Indeed, only those have enough transactions and other prerequisites to allow for statistical analysis of dishonest behavior (see next section). The rightmost column of Table 1 reports descriptive statistics for this subset.

On any given day, clerks operate, on average, on 5 of these markets.

This context has the advantage that corruption is well-defined. Clerks can behave dishonestly by transferring additional revenue to specific service providers in return for a kickback. Such collusion requires either creating fictitious claims, or misallocating them – in other words, deviating from the company’s allocation policy and allocating more claims to their chosen service provider instead of following company policy. Creating fictitious claims is virtually impossible, for management monitors all claims that involve financial transactions. As such, the only kind of corrupt behavior is claim misallocation. Some of such schemes have been detected by company management in the past, with the last prominent incident dating back to 2012: a couple was colluding with several providers, who got caught after they approached an honest provider. The provider signaled the incident to management, who then used testimonies from other clerks for corroboration.

The incident highlights a weakness in managerial processes: manually detecting instances of collusion between a clerk and a service provider is a daunting task. The call-center processes an average of 1,017 claims daily; monitoring their allocation would require listening to hours of recordings of phone conversations. Such lack of oversight leaves room for fraudulent behavior.

To address the problem, the company updated its software in November 2016 to feature a system of random draws in order to limit clerks’ discretion during claim allocation in selected markets. To allocate a claim, the new system has clerks select a market. The software randomly draws a provider within the market; that is, a provider able to service claims of this type in this particular city. The draw is weighted, with weights defined by management so as to balance revenue allocation according to their needs. Clerks must call the selected provider and check whether it is available. If it is, then they should allocate the claim to that provider. If not, they must input a reason, and may skip – ask for another draw. The process ends when they reach an available provider. Reasons for unavailability include inability to reach the provider, the provider already operating at full capacity, or incapacity to service particular claims for technical reasons, such as lack of adequate equipment. Through this process, it takes an average

---

3 Evidence collected from interviews with management. Interview transcripts are available upon request.
Figure 3: **Observed distribution of chain lengths under the random draw system.** The distribution is skewed to the left: 94 and 91 percent of claims are allocated after at most two draws in the full sample and subset respectively.

Figure 4: **Schematic floor plan of the call-center.** Desks sit one person. The space has an area of 4,725 squared feet. Clerks operate in a single room. They are separated in two functionally and spatially distinct divisions.

of 1.4 draws to award a claim in the subset of markets considered for the analysis (Figure 3). Clerks may also *force* out of the rotation, i.e. terminate the random draw, and select a provider manually, which occurs in 30 percent claims from the subset. Clerks should force in three instances: when instructed to do so by management, for recurring claim where one customer always deals with the same provider, and when a customer jointly files two claims that could be serviced by the same provider. In this case, the provider should be selected through the random draw system for the first claim, and is forced for the second claim.

Although the random draw system does not completely eliminate discretion during provider selection, it makes collusion bear clear empirical implications. Because it is costly, management seldom verifies the reasons clerks invoke for skipping or forcing. Suppose that clerk *i* colludes with provider *k* in some market. Clerk *i* may then pretend that the providers drawn by the software were unavailable, and skip to *k*. Alternatively, *i* may pretend that she was following instructions from management, and force to *k*. As such, if *i* colludes with *k*, she should skip non-*k* providers more than honest clerks, and/or her rate of services forced to *k* should be higher than that of honest clerks in this market.

Peers matter because they may monitor or corrupt each other. Clerks all work in one
The call-center is divided into two functionally and spatially separate divisions, that handle different categories of services. Clerks are assigned to divisions, within which they typically handle several markets. Clerks operating on similar markets often sit close to one another in order to exchange information, typically about providers’ availability. As such, clerks that operate within the same market are able to verify their peers’ information. They overhear each others’ conversations and, when operating within the same markets, also hold private information about providers’ availability that they may leverage to verify whether their colleagues are telling the truth when claiming that some provider was unavailable, potentially reporting dishonest clerks to management. Conversely, dishonest clerks may also induce their honest colleagues to participate in their collusion. In this regard, the 2012 fraud incident is exemplary: the couple was operating on the same markets, and some of their peers provided management with evidence of their collusion.

The internal software allows reconstructing interactions between clerks. Clerks can only access the company software from the call-center floor. As such, if a clerk awards a claim at some point in time, she must have been in the room at that time. Similarly, if a clerk awards a claim in some market at some point in time, she must have been operating in that market at that time. This metric underestimates clerk attendance, since a clerk may be in the room/operating in some market but not allocate claims. However, because calls are allocated at random by dispatch, this measurement error should affect all clerks equally.

I use the data to construct a social network among clerks for each market-month. I focus on markets because only clerks operating within the same market matter for dishonest behavior. I consider months because they are a natural time unit within the call-center: clerks are typically hired at the beginning of the month, and leave the company at the end of the month, after getting paid their wages. Similarly, management revises sampling weights for the random draw system at the beginning of the month. I define two clerks to have spent one hour together on some market if they both allocated at least one claim within the same hour on that market. As a baseline, I define two clerks to be connected on some market-month if they spent at least one hour together, and vary that threshold as a robustness check.

Because this paper uses a private company to make claims about corruption in bureaucracies, it is important to assess the extent to which lessons learnt from this call-center may travel to public-sector organizations. Bureaucracies are diverse, and espouse Weber’s (1948) idealtype to varying degrees (Rauch and Evans 2000). The case has immediate counterparts in bureaucracies, which also operate call-centers, such as 911.

The case shares some features with Weberian bureaucracies: there is a strictly hierarchical organization between clerks and management with formal lines of authority, and a fixed area of operations: servicing customers’ claims by allocating them to service providers. Those tasks are, furthermore, executed continuously, since the call-center operates 24-7.

The organization displays, however, some important differences with the Weberian idealtype: call-center clerks have little career prospects and although not fully discretionary, hiring and firing decisions are much less regulated than in the public sector. Hiring decisions are based upon an evaluation of the candidate’s resume and an in-person interview instead of a formal exam. Firing decisions are regulated by law, but managers have significantly more discretion than in
the public sector. As such, while clerks share some characteristics with Weberian bureaucrats – namely that they are educated, white-collar workers with low-powered wage incentives\footnote{Virtually all clerks have completed some tertiary education. They are paid a flat rate with a yearly bonus that is capped at one month salary and awarded based on objective measures of performance derived from the company’s internal software.}, they also display some significant differences: clerks are all relatively young (28 years old on average), and turnover is high (47.6 percent in 2017).

