Subjective Performance Evaluation, Influence Activities, and Bureaucratic Work Behavior: Evidence from China

Alain de Janvry*, Guojun He•, Elisabeth Sadoulet‡,

Shaoda Wang†, Qiong Zhang▪

[September 2020]

Subjective performance evaluation is widely used by firms and governments to provide work incentives. However, delegating evaluation power to local senior leadership could induce influence activities: agents might devote much effort to pleasing their supervisors, rather than focusing on productive tasks that benefit their organizations. We conduct a large-scale randomized field experiment among Chinese local government employees and provide the first rigorous empirical evidence on the existence and implications of influence activities. We find that employees do engage in evaluator-specific influence to affect evaluation outcomes, and that this process can be partly observed by their co-workers. However, introducing uncertainty in the identity of the evaluator discourages evaluator-specific influence activities and significantly improves the work performance of local government employees.

Keywords: subjective evaluation, influence activities, civil servants, work performance

JEL: M12; D73; F63

* University of California, Berkeley. Email: alain@berkeley.edu.
• Hong Kong University of Science and Technology. Email: gjhe@me.com.
‡ University of California, Berkeley. Email: esadoulet@berkeley.edu.
† University of Chicago. Email: shaoda@uchicago.edu.
▪ Renmin University of China. Email: zhangqiong8@ruc.edu.cn.

I. Introduction

For a large share of jobs in modern economies, objective performance measures are difficult to obtain, leading employers to rely heavily on supervisors' subjective evaluations to provide work incentives (Prendargast, 1999; Deb et al., 2016). This is particularly ubiquitous in the public sector, due to the inherent problems of measurability and the multiplicity of tasks for most civil service jobs (Olken and Pande, 2013; Finan et al., 2015).

While subjective performance measures could potentially improve contractual power (Gibbons and Murphy, 1992; Baker et al., 1994), they also open the door to influence activities: employees can take actions to affect the evaluator's assessment in their favor, which might be detrimental to the interests of the organization (Milgrom and Roberts, 1988; Milgrom, 1988). For the organization, influence activities thus create a tradeoff between taking advantage of the supervisor's local information and having the supervisor captured by his subordinates, undermining the organization's ability to fully incentivize the agents to perform their jobs. Despite being consistent with anecdotal examples and case studies, the long-standing theoretical discussion on the existence and consequences of influence activities is yet to be backed up by rigorous empirical evidence (Oyer and Schaefer, 2011; Lazear and Oyer, 2012).

Empirically studying influence activities is challenging for at least three reasons. First, behaviors such as buttering up supervisors or providing personal favors to them are by nature difficult to observe -- agents will try their best to hide such behaviors because they are usually deemed inappropriate. Second, even if these behaviors are observed, it is difficult to conclude that they are driven by intentions of improving evaluation outcomes (instead of simply being friendly), making it difficult to classify them exclusively as influence activities. Third, even if the existence of influence activities is established, quantifying the effects of these activities on work performance still requires exogenous variation in such behaviors across agents.

In this paper, we conduct a large-scale field experiment in two Chinese provinces, which aims at addressing these three challenges and provides the first rigorous empirical evidence on the existence and consequences of influence activities in the workplace. Our experiment focuses on China's "3+1 Supports" program, which hires more than 30,000 college graduates annually to work in rural township governments on two-year contracts. These junior government employees are referred to as College Graduate Civil Servants (CGCSs) in this paper.

An important institutional feature of the Chinese governance system is its dual-leadership arrangement (Shirk, 1993), whereby every government organization/subsidiary has two leaders: a "party leader" (i.e., party secretaries at various levels) and an "administrative leader" (i.e., the head in a village, the mayor in a city).[1] As a result of this dual system, every CGCS reports to two supervisors who both assign him job tasks and provide performance feedback on a regular basis.[2] In the *status quo,* every CGCS is evaluated by one of his two supervisors every year. The evaluation outcome will determine whether the CGCS can be promoted to a "tenured" position upon completing his two-year term, a highly sought-after outcome for most CGCSs due to the prestige of permanent civil service jobs in China.

Exploiting this unique setting, we collaborate with two provincial governments in China to randomize two performance evaluation schemes among their 3,785 CGCSs. In both schemes, we randomly select one of the two supervisors to be the "evaluator." The only difference is that, in the "revealed" scheme, we announce the identity of the evaluator to the CGCS at the beginning of the evaluation cycle, so that throughout the year the CGCS knows whose opinion is 100% responsible for his promotion; in the "masked" scheme, we keep the identity of the evaluator secret until the end of the evaluation cycle, so that throughout the year the CGCS perceives each supervisor to have a 50% chance of determining his promotion.

We find that, in the revealed scheme, the randomly selected evaluating supervisor gives significantly more positive assessments of CGCS performance than his non-evaluating counterpart. In addition, when we ask colleagues of the CGCS to speculate which of the two supervisors would give higher assessments to the CGCS, they are also more likely to (correctly) point to the randomly selected evaluator, whose identity was not revealed to them. These results are consistent with a scenario where the agent is able to engage in evaluator-specific influence activities to improve evaluation outcomes, and such influence activities can be at least partially observed by his co-workers.

In the masked scheme, we find no such asymmetries in supervisor assessments and colleague perceptions. Furthermore, we find that the masked scheme, by discouraging influence activities and inducing productive efforts, improves CGCS performance according to a series of measures, such as average colleague assessment, non-evaluating supervisor

[1] The two leaders have large overlaps in their responsibilities, introducing *de facto* checks and balances. See Li (2018) for information on the institutional details of the duality system.

[2] For simplicity, we use the male pronouns (he/his/him) for the CGCS throughout this paper. 62% of the CGCS are females.

assessment, likelihood of being recommended for tenure, and monthly bonuses determined by objective performance indicators. The return to masking evaluator identity is comparable to the performance gap between four-year (regular) college graduates and three-year (community) college graduates, suggesting that the "masked scheme" generates economically significant efficiency gains. Together with the results from the revealed scheme, our findings suggest that influence activities are prevalent in China's administrative system and are detrimental to local bureaucratic performance.