While these differences distinguish the case from the Weberian idealtype, this does not prevent extending conclusions to real bureaucracies. First, bureaucracies conform less and less to the idealtype. In developed countries, the advent of New Public Management \cite{Hood1991} has resulted in more flexible labor management, with steady increases in turnover rates \cite{DahlstromHolmgren2017} for Sweden, and existing evidence from developing countries show that bureaucrats’ mobility across agencies is comparable to the case’s turnover rates \cite{IyerMani2012} for India. Finally, the fact that labor is less mobile in bureaucracies than in this organization suggests that the conclusions of this study are all the more relevant for the public sector. Indeed, while private sector management may exploit increased flexibility to curtail corruption by selecting on good apples ex-ante and firing bad apples ex-post, public sector management does not have this luxury. As such, aiming at reducing corruption through changes in the organizational structure while holding staff constant is all the more relevant in the public sector.

\section{Measuring fraud}

This section devises a novel, simple test to show statistical evidence of fraud in the allocation of claims. The test assumes that all clerks are honest and derives null distributions for two behaviors of interest: (1) whether the claim is awarded through suspicious means – that is, either after a disproportionately high amount of skips, or is forced; and (2) the distribution of revenue among service providers. I then compare each clerk’s observed behavior to that null distribution, and single out the clerk-provider dyads that are most unlikely to conform to the null distribution, and hence most likely to be dishonest. Finally, I validate this metric by verifying that it conforms to expectations about fraudulent behavior.

\subsection{Approach}

As discussed in the previous section, this setting has the advantage that dishonest behavior is well-defined and bears clear empirical implications. Dishonest behavior by clerks is the preferential allocation of revenue to selected service providers, in excess of what is determined by company policy, and in return for a kickback. If clerk $i$ colludes with provider $k$ and most clerks operating on that market are honest, then three conditions should be verified:

1. Clerk $i$ should allocate large sums of money to provider $k$.

2. Clerk $i$ should allocate disproportionately high shares of revenue to provider $k$, compared to other clerks operating on that market.
3. When allocating claims to provider $k$, clerk $i$ should be more likely than other clerks to use suspicious means.

These three conditions are necessary for dishonest behavior. If condition (1) is not met, then clerk $i$ would get too small a kickback for her dishonest behavior to be profitable. If condition (2) is not met, then clerk $i$ would not be allocating revenue to provider $k$ in excess of company policy. Regarding condition (3), recall that the random draw system used to govern the allocation of claims may be exploited to award claims fraudulently in two ways: a clerk may skip until her favorite provider gets drawn by the system, resulting in disproportionately long chains, or she may force to award the claim to her chosen provider. Claims awarded after long chains or through force are suspicious; in other words, they suggest dishonest behavior, but are not definite evidence, since the (unverifiable) conditions that justify such suspicious awards might have been met. However, claims not awarded through suspicious means are definitely not fraudulent.

I operationalize those three conditions as follows. For condition (1), I use a cutoff of $500. Since clerks have an average wage of $424 and are likely getting a small fraction of revenue allocated to the provider as a kickback, allocating less than $500 to some provider would yield to small a kickback for dishonest behavior to be profitable. I operationalize conditions (2) and (3) using null models. These null models assume that all clerks are honest. Under this assumption, all claims awarded on the market may be used to pin down honest patterns for claim allocation. For condition (2), let $\mathcal{K}$ be the set of providers operating on that market, $y_c = k \in \mathcal{K}$ be the provider to which claim $c$ was awarded, and $u_c$ a vector of features of claim $c$. I use the following multinomial logistic regression to estimate model $M_2$, the null model for claim allocation:

$$M_2: \Pr(y_c = k) = \frac{\exp(u_c^\prime \beta_k)}{\sum_{k' \in \mathcal{K}} \exp(u_c^\prime \beta_{k'})}$$

For condition (3), recall that more than 90% of claims are awarded after random draws are allocated in less than two draws. As such, I define $z_c = 1$ if claim $c$ was awarded suspiciously; that is, if claim $c$ was forced, or if it was awarded after 3 draws or more, and $v_c$ a vector of features of claim $c$. I use the following logistic regression to estimate $M_3$, the null model for suspicious allocation to provider $k$:

$$M_3: \Pr(z_c = 1) = \frac{\exp(v_c^\prime \gamma_k)}{1 + \exp(v_c^\prime \gamma_k)}$$

I then use these null models to compute how much the set of claims awarded by clerk $i$ to provider $k$ meet the three above conditions. Let $\mathcal{C}_{ik}$ the set of claims that $i$ allocated $k$, and $m_c$ the monetary value of claim $c$. As such, $r_{ik} = \sum_{c \in \mathcal{C}_{ik}} m_c$ is the revenue allocated by $i$ to $k$. Define $s^1_{ik} = 1\{r_{ik} > 500\}$, an indicator for dyad whether $ik$ meets condition (1). Define $s^2_{ik} = \Pr(r < r_{ik}|M_2)$ the probability of $i$ allocating less revenue to $k$ than $r_{ik}$ under null model $M_2$. Let $\mathcal{C}_i$ the set of claims allocated by clerk $i$ and $r_i = (r_{ik})_{k \in \mathcal{K}}$ the vector of revenues allocated by $i$ to all providers operating on the market. Under $M_2$, we have $r_i \sim \sum_{c \in \mathcal{C}_i} m_c \text{Multinom}(u_c, \beta)$. I simulate draws from the distribution of $r_i$ to compute $s^2_{ik}$ for each clerk-provider dyad. Finally, let $f_{ik} = \frac{\sum_{c \in \mathcal{C}_{ik}} z_c}{|\mathcal{C}_{ik}|}$ the fraction of suspicious claims that
i awarded to k, and define $s_{ik}^3 = \Pr(f < f_{ik}|M_3)$ the probability of the dyad $ik$ having a lower rate of suspicious claims. Under $M_3$, we have $f_{ik} \sim \sum_{c \in C_{ik}} \text{Bernoulli}(v_{c,ik})$. I simulate draws from the distribution of $f_{ik}$ to compute $s_{ik}^3$ for each clerk-provider dyad. I then use $s_{ik}^1$, $s_{ik}^2$, and $s_{ik}^3$ to derive $s_{ik}$, the extent to which dyad $ik$ jointly meets conditions (1) to (3). This quantity writes

$$s_{ik} = s_{ik}^1 s_{ik}^2 s_{ik}^3$$

The quantity $s_{ik} \in [0,1]$, that I call an “s-score”, measures how suspicious dyad $ik$ is, with higher values denoting more suspicious dyads. Specifically, it captures the extent to which dyad $ik$ meets conditions (1) to (3) under null models $M_2$ and $M_3$. The quantity relates to the familiar notion of a p-value. The components $s_{ik}^2$ and $s_{ik}^3$ are in fact inverse one-tailed p-values: instead of quantifying the probability of observing, under the null, a quantity more extreme than the observed quantity, those components quantify the probability of observing a quantity less extreme than the observed quantity, such that higher values of the component reflect that the observed quantity is more atypical under the null. Those components are one-tailed inverse p-values because conditions (2) and (3) only require upward deviations from the null.