Turning to heterogeneity, we find that the effect of the masked scheme is stronger when the CGCS is more risk-averse, and when the two supervisors have more aligned preferences for performance, assign similar amounts of job tasks, and have smaller information asymmetry. As discussed in Section III, these patterns can be rationalized in a simple conceptual framework, where introducing uncertainty in the identity of the evaluator makes supervisor-specific influence costlier and riskier, therefore incentivizing the agent to reallocate efforts from evaluator-specific influence activities to productive tasks that are appreciated by both supervisors.

We conduct a battery of additional tests to rule out alternative interpretations of our findings. We find that the assessment asymmetry under the revealed scheme is not driven by any behavioral change of the evaluator, nor by any additional information about CGCS performance being presented to him. We also find that the "improved performance under the masked scheme" cannot be explained by the CGCS engaging in even more influence activities toward his supervisors and colleagues. Taken together, all the empirical results consistently support our proposed mechanism, namely the CGCS's differential efforts at impressing the evaluator instead of the non-evaluator, over other explanations.

This paper speaks to three strands of literature. First and foremost, it provides the first rigorous empirical test for the existence and implications of influence activities in the workplace. As pointed out by Lazear and Oyer (2012), while a large theoretical literature has studied how agents try to engage in influence activities in the workplace (e.g., Milgrom and Roberts, 1988; Milgrom, 1988; Meyer et al., 1992; Schaefer, 1998, Alonso et al., 2008; Powell, 2015), there is a lack of rigorous empirical evidence, aside from anecdotes and case studies, to verify these arguments.[3] Our paper fills this gap by providing field experimental evidence on

[3] Rasul and Rogger (2016) finds a negative correlation between incentives/monitoring practice and public project completion in Nigeria, which is stronger for more experienced bureaucrats. This empirical pattern is

influence activities among Chinese local government employees, as well as quantifying the causal impact of reducing influence activities on job performance. More broadly, subjective performance evaluation is ubiquitous across both the private and public sectors, and has been investigated extensively by a large body of theoretical work (Gibbons and Murphy, 1992; Baker et al., 1994; Prendergast and Topel, 1996; MacLeod, 2003, Maestri, 2012; Deb et al., 2016). However, empirical evidence on the effectiveness and limitations of subjective evaluation is still largely missing, with only a handful of exceptions (Chevalier and Ellison, 1999; Hayes and Schaefer, 2000). Our paper contributes to this literature by showing how influence activities can undermine the power of subjective performance evaluations.

Second, this paper adds to a growing empirical literature on the personnel economics of the developing state, specifically on incentivizing public employees (Finan et al., 2016). Most of the existing work on this topic focuses on the role of financial incentives,[4] with only a few exceptions studying non-pecuniary incentives such as transfers and postings (Banerjee et al., 2012), social incentives (Ashraf and Bandiera, 2018), and intrinsic motivation (Ashraf at al., 2014). Our paper adds to this line of work by exploiting the (implicit) career incentive involved in performance evaluations, which is a prevalent form of motivation in public sectors due to an often compressed wage structure, but has rarely been studied in the literature.[5] In addition, we show that holding the "career reward" fixed, a slight refinement of the performance evaluation practice can lead to a substantial improvement in bureaucratic performance, indicating a highly cost-effective way of enhancing state effectiveness.

Third, our paper relates to the research agenda on Chinese political meritocracy. Since Li and Zhou (2005), a large number of empirical studies have tried to investigate how the design of various performance indicators, such as fiscal revenue (Lü and Landry, 2014), environmental standards (He et al., 2020), and population control (Serrato et al., 2019), affect the behaviors of provincial and prefectural officials in China. However, existing evidence has focused almost exclusively on high-level government officials, leaving incentives and constraints for the vast majority of local bureaucrats under-researched, even though they could

consistent with bureaucrats learning to engage in influence activities over time. Our paper complements Rasul and Rogger (2016) by experimentally altering the bureaucrats' incentives to engage in influence activities, which allows us to causally evaluate the existence and consequences of these activities in the public sector.

[4] See Finan et al. (2016) for a summary.

[5] Previous research focuses on the selection effect of career incentives; see Ashraf et al. (2020) for example. Our paper complements this line of work by investigating the "intensive margin" impact of career incentives, while holding selection fixed.

differ substantially from those for high-level leaders.[6] Our paper intends to unravel the black box of incentive schemes for grassroots bureaucrats in China, who are the building blocks of state capacity and play key roles in public service delivery. More broadly, this paper also adds to an emerging literature on bureaucratic performance in developing countries (Bertrand et al., 2019; He and Wang, 2017).

The remainder of this paper is organized as follows. In Section II, we introduce the institutional background, design, and implementation of our field experiment. In Section III, we lay out a simple conceptual framework to help rationalize the experimental design, and generate testable hypotheses to guide the empirical analysis. In Section IV, we present experimental results testing the theoretical predictions. In Section V, we discuss potential alternative interpretations for our findings. Section VI concludes.

II. Background and the Experiment

A. Institutional Background

Since the early 2000s, the Chinese government has launched several large-scale public employee assigment programs, which altogether hired more than one million college graduates to work with local governments in rural areas, in the hope that their modern human capital and independence from local interest groups could help improve state effectiveness at the grassroots level. For example, in the College Graduate Village Officials (CGVO) program, new college graduates were hired as village officials on a contractual basis, and the arrival of CGVOs in rural villages improved policy implementation and reduced leakages in poverty subsidy distribution (He and Wang, 2017).