Computing the s-score requires estimating models $M_2$ and $M_3$, which in turn requires selecting covariates for vectors $u$ and $v$. Given that months are a natural time-unit in this setting, I calculate monthly u-scores, and estimate model parameters monthly. This poses a tradeoff. On the one hand, one would like to include a lot of covariates, such that the null model accounts for many sources of variation across clerks. On the other hand, including too many covariates may make the model too complex and prevent its estimation. Indeed, even within the subset considered for analysis, some markets or some providers within those markets only register few monthly transactions, preventing the estimation of overly complex models. As a result, both $M_2$ and $M_3$ consider six time-bins that match patterns of activity used by management to allocate employees to work shifts. I divide each day in three time-intervals: morning (7:00a to 2:59p), afternoon (3:00p to 8:59p), and evening (9:00p to 6:59a), and divide the week into weekdays and weekends. Model $M_2$, which considers claim allocation, also includes log-claim value, for some providers are often selected to handle more complex, more valuable claims.

### 3.2 Results

Figure 5 reports the s-scores of those clerk-provider dyads where monthly revenue is greater than $500. Indeed, only those dyads satisfy the first condition for dishonest behavior; that is, have large enough revenue for fraud to be profitable for a clerk. The left panel shows that there is variation in both components $s^2$ and $s^3$ of the s-score – that is, the components that capture, respectively, the extent to which one clerk’s distribution of allocated revenue is overly concentrated in favor of one service provider, and whether those transactions have disproportionately many suspicious transactions, – with little correlation (.11) between them.

I use the s-score to estimate the proportion of revenue that was misallocated over the period. To do so, I use a threshold in s-scores to define dishonest clerks. Specifically, if clerk $i$ on market $m$ had a maximum s-score above .3, then I define $i$ to be dishonest on market $m$. If $i$ is dishonest

---

5 I use a threshold of .3 to be consistent with the threshold used in the main empirical results in section 4.2
Figure 5: A measure of fraud. Both panels report the $s$-score of those clerk-provider dyads where monthly revenue is greater than $500 (i.e. where component $s^1 = 1$). Simulations to derive components $s^2$ and $s^3$ use 1,000 draws. Left: there is variation in both components of the $s$-score. Right: most clerk-provider dyads have low $s$-scores.

3.3 Validation

Of course, the $s$-score comes with limitations, as it is only a statistical indication of suspicious behavior. First, it does not measure actual dishonest behavior but only suspicious behavior, since we cannot observe the kickbacks that providers transfer to clerks. Second, the metric is conservative: it understates dishonesty because it is based on the assumption that all clerks are honest. If one clerk is dishonest and makes up most of the transactions on the market, she
would have most weight in defining the null models of behavior. She would therefore wrongly appear to be honest. Honest clerks, on the other hand, would still appear as honest because (1) they would have lower rates of suspicious transactions, and (2) the fact that they have few transactions would introduce a lot of noise in estimating their deviations from the null model of claim allocation. Similarly, if all clerks are dishonest, they would all appear to be honest, since they define the null model. That the metric is overly conservative is, however, an advantage, as it increases our confidence that dyads with high \( s \)-scores actually capture true cases of dishonest behavior. Overall, these limitations make validating the metric all the more relevant. Since direct validation of the measure through actual investigations of dishonest behavior proved not feasible, I validate the metric by checking whether it conforms to common expectations about dishonest behavior. I conduct five such tests.

**Test 1** (Dishonest behavior is rare). Figure 5b shows that the distribution of \( s \)-scores among dyads that record more that $500 of monthly revenue is largely skewed to the left: the median \( s \)-score is .18, while the third quartile, ninth decile, and ninety-ninth percentile are, respectively, .43, .68, and .96.

![Cumulative distribution of dyads by clerk-market-month](image)

Figure 6: **Cumulative distribution of dyads by clerk-market-month.** This figure only considers the clerks that have, on a given market in a given month, at least one dyad with revenue greater than $500. Clerks have few dishonest partners: the overwhelming majority of clerks that have a dishonest partner only have one.

**Test 2** (Dishonest clerks only partner with a few dishonest service providers). The metric would fail to capture fraud if it showed that within a given market, a dishonest clerk partners with many dishonest service providers. Indeed, recall that claim misallocation redistribute revenue away from some providers to dishonest providers. If one partners with too many dishonest providers, there is too little honest providers left to redistribute revenue from. Figure 6 shows that within a market, clerks have very few dishonest partners. Ninety percent of clerks have at most one service provider with an \( s \)-score greater than .3, and 98 percent have at most one provider with an \( s \)-score greater than .6.
Bad apples in a barrel

Romain Ferrali

Figure 7: **Effect of experience on dishonest behavior.** The $x$-axis represents the average experience of a clerk over the period she has been observed, while the $y$-axis plots the maximum $s$-score of any clerk-provider dyad she has been involved in. The black line is a loess fit. It takes a modicum of experience to engage in dishonest behavior: clerks that have less than one year of experience all have low $s$-scores, while some of the clerks that have more than one year of experience have high $s$-scores.

**Test 3** (Dishonest behavior requires a modicum of experience). Engaging into dishonest behavior requires knowing the ropes of the trade and building relationships with those dishonest service providers. As such, having some experience is a necessary – but not sufficient – condition for dishonest behavior: while some experienced clerks may display signs of dishonest behavior, no unexperienced clerk should. Figure 7 confirms the pattern: while clerks that have less than one year of experience all have low $s$-scores, some of the clerks that have more than one year of experience have high $s$-scores.
<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>s-score</strong></td>
<td>-1.565**</td>
<td>-0.582</td>
</tr>
<tr>
<td></td>
<td>(0.626)</td>
<td>(0.642)</td>
</tr>
<tr>
<td><strong>log(revenue)</strong></td>
<td>-0.307***</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.062)</td>
</tr>
<tr>
<td><strong>experience</strong></td>
<td>-0.002</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.004)</td>
</tr>
<tr>
<td><strong>Observations</strong></td>
<td>678</td>
<td>678</td>
</tr>
<tr>
<td><strong>Log Likelihood</strong></td>
<td>-318.668</td>
<td>-307.409</td>
</tr>
<tr>
<td><strong>Wald Test</strong></td>
<td>5.710** (df = 1)</td>
<td>27.070*** (df = 3)</td>
</tr>
</tbody>
</table>

*Note:* *p<0.1; **p<0.05; ***p<0.01*

Table 2: **Survival models for the duration of clerks’ employment spells.** Time-periods are months, and all models use time-varying covariates, with s-score defined as a clerk’s maximum s-score on a given month, experience defined in months, and log revenue defined as the log of total revenue handled by a clerk on a given month. Standard errors are clustered at the clerk level. Without controls, dishonest clerks are less likely to get fired (model 1). Controlling for revenue and experience, the length of dishonest clerks’ employment spells is not significantly different from that of honest clerks (model 2).

**Test 4** (Dishonest clerks are no less likely to get fired). If the measure picks up dishonest behavior, then dishonest clerks should not be less likely to get fired than honest clerks. Specifically, if the company is generally aware of dishonest behavior, then dishonest clerks should be more likely to get fired than honest clerks. If the company is generally unaware of dishonest behavior, then they dishonest clerks should be just as likely to get fired as dishonest clerks. Table 2 estimates survival models for the duration of clerks’ employment spells at the company. It shows that without controls, the employment spells of dishonest clerks are longer than those of honest clerks. However, after controlling for revenue handled by clerks and their experience within the firm, the length of dishonest clerks’ employment spells is not significantly different from that of honest clerks.

**Test 5** (There is little correlation between dishonest behavior and provider quality). The measure may confound dishonesty and reckless behavior. Suppose that some clerk i knows that some provider k is better than other providers, and transfers many claims to k, without following company rules regarding claim allocation. This clerk would appear to have a large s-score, although she would not be doing anything dishonest.

To test for whether the metric picks up reckless behavior, I leverage measures of provider quality devised by the company. If the measure captures reckless behavior, then there should be a positive correlation between s-scores and provider quality. If the measure captures dishonest behavior then there should be a null to negative correlation between s-scores and provider quality.
I use two measures of provider quality. First, the company conducts yearly audits of its providers. This is an objective measure in that the company verifies whether each provider conforms to a series of pre-defined quality standards. Second, the company continuously conducts satisfaction surveys of a random subset of its customers over the phone. I use these surveys to construct a subjective, quarterly measure of provider quality by averaging customer satisfaction by quarter. Some customers mistakenly believe that they can get additional benefits by reporting that they are dissatisfied. As such, the company verifies the claims of each dissatisfied customer, in order to establish whether the complaint was proven or not. I only consider satisfied customers and dissatisfied customers whose dissatisfaction was proven.

<table>
<thead>
<tr>
<th>Dependent variable:</th>
<th>yearly ratings</th>
<th>quarterly ratings</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
</tr>
<tr>
<td></td>
<td>(3)</td>
<td>(4)</td>
</tr>
<tr>
<td></td>
<td>(5)</td>
<td>(6)</td>
</tr>
<tr>
<td>s-score</td>
<td>−0.003</td>
<td>−0.006***</td>
</tr>
<tr>
<td></td>
<td>(0.004)</td>
<td>(0.002)</td>
</tr>
<tr>
<td></td>
<td>−0.008***</td>
<td>(0.003)</td>
</tr>
<tr>
<td></td>
<td>0.005</td>
<td>(0.006)</td>
</tr>
<tr>
<td></td>
<td>0.009**</td>
<td>(0.004)</td>
</tr>
<tr>
<td></td>
<td>0.008**</td>
<td>(0.004)</td>
</tr>
<tr>
<td>Constant</td>
<td>0.757***</td>
<td>0.713***</td>
</tr>
<tr>
<td></td>
<td>(0.004)</td>
<td>(0.006)</td>
</tr>
<tr>
<td>Market FE</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Month FE</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Clerk FE</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Mean DV</td>
<td>0.757</td>
<td>0.715</td>
</tr>
<tr>
<td>SD DV</td>
<td>0.037</td>
<td>0.089</td>
</tr>
<tr>
<td>Observations</td>
<td>3,616</td>
<td>4,239</td>
</tr>
<tr>
<td>R²</td>
<td>0.0004</td>
<td>0.087</td>
</tr>
</tbody>
</table>

Note: *p<0.1; **p<0.05; ***p<0.01

Table 3: Correlation between s-scores and provider quality. These models regress the quality rating of a provider on the s-score of each dyad with revenue above $500 she is involved in. Models 1 to 3 use objective, yearly quality ratings, models 4 to 6 use subjective, quarterly quality ratings. Standard errors are clustered at the month and market levels. There is little correlation between provider quality and s-scores, with effect size of about one tenth of a standard deviation of the dependent variable.

Table 3 reports the correlation between the monthly s-score of a clerk-provider dyad whose revenue is higher than $500 and provider quality, normalized to fall between 0 and 1. Using yearly quality ratings (models 1 to 3) shows a null to negative correlation between provider quality and s-scores. Comparing within market and within month (model 2), low-quality providers are involved in more dishonest dyads. Comparing within clerk (model 3), clerks are more dishonest with low-quality providers. Moving to quarterly, subjective quality ratings (models 4 to 6) reveals the opposite pattern: there is a null (model 4) to positive (models 5 and 6) correlation between s-scores and provider quality. Note, however, that in all models, even though some providers that service few claims are, by virtue of random sampling, seldom surveyed. As such, I only consider providers that have been surveyed more than 5 times on a given quarter.
correlations are significant, effect sizes are minuscule: moving from perfectly honest to perfectly dishonest only increases provider quality by a tenth of a standard deviation. Therefore, there is little correlation overall between dishonesty and provider quality.

Figure 8: Correlation between $s$-scores and provider quality by clerk type. A clerk is defined to be dishonest (bad) if her maximum $s$-score on a given month is above some threshold, reported on the top row of the $x$-axis. For each threshold, I reestimate models (3) and (6) in Table 3 (left and right panel respectively), but add an interaction term between clerk type and $s$-score. Points report point estimates for good and bad types. Bars represent 90 and 95 percent confidence intervals clustered at the month and market levels. The $x$-axis also reports the number of good and bad types for each model (second and third rows). There is little correlation between provider quality and $s$-scores. That correlation is minuscule for bad types, and comparable between high and low types for thresholds above .5.