In this paper, we focus on the "3+1 Supports" initiative -- a human capital building program for local governments launched in 2006 by the Ministry of Human Resources and Social Security.[7] Through this program, college graduates are hired to work as temporary civil

[6] For instance, a key distinction is that job tasks for low-level bureaucrats are much more difficult to quantify with objective measures such as GDP growth and environmental quality, so most grassroot bureaucrats are only rewarded based on subjective evaluations by their supervisors.

[7] In Chinese, the initiative corresponds to the "*San Zhi Yi Fu* (三支一扶)" program. Six other ministries and departments co-sponsored the program, including the Ministry of Education, the Ministry of Finance, the Ministry of Agriculture, the National Health Commission, the State Council Leading Group Office of Poverty Alleviation and Development, and the Communist Youth League Central Committee.

servants in rural townships. They assume four types of positions: township government clerks focusing on poverty alleviation, township government clerks focusing on agricultural support, teachers in township primary schools, and nurses in township clinics. By the end of 2018, more than 350,000 college graduates had been hired as "College Graduate Civil Servants" (CGCSs) through this program.

The CGCSs are recruited nationwide on a yearly basis. In May, before the end of the school year, each provincial government announces vacancies on its website and invites college graduates to apply. In most provinces, the procedure for CGCS recruitment is similar to that for recruiting regular state employees: applicants first take a comprehensive written exam, which is similar to the Administrative Aptitude and Essay Writing Tests in the National Civil Service Exam. High-scoring applicants are then interviewed, and top-ranked candidates (based on combined scores) are recruited. Some provinces forgo tests and interviews, and screen applicants simply based on their application materials.

Admission to CGCSs is highly competitive. In most provinces, the acceptance rate for the "3+1 Supports" program is consistently below 10%. For example, Shandong province had around 1,500 positions in 2017 and attracted over 31,000 applicants (acceptance rate < 5%); in Guangxi province, the government planned to hire 800 CGCSs in 2017 and the total number of applicants exceeded 13,600 (acceptance rate < 6%). Such intense competition ensures the high quality of selected CGCSs.

The job tasks of a CGCS are similar to those of a regular entry-level township civil servant. For instance, for CGCSs in clerical positions, like for regular rural civil servants, job tasks tend to be a combination of routine paperwork, visits to villages, interactions with villagers, and other case-based assignments from supervisors. The multi-dimensional and vaguely-defined nature of these job tasks makes it infeasible to objectively compare job performance across different individuals.[8] As a result, in the *status quo*, the evaluation of a CGCS relies solely on the evaluating supervisor's subjective assessment, which is also the norm for the vast majority of regular civil service jobs in China as well as across the world.

[8] For CGCSs in more specialized positions like township clinic nurses or primary school teachers, job tasks are also similar to those of their colleagues who are formal public employees. While some dimensions of these jobs are better defined than those of clerical jobs, objective performance evaluation remains difficult. For example, student scores cannot be used to incentivize teachers due to lack of unified written exams at the primary school level.

The only major difference between a CGCS position and a regular civil servant position is that the former is based on a two-year contract while the latter is "tenured."[9] The majority of CGCSs are eager to be promoted to tenured positions upon finishing their two-year terms, which would only be approved by the government with a satisfactory supervisor evaluation.[10] As a result, aspiring CGCSs have exceptionally strong incentives to impress their supervisors in order to improve evaluation outcomes. While such incentives could encourage greater work efforts, they might also induce influence activities that are misaligned with the government's interest. Simple examples of such non-productive influence activities include picking up the supervisor's kids from school, making coffee for the supervisor, doing personal chores for the supervisor, etc.

Every CGCS reports to two supervisors: a party leader and an administrative leader. This is due to China's unique dual-leadership governance structure: in principle, the administrative leader is in charge of the day-to-day operation of the government entity, while the party leader oversees the process and has the final say in the most high-stakes decisions. These two leaders have the same official ranking, but the party leader is normally perceived to have an edge in authority. At the grassroots level, such as a township (which is the lowest layer of formal bureaucracy), the division of labor between the two leaders often becomes less clear, and there tends to be substantial overlap in their roles. This dual arrangement provides *de facto* checks and balances in China's local governance system (Li, 2018), and is adopted by various levels of administrative units (ranging from the central government to the village committees). It is also enforced in public institutions such as schools, hospitals, and state-owned firms, as long as there are more than three Communist Party members among the employees.

Under the current evaluation scheme, when a CGCS is first assigned to a township by the provincial Department of Human Resources, he is explicitly told that he needs to work under the supervision of both leaders in his unit, but the Department of Human Resources has designated one of the two leaders as the "evaluator" who is responsible for evaluating the CGCS's performance at the end of the year. The CGCS, therefore, knows whose opinion matters for his career development, starting at the beginning of his appointment.

[9] In this setting, "tenure" corresponds to the "*Bian Zhi* (编制)" status, which is essentially a permanent contract provided by the government.

[10] In the provinces where our study took place, on average about 40% of the CGCSs subsequently become permanent civil servants.

B. Experimental Design

In this section, we explain the experimental design and discuss the intuitions for our main hypotheses. A formal rationalization of the experiment is presented with a conceptual framework in Section III.

The experimental design is illustrated in Figure 1. In collaboration with two provincial governments in China, we randomize two subjective performance evaluation schemes, a "revealed" scheme and a "masked" scheme, across all their 3,785 CGCSs employed in 2017.

For every CGCS in our sample, one of his two supervisors is randomly selected to be the evaluator, meaning that this supervisor's assessment will be given a 100% weight in the final evaluation outcome, which in turn affects the CGCS's future career development. For the randomized non-evaluator, we also collect his assessment of CGCS performance, but this assessment will be given no weight in determining the evaluation outcome. In both schemes, we never directly inform a supervisor whether or not he is chosen as the evaluator, nor do we inform the CGCS's colleagues.