Figure 8 considers honest and dishonest clerks separately. Indeed, the measure would prove problematic if it showed a positive correlation between provider quality and $s$-scores for dishonest clerks. I define a clerk to be dishonest on a given month if her maximum $s$-score during that month is above some threshold. I then add to models 3 and 6 in Table 3 an interaction term between a clerk’s type and the $s$-score of a dyad she is involved in. The figure shows the same pattern as in the previous exercise: even when considering dishonest clerks only, there is little correlation between the $s$-score of the dyads they are involved in and the quality of the providers involved in those dyads. When types are defined using high enough thresholds (above .5), effect sizes are comparable between honest and dishonest types.
4 Peer effects among good and bad apples

While the previous section derived the s-score, a statistical measure of dishonest behavior, this section examines how such behavior is affected by social relationships in the workplace.

4.1 Approach

Section 1 showed that the organizational structure that best mitigates corruption varies widely depending on the direction of influence – namely, whether being exposed to the other type triggers reactions of contagion or differentiation. The aim is to establish whether social relationships in the workplace result into contagion or differentiation. In other words, the aim is to identify the parameters $\beta_p$ and $\beta_q$ from the theoretical model in section 1.

Identifying peer effects is a notoriously hard problem (Manski, 1993; Fowler et al., 2011). Unless one controls for all relevant confounders, the correlation between the outcome of agent $i$ and her neighbors’ outcome need not indicate influence; namely, that $i$’s neighbors caused $i$’s outcomes. If the analyst cannot control for all relevant confounders, two sources of bias have been widely discussed (VanderWeele and An, 2013): homophily, the fact that peers often share similar characteristics, which may drive the observed outcome; and contextual confounding, the fact that peers often evolve in a similar context that may be driving the outcome.

Features of the firm considered in this paper provide a natural experiment that alleviates concerns for bias affecting the results. First, such a small, self-contained environment is largely homogeneous, in the sense that contextual shocks may affect either all, or none of the clerks, suggesting that contextual confounding will probably have little impact. Second, management style alleviates concerns for homophily driving the results. Homophily may arise if types sort into markets; in other words, if good types tend to work with good types, while bad types tend to work with bad types. As is often the case in call-centers, management is quite authoritarian; they decide how to affect clerks to work shifts depending on their own scheduling priorities. During a work shift, they assign clerks to markets depending on their own business priorities. Management largely ignores clerks’ types – that is, whether a clerk is honest or dishonest. As such, the assumption that clerks are assigned to markets independently of their types seems plausible.

The estimation strategy allows identifying the peer effects of good types on bad types, and that of bad types on good ones, following the model introduced in section 1. Slightly adapting notation from that model, let $y_{imt} = 1$ if clerk $i$ in market $m$ during month $t$ is a bad type, and $y_{imt} = 0$ otherwise. I estimate the following linear probability model:

$$\text{switch}_{imt+1} = \alpha_m + \alpha_t + \beta y_{imt} + \gamma_0 n_{imt} + \gamma_1 n_{imt} y_{imt} + \delta x_{imt} + \epsilon_{imt},$$

(5)

where $\text{switch}_{imt+1} = 1\{y_{imt+1} \neq y_{imt}\}$ equals to 1 if $i$ changes types between months $t$ and $t+1$, and 0 otherwise, $\alpha_m$ and $\alpha_t$ are market and month fixed effects, and $n_{imt} = \sum_{j \in N_{imt}} 1\{y_{jmt} \neq y_{imt}\}$ the number of agents $j$ that belong to $N_{imt}$, $i$’s neighborhood on market $m$ during month $t$ whose type is different from $i$. Furthermore, $x_{imt}$ is a vector of control variables, and $\epsilon_{imt}$ is the error term.

The model in equation 5 allows estimating the parameters $\beta_p$ and $\beta_q$ in equations 1 and 2.
from the theoretical model. The empirical model regresses the probability that agent \(i\) changes type in period \(t+1\) on the number of neighbors from the opposite type she has in period \(t\). As such, the marginal effect of a bad type on a good type is \(\beta_q = \gamma_0\), while the marginal effect of good type on a bad type is \(\beta_p = \beta + \gamma_1\).

This model contains several features that allow for more credible estimation of peer effects. First, the market- and month-level fixed effects control for potential contextual effects occurring in specific months or markets. Since those effects may introduce heterogeneity, I cluster standard errors at the market and month level. The vector of controls \(x_{imt}'\) controls for as many observables as possible in order to account for homophily.\(^7\)

Since the model analyzed in section \(^1\) requires a discrete measure of fraud, I turn the continuous measure of dishonest behavior derived in the previous section into a binary indicator. I consider the maximum \(s\)-score of any clerk-provider dyad that includes clerk \(i\), and set \(y_{imt} = 1\) if it is above some threshold. With \(K_{mt}\) the set of providers operating on market \(m\) in month \(t\), and for a threshold \(\bar{s} \in [0,1]\), we have \(y_{imt} = 1\{\max_{k \in K_{mt}} s_{ikt} \geq \bar{s}\}\). During estimation, I consider a range of such thresholds. Finally, note that by definition of the \(s\)-score, if clerk \(i\) has no dyad with revenue higher than $500, then \(y_{imt} = 0\). As such, I estimate the model in equation \(^5\) only for those clerks who have at least one such dyad at time \(t+1\).

Although features of the firm provide suggestive evidence against bias, I conduct in section \(^4.3\) several validity checks to verify that there is little homophily. Specifically, I check whether clerks are indeed assigned to markets independently of their type, and conduct a permutation test to see whether observed patterns of tie formation suggest that homophily is at play.

### 4.2 Results

Figure \(^9\) reports the main result: for a high enough threshold in \(s\)-scores, honest types are contaminated by dishonest types, while dishonest types differentiate when exposed to honest types. For a threshold of \(.3\), we have \(\beta_q = .059\) and \(\beta_p = -.012\) in the specification with controls.\(^8\) In other words, one additional dishonest neighbor increases the chances that an honest agent turns dishonest by 5.9 percentage points, while one additional honest neighbor decreases the chances that a dishonest agent turns honest by 1.2 percentage points. Effect sizes are similar in the specifications with and without controls. For bad types, the magnitude of the effects increases as we adopt more stringent definitions for dishonest types. In fact, effects only show for \(s\)-scores above \(.2\). This suggests that only the most dishonest agents differentiate from honest types. Results for thresholds above \(.7\) become very unstable because there are too few clerks with \(s\)-scores sufficiently high for estimation to be precise enough. In the remainder of this paper, I define good and bad types using only that threshold of \(.3\).

I examine next how results vary as we adopt more stringent definitions for social ties. Recall from section \(^2\) that we defined a tie between clerks \(i\) and \(j\) on market \(m\) if they had spent at least one hour operating together on that market. I now change this definition, using thresholds ranging from 1 to 4 hours. Figure \(^10\) shows that results have a similar magnitude irrespective

\(^7\)That vector includes a series of variables measured at month \(t\). Specifically, it contains \(i\)'s degree on market \(m\), the log revenue processed by \(i\) on any market, the log revenue of market \(m\), the maximum \(s\)-score of \(i\) on any market, the number of hours worked by \(i\) on market \(m\), and on any market.

\(^8\)See Appendix \(B\) for a regression table.
Figure 9: **Average marginal effect of a neighbor of the opposite type on switching types.** The $x$-axis represents the threshold in $s$-score used for type definition to estimate the model in equation 5 as well as the number of good and bad types implied by such threshold. Points represent, for each model, the average marginal effect of an additional neighbor of the opposite type on switching types; bars are 90 and 95% confidence intervals clustered at the month and market levels. All models include month and market fixed effects. Panel (a) reports models without controls, panel (b) reports models with the controls discussed in footnote 7. For thresholds between .3 and .6, honest types are contaminated by dishonest types, while dishonest types differentiate from honest types.

of the definition used. For honest types, this occurs irrespective of whether the specification includes controls or not. For dishonest types, effects get stronger without controls, but are of comparable magnitude after introducing controls. Using, however, more stringent definitions for social ties reduces the number of recorded social ties which, in turn, increases the uncertainty associated with the estimates.

### 4.3 Validity checks

Having established that there is social influence within the firm, I now conduct several validity checks to make sure that the estimates are unbiased. In this setting, the main threat to identification is homophily: it might be the case that honest and dishonest clerks sort into specific markets and manage to only interact with clerks of their own type within a given market. Robustness checks all investigate this particular source of bias.

The first series of tests are substantive. They investigate whether there is evidence for sorting into specific markets. To do so, I first compare, for each month, the clerks that enter new markets to the ones who do not. In the absence of sorting, there should be no correlation
between a clerk’s type during month $t$, measured as her maximum $s$-score that month, and whether she enters a new market during month $t + 1$. I then consider only those clerks that move at time $t$ and look into the market(s) they join at time $t + 1$. In the absence of sorting, there should likewise be no correlation between that clerk’s type and the degree of honesty within the market she joins at $t + 1$, measured as the maximum $s$-score within that market at time $t$. Table 4 shows results that confirm the expectation. All models include month fixed-effects to compare within month. Conditional on clerk characteristics that determine moving, there is no correlation between a clerk’s type and entering a new market (model 2). Conditional on moving, there is no significant correlation between a clerk’s type and the degree of honesty in the market(s) she joins in the specification without controls, and in the specification that controls for clerk and market characteristics (models 3 and 5).

As a final validity check, I use a permutation test for homophily adapted from LaFond and Neville (2010). Their intuition is simple: if there is homophily, then over time, ties should form between individuals sharing similar values for some attribute – in our case, whether clerk $i$ is honest. Their permutation procedure decorrelates one’s adoption pattern from that of her
Clerks entering new markets. Clerk controls are measured at month $t$ and include clerk experience in months, log revenue, and number of hours worked that month. Market controls are also measured at month $t$ and include log market revenue, and the number of clerks and of firms operating on that market. All models cluster standard errors at the month and employee level. Conditional on clerk characteristics that determine moving, there is no correlation between clerk type and entering new markets (model 2). Conditional on moving, there is no correlation between clerk type and the degree of honesty in the destination market (models 3 and 5).

neighbors. Considering periods $t$ and $t+1$, they hold constant the number of individuals that keep and drop the attribute between $t$ and $t+1$, but permute their identity. In other words, they randomize, among the agents that had the attribute at $t$, the ones who drop it and randomize, among the agents that did not have the attribute at $t$, the ones who adopt it. The authors consider only one graph over multiple periods. As such, their test statistic is the variation between $t$ and $t+1$ of the $\chi^2$ statistic of the association table between whether the dyad $i$ and $j$ has a tie, and whether that dyad has the same attribute value. In order to use that test jointly for several networks, I use a parametric specification:

$$g_{ijmt} = \alpha_t + \alpha_m + \beta_0t + \beta_1y_{ijt} + \gamma ty_{ijt} + \epsilon_{ijmt},$$

with $g_{ijmt} = 1$ if there is a tie between clerks $i$ and $j$ in market $m$ during period $t$, $y_{ijt}$ a binary variable that equals 1 if $i$ and $j$ have the same type at period $t$, and $t \in \{0, 1\}$ a binary variable that equals 0 in the first period, and 1 the next period. The model compares within months and within markets. As such, I include month fixed-effects $\alpha_t$, and market fixed-effects $\alpha_m$. The statistic of interest in this test is the parameter $\gamma$ that captures whether the effect of having the same attribute value increases between the first and the second period. Note that there needs to be enough individuals capable of switching types for there to be enough permutations. The main analysis only considered those clerks that had at least one dyad with revenue greater than $500 at t + 1$. I now consider only those market-months that had 5 such clerks or more at $t + 1$, and consider only those clerks within the market.

Figure 11 reports the results, showing that there is little evidence for homophily. Having similar attribute values increase the chance of tie creation by about 0 percentage points, which is not significantly different from what is to be expected from a process where adoption decision
Figure 11: **Permutation test for homophily.** The figure reports the density of test statistics for a permutation test for homophily adapted from [LaFond and Neville (2010)](https://www.jstor.org/stable/10.1093/academ傲journals/kxy040), using 1,000 permutations. The black bar represents the observed test statistic, the dotted bars represent the 2.5 and 97.5 percentiles. Having similar attribute values has about no effect on tie creation, which is not significantly different from what is to be expected from a process where adoption decision is uncorrelated among neighbors.

Having established that good types tend to be contaminated by bad types, while bad types tend to differentiate from good types, I discuss in the conclusion the implications of these findings for optimal organizational design.