Two-thirds of the CGCSs in our sample are assigned to the "revealed" scheme. In this scheme, we inform the CGCS about the identity of his evaluating supervisor at the beginning of the evaluation cycle. This essentially mimics the current system of CGCS performance evaluation, where the agent is informed *ex-ante* about the evaluating supervisor. The key difference is that, in the current system, the evaluator is endogenously chosen from the two supervisors, typically through an opaque process combining supervisor opinions, division of labor between supervisors, and other idiosyncratic factors. In our "revealed" scheme, by randomly selecting the evaluator, endogeneity in evaluator selection is eliminated.

We exploit the revealed scheme to test whether knowing the evaluator's identity generates asymmetry in supervisor assessments. Since the evaluator is randomly selected from the two supervisors, in the absence of any evaluator-specific influence activities, both supervisors should on average give similar assessments of CGCS performance. However, if the CGCS indeed engages in evaluator-specific influence activities, we should observe an asymmetry in the two supervisors' assessments of the same CGCS. At the end of the evaluation cycle, we also asked the CGCS's colleagues to speculate who among the two supervisors would be more positive about the CGCS's performance. While we never informed colleagues which supervisor was randomized as the evaluator, to the extent that colleagues can observe some

of the evaluator-specific influence activities extended by the CGCS, they will be systematically more likely to (correctly) speculate that the evaluator will give more positive assessments than the non-evaluator.

The remaining one-third of the CGCSs are assigned to the "masked" scheme. In this scheme, while we still randomize one of the two supervisors as the evaluator, we do not inform the CGCS about the identity of the evaluator until the end of the evaluation cycle. Therefore, from the CGCS's perspective, each supervisor will have a 50% chance of determining his evaluation outcome. Compared to the revealed scheme, the masked scheme reduces the relative return to supervisor-specific influence activities; if the CGCS puts effort into influencing a specific supervisor, there is a 50% chance that this supervisor will not end up evaluating his performance, reducing the expected return from engaging in influence activities. As a result, under the masked scheme, a CGCS has incentives to reallocate his efforts from influence activities toward productive activities that will be appreciated by both supervisors; this should improve overall work performance.

Exploiting the randomization of CGCSs into the "revealed" vs. "masked" schemes, we can test whether introducing uncertainty in evaluator identity indeed improves CGCS performance. Our benchmark performance indicator is the average assessment given by other colleagues. We define "colleagues" as co-workers in the same office as the CGCS who were not hired through the "3+1 Supports" program. We consider the colleagues' assessments as a credible performance measure in this context for three reasons. First, the colleagues are randomly chosen from the same office where the CGCSs work. They work closely with CGCSs and can thus best observe the CGCSs' performances. Second, there is no obvious conflict of interest between the CGCSs and their colleagues. Unlike the CGCSs, who work in the office only under 2-year contracts, most colleagues already have tenure and have worked in the office for many years. As a result, the CGCSs and their colleagues do not directly compete with each other for career advancement. Finally, the CGCSs do not have obvious incentives to influence their colleagues for evaluation purposes; at the beginning of the experiment, the provincial governments explicitly told each CGCS that only the evaluating supervisor's opinion would count for promotion.[11]

[11] Most CGCSs, in fact, did not even expect that we would survey their colleagues until the enumerators were sent to their workplaces at the end of the experiment.

To complement colleague assessments, we also measured CGCS performance in three other ways. First, we elicited performance assessments given by both the evaluating supervisor and the non-evaluating supervisor. Second, to make sure that the performance results were not driven by "cheap talk," we asked the supervisors and colleagues to make a recommendation to the provincial government as to whether the CGCS should be promoted to a permanent position in the current work unit. Third, we tried to benchmark performance objectively using the actual salaries received by the CGCSs. While it is difficult to measure performance objectively due to the multi-dimensional and vaguely-defined job tasks for most CGCSs, for some CGCS positions, a modest amount of "monthly bonus" is linked to well-defined objective performance indicators.[12] Therefore, we can compare the actual salaries received by CGCSs between the two schemes, and infer the differences in objective performance measures (based on the bonus pay algorithms).

C. Implementation

Our experiment is conducted in collaboration with two of the largest provinces in China, with a combined population of more than 150 million. Province A is coastal and more developed, while Province B is inland with lower average income. Province A recruits CGCSs using an automatic scoring system, while Province B recruits through traditional written tests and interviews. Our sample covers all 3,785 CGCSs employed by these two provinces as of September 2017 (cohorts of 2016 and 2017). Our research team was appointed by the two provincial Human Resources Departments as the third-party evaluator for their "3+1 Supports" programs to help launch new performance evaluation schemes of which the provincial governments officially informed all the CGCSs. This high-level endorsement helped ensure that the vast majority of CGCSs were well aware of the high stakes involved in the evaluation outcomes under our newly introduced evaluation schemes.

The baseline survey was carried out in September 2017, one month after the 2017 CGCS cohort finished job training and were assigned to their positions. Every CGCS was then randomized into one of the two evaluation schemes. The randomization was conducted at the work unit level instead of the individual level.[13] Different CGCSs working in the same unit

[12] For example, CGCSs who serve as nurses receive bonuses based on the number of night shifts they take.
[13] In Chinese, a work unit corresponds to a "*Gong Zuo Dan Wei* (工作单位)."

(i.e., an organization branch led by the same set of supervisors) were assigned to the same scheme. This is at the request of our government partners, to ensure that the evaluation outcomes of CGCSs working in the same unit could be fairly compared to each other. In this setting, because 83.9% of the work units had only one CGCS assigned, randomizing at the work unit level instead of the individual level did not hurt our statistical power in any substantial way.

Based on the randomization, in September 2017, we informed every CGCS about the evaluation scheme to which he had been assigned. Specifically, if a CGCS was randomized into the revealed scheme, we notified him that "*among your two supervisors A and B, we randomly selected supervisor A to be your evaluator, whose opinion will be collected by the end of this evaluation cycle and provided to the provincial Human Resources Department for their review.*" If a CGCS was randomized into the masked scheme, we notified him that "*among your two supervisors A and B, we will randomly select one of them to be your evaluator. The randomization will be realized at the end of this evaluation cycle, at which time the evaluator's opinion will be collected and provided to the provincial Human Resources Department for their review."* The individualized notification letters are translated in Appendix C.

To ensure the credibility of our intervention, the two provincial governments sent formal notifications with official stamps to every CGCS. The government notifications emphasized the importance of this "third-party" performance evaluation and confirmed the design of the evaluation schemes that we sent to the CGCSs. We reminded the CGCSs about their evaluation schemes in January 2018.

The end-line survey was carried out in June 2018, and consisted of three parts: colleague assessment, supervisor assessment, and self-assessment. When the enumerators visited the office where a CGCS worked, if there were fewer than five colleagues in the office, all of them were invited to fill in the colleague questionnaire; if there were more than five colleagues, the surveyor randomly sampled five of them to fill in the colleague questionnaire, using a random number generator.[14] To protect the privacy of colleagues and encourage truth-telling, colleague questionnaires were strictly anonymous, and CGCSs were not allowed to communicate with colleagues during the entire process. The CGCS survey was also conducted on-site, but independently from the colleague survey to avoid interference. Supervisor assessment was

[14] If a colleague was not at the office when the enumerator visited, his contact information was collected and he was surveyed over the phone the following day. To ensure data accuracy, the leader of the surveying team randomly called some of the surveyed colleagues on the following days to verify the sampling procedure and the answers collected.

completed online, with an individual-specific link for each supervisor, listing all the CGCSs in his unit.

In the colleague and supervisor surveys, we collected information on the main characteristics of the colleague/supervisor, their interactions and familiarity with the CGCS, the job tasks of the CGCS, and their assessments of the CGCS along various dimensions. Specifically, we asked for an overall assessment of CGCS performance, as well as a "revealed preference" measure asking each colleague/supervisor whether he recommends that the CGCS be promoted to a permanent civil servant in the current work unit.

The end-line CGCS survey followed a similar structure by asking about interactions with supervisors/colleagues and self-assessments along multiple dimensions. In addition, we also asked a series of questions related to future career plans and satisfaction with the "3+1 Supports" program.

D. Balance and Attrition Tests

To ensure that the randomization was well executed, we conduct a battery of balance tests: in the revealed scheme, supervisor characteristics should be balanced between the evaluating and non-evaluating supervisors; between the two evaluation schemes, CGCS characteristics, supervisor characteristics, and colleague characteristics should all be balanced. As shown in Appendix Tables A1-A4, these balance tests are all satisfied.

Between the baseline and the end-line surveys, we lost 918 (24.3%) CGCSs in the sample. The main cause for attrition was that some CGCSs or their supervisors were re-assigned to different job posts during our study period (14.9%). For example, a CGCS could be relocated from one township to another because of changes in government priorities. The supervisors also could retire or be promoted or rotated to other institutions. Such job changes would break the supervisor-subordinate relationship defined by our intervention, and thus invalidate the experimental design. In addition, some CGCSs passed the formal civil service exams or got admitted to graduate schools, and decided to quit their jobs during our experiment (7.4%). To test whether our experiment suffers from any attrition bias, we estimate the relationship between attrition status and treatment status in Appendix Table A5. We find that our treatment does not lead to selective attrition.

III. Conceptual Framework

In this section, we present a simple conceptual framework to rationalize the experiment and derive the main propositions that will guide the empirical analysis.

Assume a CGCS works on a job with a productive dimension x, which can be observed by his supervisors and co-workers, but cannot be verified quantitatively. The organization therefore relies on a subjective performance evaluation scheme, where the agent's reward depends on the assessments given by his supervisors.

To mimic our empirical setting, we assume that there are two supervisors, $j \in \{1, 2\}$. In addition to working on the productive dimension of the job (x), the CGCS also has the option of exerting (supervisor-specific) influence activities u_j to please supervisor j, in order to improve his assessment score:

$$Y_j = v(x) + u_j,$$

where $v(x)$ is strictly increasing and strictly concave. Assume that the costs of producing objective performance and influence activities are both linear: $C(x) = ax; C(u_j) = b_j u_j$. For tractability, let $b_1 = b_2 = b$.

Each CGCS maximizes his utility subject to a time constraint:

$$Max_{x,u_{j_{\{j=1,2\}}}} \; V = v(x) + \sum_j s_j u_j - ax - \sum_j b u_j,$$

s.t.

$$x + \sum_j u_j = T,$$

where s_j is the probability of each supervisor j's assessment being used to determine the CGCS's reward in the performance evaluation scheme ($\sum_{j\in\{1,2\}} s_j = 1$). T is the total time budget for an individual.

Other colleagues also observe CGCS performance, but their opinions are not included in the performance evaluation scheme, so the agent is not incentivized to adjust his efforts to improve colleague assessments. Colleagues therefore receive no influence activities from the CGCS, and base their assessments solely on the productive dimension:

$$Y_c = v(x).$$

Suppose that one of the two supervisors is randomly chosen to evaluate CGCS performance, and the other supervisor's opinion bears no weight in the evaluation. When we inform the CGCS about the identity of the evaluator (*revealed scheme*), the CGCS knows exactly whose opinion matters for his career development: $s_1 = 1, s_2 = 0$ or $s_1 = 0, s_2 = 1$. When we do not inform the CGCS about the identity of the evaluator until the end of the evaluation cycle (*masked scheme*), the CGCS perceives each supervisor as equally likely to determine his career development: $s_1 = s_2 = \frac{1}{2}$. Solving the CGCS's maximization problem in the two schemes, under simple regularity conditions,[15] we can derive the main testable hypotheses that will guide the empirical investigations.