## 5 Conclusion

This paper answered one simple question: within an organization, how much to good apples deter bad apples, and how much do bad apples corrupt good apples? I derived a theoretical model to highlight that depending on the strength and direction of such relationship, the structure of the organization that best curtails corruption varies widely. I then showed that in this field setting, social interactions deter corruption: honest agents are contaminated by their dishonest colleagues, and tend to increase their dishonest behavior after interacting with dishonest colleagues. Conversely, dishonest employees differentiate from their honest colleagues: they tend to become more dishonest after interacting with dishonest colleagues.

The results provide micro-foundations previous macro-empirical on corruption in organizations, and force us to put them in a broader perspective. Similar to this paper, both [Evans (1995)](https://www.jstor.org/stable/10.1093/academ傲journals/kxy040) and [Carpenter and Moss (2013)](https://www.jstor.org/stable/10.1093/academ傲journals/kxy040) have examined cases in which dense social ties within the bureaucracy, and between the regulator and the industry may lead to increased corruption and regulatory capture. This paper, however, uses stronger micro-foundations that allow deriving the optimal organizational structure resulting from these micr-level interaction: in light of the empirical results, propositions [1] and [5] tell us that the optimal organization is a minimally connected one. Further research should, however, verify that the micro-mechanisms posited by the model are indeed at play in the cases examined by previous work, and examine whether some empirical support can be found for the other cases highlighted by propositions [4] and [5].

The findings have two important implications. First, they show that organizational structure
affects the emergence and extent of corruption and call into question the soundness of some popular policies that aim at tackling corruption through organizational change. By structuring the interpersonal relationships that take place within them, organizations give birth to patterns of contagion that spread corruption. This may explain why policies that combat corruption through major reforms of the organizational chart have proven so popular in recent years (e.g. Bennet, 2012; Friedman, 2012; Hausman, 2011). However, should the results from this paper travel to other organizations, they call into question the soundness some of such policies. In particular, one such popular policy is the creation of one-stop shops, a type of organizational unit that regroups previously disparate divisions into a single organization, therefore increasing the amount of social interactions between divisions. This might have a positive impact in terms of productivity, for it allows for better interactions among divisions but also results in increased opportunities for corruption to spread.

Second, coupling the empirical findings with theory have important implications for both organizational design and extant micro-empirical work. Coupling the empirical results with proposition 4 tells us that increasing the amount of social interactions within the organization should increase corruption. As such, the optimal organization is the minimally connected organization. However, propositions 4 and 5 also told us that the structure of the optimal organization varies widely depending on the direction, relative magnitude, and preconditions of the organization. These results tell us that existing micro-empirical results, typically coming from the lab, and focusing on one side of the relationship – that is, either how good apples react to bad apples, or how bad apples react to good apples – are useful but incomplete. They are useful in that they measure this interaction with high internal validity, since the lab affords both control and high-quality measurement. They are, however, incomplete in that they prevent us from deriving conclusions as to optimal organizational design. Measuring the relative magnitude of these two effects makes little sense in a lab setting, where treatments may be made arbitrarily strong. As such, future research should move to field settings in order to calibrate these two effects more realistically.
References


**URL**: https://www.dropbox.com/s/2lz4c4ddl0w4cj9/corruption_ferrali.pdf?dl=0


International Monetary Fund - Staff Team from the Fiscal Affairs Department and the Legal Department. 2016. Corruption: Costs and mitigating strategies - IMF Staff Discussion Note no. SDN/16/05. Technical report.


A Proofs

Proof of proposition 1. Solving for $H(0) = 0$, we find $H(0) = 0 \iff \alpha_q = 0$. Similarly, solving for $H(1) = 1$, we find $H(1) = 1 \iff \alpha_p = 0$. \hfill \qedsymbol

Proof of proposition 2. Note that
\[
\frac{\partial \rho(d)}{\partial \theta} = \frac{d [\beta_q p(0, d) + \beta_p q(0, d)]}{[p(\theta, d) + q(\theta, d)]^2}
\]
Because $p(0, d), q(0, d) > 0$, if $\beta_p$ and $\beta_q$ have the same sign, then $\frac{\partial \rho(d)}{\partial \theta} \geq 0 \iff \frac{\partial \rho(d)}{\partial \theta} \geq 0$ for any $d, d' \in \{1, \ldots, |N| - 1\}$. As such, $H_g$ is monotonic. Furthermore,
\[
\frac{\partial^2 \rho}{\partial \theta^2} = -2d \frac{\beta_q p(0, d) + \beta_p q(0, d) (\beta_q - \beta_p)}{[p(\theta, d) + q(\theta, d)]^3}
\]
A similar argument shows that $H_g'$ is monotonic.

If $H_g$ and $H_g'$ are monotonic and $\alpha_p, \alpha_q \neq 0$, then $H_g$ has a unique fixed point. \hfill \qedsymbol

Proof of lemma 2. It suffices to show that $\theta' \geq \theta$ and $\overline{\theta}' \geq \overline{\theta}$.

The cases where $\theta = 0$ and $\overline{\theta} = 1$ are trivial. We have $\theta' \geq 0 = \theta$, and since $\overline{\theta}' \leq 1$, it must be that $\overline{\theta}' = \overline{\theta} = 1$.

Consider $\theta > 0$. I show that $\theta' \geq \theta$. Since $H_o(0) > 0$, it must be that $H_o(\theta) > \theta$ for any $x \in [0, \theta)$. So $H_o'(\theta) \geq H_o(\theta) > \theta$ for any $x \in [0, \theta)$. It must therefore be that $\theta' \geq \theta$.