Proposition 1: *Under the revealed scheme, the agent extends evaluator-specific influence activities, and the evaluating supervisor gives a higher assessment than the non-evaluating supervisor.*

Without loss of generality, assume that $s_1 = 1, s_2 = 0$. So we can re-write the problem as:

$$Max_{u_1,u_2}\ E[V] = v(T - u_1 - u_2) + u_1 - a \cdot (T - u_1 - u_2) - bu_1 - bu_2$$

Solving this leads to:

$$x_r = v'^{-1}(1 + a - b);\ \ u_1 = T - v'^{-1}(1 + a - b);\ \ \ u_2 = 0.$$

Therefore we have: $u_1 > u_2;\ Y_1 > Y_2$.

The intuition is that, when the agent knows the identity of the evaluator, he exerts more evaluator-specific influence, which leads to a more positive assessment from the evaluating supervisor.

Proposition 1.1: *Under the revealed scheme, an agent engages in more influence activities when the cost of doing so is lower.*

If an agent has a personality that reduces the psychological cost of engaging in influence activities, b is lower. Since $v(x)$ is monotonically increasing and concave, $v'(x)$ and $v'^{-1}(x)$ are both monotonically decreasing. Therefore, we have: $\frac{dx_r}{db} = \frac{dv'^{-1}(1+a-b)}{db} > 0;\ \frac{du_1}{db} = -\frac{dv'^{-1}(1+a-b)}{db} < 0$.

Proposition 2: *Under the masked scheme, total influence activities decrease, while productive performance improves.*

[15] Specifically, we assume that $v'^{-1}(1 + a - b) > 0$ and $v'^{-1}\left(\frac{1}{2} + a - b\right) < T$, which ensures that there will be interior solutions for x and u in both schemes.

Since $s_1 = s_2 = \frac{1}{2}$ under the masked scheme, we can re-write the problem as:

$$Max_{u_1,u_2}\ E[V] = v(T - u_1 - u_2) + \frac{1}{2}(u_1 + u_2) - a \cdot (T - u_1 - u_2) - b(u_1 + u_2).$$

Solving this leads to:

$$x_m = v'^{-1}\left(\frac{1}{2} + a - b\right); u_1 + u_2 = T - v'^{-1}\left(\frac{1}{2} + a - b\right).$$

Since $v(x)$ is monotonically increasing and concave, $v'(x)$ and $v'^{-1}(x)$ are both monotonically decreasing. Therefore, $x_m > x_r; u_{1m} + u_{2m} < u_{1r} + u_{2r}$. This suggests that when we switch from the revealed scheme to the masked scheme, objective performance (x) increases, while total influence activities ($u_1 + u_2$) decrease.

The intuition is that, when evaluator identity is masked, the expected return to influence activities (u_j) is reduced by half, while the expected return to productive efforts (x) remains unchanged. This encourages the CGCS to reallocate his efforts from influence activities to productive performance.

This simple conceptual framework can be extended in several different ways, which provide additional predictions on the heterogeneous treatment effects of masking evaluator identity. These additional hypotheses are discussed intuitively below and proved formally in Appendix B.

Proposition 2.1: *When the CGCS is risk-averse, masking evaluator identity is more effective in improving performance.*

When masking evaluator identity, we are introducing uncertainty in the return to influence activities. As a result, risk-averse CGCSs have more incentives to reallocate efforts from the risky investment (influence activities) to the safe investment (productive dimension).

Therefore, we hypothesize that the return to the masked scheme is higher among the more risk-averse CGCSs.

Proposition 2.2: *When job tasks are multi-dimensional, masking evaluator identity is more effective when the two supervisors' subjective weights for different dimensions of performance are more consistent.*

As illustrated in the baseline model, the masked scheme improves CGCS performance because it increases the relative return to productive efforts, as compared to supervisor-specific influence activities. When the two supervisors have heterogeneous preferences about different dimensions of productive performance, from the CGCS's perspective, the return to productive performance also becomes more "supervisor-specific," which weakens the reason

that they perceived productive efforts to be more desirable than influence activities. Therefore, when the two supervisors have more aligned preferences regarding different dimensions of productive performance, the return to the masked scheme is larger.

Proposition 2.3: *If the two supervisors give different weights to the same productive task, when the asymmetry is larger, the revealed scheme is more effective, and the return to adopting the masked scheme is lower.*

Suppose there are two supervisors; one gives a larger weight to productive performance (x) and the other gives a smaller weight. In the revealed scheme, if the former (latter) is chosen as the evaluator, the CGCS will exert more (less) efforts on productive performance. Since the production function of productive performance is concave, such dispersion of weights on average leads to better performance than when the two supervisors have more similar weights, holding the average weight constant. In the masked scheme, the CGCS no longer knows which supervisor will evaluate him, so he no longer knows what the weight for productive performance will be. A risk-neutral CGCS then decides efforts based on only average weight, ignoring the difference in weights.

Therefore, we hypothesize that greater asymmetry in supervisors' weights for the same productive task leads to better performance in the revealed scheme, but not in the masked scheme.

Proposition 2.4: *If the two supervisors have imperfect information about job performance, the larger the information asymmetry is between the two supervisors, the less effective the masked scheme is.*

The intuition is similar to that of *Proposition 2.3*. Incompelete information about performance discourages productive efforts. In the revealed scheme, an information asymmetry between the two supervisors would increase the average effort from the CGCS due to the concavity of the production function. In the masked scheme, a risk-neutral CGCS would act on the average information for the two supervisors, ignoring the level of asymmetry between them.

Therefore, we hypothesize that greater information asymmetry between the two supervisors regarding performance leads to better performance in the revealed scheme, but not in the masked scheme.