Consider $\overline{\theta} < 1$. I show that $\overline{\theta}' \geq \overline{\theta}$. Suppose $\overline{\theta}' < \overline{\theta}$. Since $H_o(1) < 1$, it must be that $H_o'(\theta) < \theta$ for any $x \in (\overline{\theta}, 1]$. So $H_o(\overline{\theta}) \leq H_o'(\theta) < \overline{\theta}$, a contradiction. \hfill \qedsymbol

Proof of proposition 3. Note that if $\frac{\partial \rho(d)}{\partial \theta} \geq 0$ for any $(\theta, d)$, then $\frac{\partial H_o}{\partial \theta} \geq 0$ for any $\theta$. Using lemma 1, this implies $\theta' \leq \theta$. Similarly, if $\frac{\partial \rho(d)}{\partial \theta} \leq 0$ for any $(\theta, d)$, then $\frac{\partial H_o}{\partial \theta} \leq 0$ for any $\theta$, which implies $\theta' \geq \theta$. We have:
\[
\begin{align*}
\frac{\partial \rho(d)}{\partial \alpha_p} &= -\frac{q(\theta, d)}{[p(\theta, d) + q(\theta, d)]^2} \leq 0 \\
\frac{\partial \rho(d)}{\partial \beta_p} &= -\frac{(1 - \theta) q(\theta, d)}{[p(\theta, d) + q(\theta, d)]^2} \leq 0 \\
\frac{\partial \rho(d)}{\partial \alpha_q} &= \frac{p(\theta, d)}{[p(\theta, d) + q(\theta, d)]^2} \geq 0 \\
\frac{\partial \rho(d)}{\partial \beta_q} &= \frac{\theta dp(\theta, d)}{[p(\theta, d) + q(\theta, d)]^2} \geq 0
\end{align*}
\]
\hfill \qedsymbol

Proof of proposition 4. Note that if $P'$ FOSD $P$, then $\tilde{P}'$ FOSD $\tilde{P}$, and that
\[
\frac{\partial \rho(d)}{\partial d} = \frac{\theta \alpha_p \delta_q - (1 - \theta) \alpha_q \delta_p}{[p(\theta, d) + q(\theta, d)]^2}
\]
If \( \beta_p < 0 \) and \( \beta_q \geq 0 \), then \( \frac{\partial \rho(d)}{\partial d} \geq 0 \). Since \( \rho \) is non-decreasing, \( \tilde{P}' \) FOSD \( \tilde{P} \) implies

\[
\sum_d \tilde{P}'(d)\rho(\theta, d) \geq \sum_d \tilde{P}(d)\rho(\theta, d),
\]

that is \( H_{\theta}'(\theta) \geq H_\theta(\theta) \) for any \( \theta \). By lemma 1, \( \theta' \leq \theta \).

If \( \beta_p < 0 \) and \( \beta_q \geq 0 \), then \( \frac{\partial \rho(d)}{\partial d} \leq 0 \). If \( \theta' \) is non-decreasing, FOSD implies

\[
- \sum_d \tilde{P}'(d)\rho(\theta, d) \geq - \sum_d \tilde{P}(d)\rho(\theta, d),
\]

that is \( H_{\theta}'(\theta) \leq H_\theta(\theta) \) for any \( \theta \). By lemma 1, \( \theta' \geq \theta \).

**Proof of proposition 5.** Suppose that \( \beta_p, \beta_q \geq 0 \) and \( \theta^* \geq \hat{\theta} \). It suffices to show that \( \theta'^* \geq \theta^* \).

Note that

\[
\frac{\partial \rho(d)}{\partial \theta} = \frac{d[\beta_qp(0, d) + \beta_pq(0, d)]}{[p(\theta, d) + q(\theta, d)]^2}
\]

Since \( \beta_p, \beta_q \geq 0 \), then we have \( \frac{\partial \rho(d)}{\partial \theta} \geq 0 \) for any \( d \), which implies that \( \frac{\partial H_{\theta}}{\partial \theta}, \frac{\partial H_{\theta}'}{\partial \theta} \geq 0 \).

Furthermore, note that

\[
\frac{\partial \rho(d)}{\partial d} \geq 0 \iff \theta \geq \frac{\alpha_q \beta_p}{\alpha_q \beta_p + \alpha_p \beta_q} \equiv \hat{\theta}
\]

This and \( P'(d) \) FOSD \( P(d) \) implies \( H_{\theta}'(\theta) \geq H_\theta(\theta) \iff \theta \geq \hat{\theta} \).

Suppose that \( \theta'^* < \hat{\theta} \). First, it must be that \( \theta^* \geq \hat{\theta} \). Indeed, we have \( H_{\theta}(0) > 0 \). If \( \theta^* < \hat{\theta} \), then \( H_{\theta}'(\hat{\theta}) = H_\theta(\hat{\theta}) \geq \hat{\theta} \) implies that \( H_\theta \) must have another fixed point in \( (\theta^*_p, \hat{\theta}) \), contradicting proposition 2.

Suppose that \( \theta'^* \in (\hat{\theta}, \theta^*) \). Since \( H_\theta \) is increasing and has a unique fixed point at \( \theta^*_p \), it must be that \( H_\theta(\theta) > \theta \) for any \( \theta \in (\hat{\theta}, \theta^*_p) \). This implies \( H_{\theta}'(\theta'^*) \geq H_\theta(\theta') > \theta^* \), a contradiction.

The three other cases prove similarly. \( \square \)
## B Main empirical result

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dependent variable: Pr(type switch)(_{t+1})</td>
<td></td>
<td></td>
</tr>
<tr>
<td>bad type(_t)</td>
<td>0.093(^*)</td>
<td>0.147(^{***})</td>
</tr>
<tr>
<td></td>
<td>(0.052)</td>
<td>(0.046)</td>
</tr>
<tr>
<td>N opposite type(_t)</td>
<td>0.059(^{***})</td>
<td>0.059(^{***})</td>
</tr>
<tr>
<td></td>
<td>(0.005)</td>
<td>(0.005)</td>
</tr>
<tr>
<td>bad type × N opposite type(_t) (\gamma_2)</td>
<td>-0.070(^{***})</td>
<td>-0.071(^{***})</td>
</tr>
<tr>
<td></td>
<td>(0.007)</td>
<td>(0.007)</td>
</tr>
<tr>
<td>(\gamma_1 + \gamma_2)</td>
<td>-0.011(^{**})</td>
<td>-0.012(^{**})</td>
</tr>
<tr>
<td>Market FE</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Month FE</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Controls</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Observations</td>
<td>2,142</td>
<td>2,142</td>
</tr>
<tr>
<td>R(^2)</td>
<td>0.159</td>
<td>0.168</td>
</tr>
</tbody>
</table>

*Note:* \(^*p<0.1; \ ^{**}p<0.05; \ ^{***}p<0.01\)

Table 5: **Main result with a threshold in s-score of 0.3.** Estimates for the specification reported in equation 5. Controls included in model (2) are reported in footnote 7. Standard errors are clustered at the month and market levels. The significance stars for row \(\gamma_1 + \gamma_2\) are based on the results of an \(F\)-test for the null hypothesis \(H_0: \gamma_1 + \gamma_2 = 0\). Good types are contaminated by bad types (\(\gamma_1 > 0\)), while bad types differentiate from good types (\(\gamma_1 + \gamma_2 < 0\)).