IV. Main Results

In this section, we present a series of experimental results to verify the main propositions of our conceptual framework. In the revealed scheme, we find that the assessment given by the (randomized) evaluating supervisor is substantially higher than that given by the (randomized) non-evaluating supervisor, and this asymmetry in supervisor assessments is correctly predicted by the CGCS's colleagues, confirming *Proposition 1*. We find the asymmetries in supervisor assessments and colleague perceptions to be particularly salient when the CGCS has personality traits that we identify as "careerist" or "slick," as opposed to "altruistic" or "candid. "

When switching from the revealed scheme to the masked scheme, the asymmetries in supervisor assessment and colleague perceptions no longer exist. Instead, we find significant improvements in various measures of CGCS work performance, and suggestive evidence of reduced influence activities and increased work efforts, which, together, consistently support *Proposition 2*. In addition, our results also indicate that the masked scheme is more effective when the CGCS is more risk-averse, when the two supervisors have similar opinions on what constitutes good performance, when both supervisors assign similar amounts of job tasks to the CGCS, and when there is little information asymmetry between the two supervisors.

These findings are highly compatible with the interpretation of the agent undertaking influence activities (that target the evaluator and at the expense of job performance) in the revealed scheme, and reallocating efforts from influence activities toward productive dimensions in the masked scheme, as formalized in our conceptual framework. The combination of these patterns can hardly be reconciled with alternative interpretations, and we will further confront each of those remaining confounding mechanisms in Section V.

A. Proposition 1: Asymmetry in Supervisor Assessments under Revealed Scheme

In Table 1, the outcome variable is "Supervisor 1's assessment score minus Supervisor 2's score," measuring the extra positiveness of Supervisor 1 relative to Supervisor 2 toward the same CGCS. "Supervisor 1" and "Supervisor 2" are random labels we give to each CGCS's two supervisors. The explanatory variable is a dummy variable indicating whether Supervisor 1 was chosen (randomly) to be the evaluating supervisor. Here, we focus on the revealed

scheme, in which every CGCS is informed about the identity of his randomized evaluator at the beginning of the evaluation cycle.

In Column (1), we find that in the revealed scheme, if a supervisor was chosen as the evaluator at the baseline, he indeed gave a more positive assessment at the endline. In Column (2), we include a rich set of control variables in the regression, and the estimated coefficient remains unchanged, confirming that the treatment of "Supervisor 1 Evaluating" was indeed randomly assigned. This asymmetry in supervisor assessments is consistent with the agent engaging in evaluator-specific influence activities to improve evaluation outcomes.

If this asymmetry is indeed driven by influence activities, to the extent that such behaviors can at least partially be observed by other co-workers in the same office, we should expect that in the revealed scheme colleagues could update their priors on which of the two supervisors would be more positive about CGCS performance. In other words, when colleagues receive some signals of u_j, even without knowing who was chosen as the evaluator, they should still be more likely to correctly speculate that the evaluating supervisor would be more positive.

We test this prediction in Columns (3)–(4), where the outcome of interest is a dummy variable indicating whether a colleague thinks Supervisor 1 would be more positive than Supervisor 2. We see that when Supervisor 1 is randomly selected as the evaluator, colleagues are more likely to think he is going to give more positive assessments. Combined with the results in the first two columns, colleagues appear to correctly predict the direction of the asymmetry in supervisor assessments.

Our model also predicts that, when the CGCS's personality makes it (psychologically) easier for him to engage in influence activities, evaluator-specific influence activities under the revealed scheme should increase (*Proposition 1.1*). In the baseline survey, we elicited two dimensions of personality traits that are related to this hypothesis. First, we asked the CGCS, "whether you are willing to sacrifice your own career development for the welfare of the public." A positive answer is coded as "altruistic," while a negative answer is coded as "careerist." Second, we asked the CGCS, "if you do not like someone, whether you will make him aware of that." A positive answer is coded as "candid," while a negative answer is coded as "slick." We hypothesize that a careerist/slick CGCS will find it more comfortable to influence his evaluator for career advancement. As shown in Table 2, both the "asymmetry in supervisor assessments" and the "asymmetry in colleague-perceived positiveness" are substantially larger in the subsamples of careerist/slick CGCSs, confirming *Proposition 1.1*.

If the "assessment asymmetries" documented in Table 1 are indeed caused by evaluator-specific influence activities as we interpreted, then they should only exist when the CGCS knows who the "target" is. Specifically, under the masked scheme, when the CGCS no longer knows the identity of the evaluator, there should no longer be any asymmetry in supervisor assessments and colleague perceptions. Therefore, as a placebo test, in Table 3, we focus on the masked scheme where the randomly chosen evaluator's identity was not announced until the end of the evaluation cycle. As we can see in the first two columns, in the masked scheme, being selected as the evaluator no longer leads to more positive assessments than the other non-evaluating supervisor. Further, in Columns (3) and (4), we find that in the masked scheme, colleagues are no longer able to identify the evaluator as the more positive supervisor. Together, these results support our interpretation of the "assessment asymmetries" under the revealed scheme.

B. Proposition 2: Improved Performance under Masked Scheme

As suggested by *Proposition 2*, the masked scheme improves CGCS performance. We exploit the random assignment of CGCSs between the two evaluation schemes to causally test this hypothesis.

First, in Table 4, we examine colleague assessments, which we consider to be the ideal measure of CGCS performance in this setting, for reasons explained in Section II. Specifically, the dependent variable is the average colleague assessment of the CGCS's performance, which is framed relative to other civil servants employed in the same work unit. The assessment score in the questionnaire ranges from 1 to 7, representing different categories from "worse than all the colleagues" to "better than all the colleagues". Relatedly, we have an outcome variable indicating whether the colleagues think the CGCS's performance ranks in the top 10% of the organization.

Results in Columns (1)–(2) show that masking the identity of the evaluator significantly improves colleagues' assessments of CGCS performance. In Columns (3)–(4), using "top 10%" as the outcome variable, all the results remain similar: CGCSs in the masked scheme are more likely to be recognized as top performers by their colleagues.

To better interpret the economic significance of our "masked scheme" intervention, we compare the magnitude of the "masking effect" to other CGCS characteristics that strongly predict colleague assessment score. In Appendix Table A7, we report partial correlations

between subjective assessments and a rich set of CGCS characteristics. As can be seen, education and CCP membership have the strongest predictive power on colleague assessments: graduating from a four-year regular college instead of a three-year community college is correlated with a 0.15-point increase in average colleague assessment, and being a party member is associated with a 0.17-point increase in average colleague assessment. The "masking" effect (0.22-point) is therefore larger than the effect of "upgrading" three-year community college graduates to four-year regular college graduates, or replacing non-party members with party-members. Given the substantial edge in average ability associated with four-year colleges and party member status, the return to the masked scheme does appear to be economically significant.

In Table 5, we corroborate the improvement in colleague assessments with supervisor assessments. In Column (1), the outcome variable is the mean assessment of the two supervisors. We find that "masking the identity of the evaluator" significantly improves average supervisor assessment. In Columns (2) to (4), we find that the masked scheme improves mainly the non-evaluating supervisor's assessment, while the evaluating supervisor's assessment remains essentially unchanged, resulting in higher average assessment and lower deviation in assessments. These results, again, are consistent with our conceptual framework, where masking evaluator identity leads to less evaluator-specific influence activities and higher overall productive performance.

In addition to the subjective assessments given by colleagues and supervisors, we also explore other more "revealed-preference" and "objective" performance measures to further support our findings. The results are presented in Table 6.

In the first column, as a "revealed-preference" measure, we directly asked colleagues whether they recommend to the provincial government that the CGCS be promoted to a tenured position in this office after finishing his two-year term. Using this as an alternative outcome, we find that masking evaluator identity makes more colleagues think that the CGCS deserves tenure, again suggesting an improvement in performance.

We also asked each CGCS to report their total monthly remuneration, including basic wages and performance bonuses (if any), which we later verified using administrative information provided by the provincial governments. The basic wage is set by the upper-level government and should be exactly the same for all CGCSs, conditional on the county of residence, enrollment year, and position. In addition to the basic wage, each work unit has some

discretion over a modest amount of performance bonuses to reward the best performing employees. In Columns (3) and (4), we observe that on average, the CGCSs in the masked scheme earn 50-RMB (2.3%) higher salaries than those in the revealed scheme. Since the basic salary for CGCSs is fixed (matched to the entry-level permanent civil servant wage), this income gap reflects the difference in performance bonus.

During our field interviews, we were informed that the CGCSs who work as nurses in township clinics enjoy the most substantial performance bonuses, because these clinics have a "business" feature and can keep some profits to reward the most hard-working staff. For nurses, the most important factor determining their bonus differentials is the number of night shifts taken each month: every additional night shift is rewarded by about 20 RMB (about $3). In Columns (5) and (6), when we restrict the sample to CGCSs working as nurses, we find a larger than 110-RMB (6.2%) income gap between the two schemes. The compensation differential between the "revealed" and "masked" groups is therefore equivalent to nearly six additional night shifts per month. This result suggests that the performance improvement caused by the masked scheme is indeed substantial when benchmarked objectively.

As reflected by various measures, the evidence consistently suggests that CGCS performance improved in an economically significant manner under the masked scheme, confirming *Proposition 2* of our model.

C. Mechansisms behind the Masked Effects

In our model, the masked scheme improves CGCS performance by incentivizing the reallocation of efforts from influence activities toward productive tasks. Therefore, given the significant performance improvement documented in the previous subsection, we expect to also observe decreased influence activities and increased work efforts for CGCSs assigned to the masked scheme. We test these predictions in Table 7.

In the end-line survey, we asked each CGCS, "what was the most challenging part of your CGCS experience?"[16] As shown in Column (1), CGCSs under the revealed scheme were significantly more likely to report "handling the personal relationship with supervisors" as the

[16] The choices included "familiarizing myself with the local governance system," "handling the personal relationship with my supervisor," "handling personal relationships with my colleagues," "adjusting to life in the rural area," "working on tasks unrelated to my college major," "adjusting to unfamiliar work and life conditions," "getting useful work feedback," and "other challenges."

most challenging part of their experience, as compared to their peers who were randomized into the masked scheme. In contrast, as can be seen from Column (2), the proportion of CGCSs answering "handling personal relationships with colleagues" as the most challenging part of the experience is the same across the two schemes. These two results are consistent with our model, in which the CGCS engages in more influence activities under the revealed scheme than the masked scheme, and does not have incentives to influence his colleagues under either scheme.[17]

In Column (3), we provide evidence that CGCSs under the masked scheme indeed work harder. Specifically, we asked the colleagues whether they thought the CGCS was hardworking and frequently worked overtime. The result in Column (3) shows that the colleagues are significantly more likely to assess the CGCS as "hardworking and frequently work overtime" under the masked scheme, confirming our model prediction of increased work efforts.

Moreover, as can be seen in Column (4), when asked "whether you think the civil service system is meritocratic," CGCSs under the masked scheme are significantly more likely to give a positive answer. This likely reflects the increased return to productive efforts (relative to influence activities) under the masked scheme.

Combined, the results in Table 7 indicate that reallocating efforts from influence activities to productive efforts is indeed the driving force behind the improvement in CGCS performance, providing further support for *Proposition 2*. We address remaining alternative interpretations in Section V(B).

D. Heterogeneous Effects of Masking Evaluator Identity

Our model provides a series of predictions regarding heterogeneous returns to the masked scheme. In Table 8, we confront each of these hypotheses.

First, *Proposition 2.1* suggests that if a CGCS is more risk-averse, he should respond more strongly to the introduction of uncertainty in evaluator identity, and thus achieve a greater improvement in performance. In Column (1) of Table 8, we elicit the level of risk aversion of

[17] Relatedly, as shown in Appendix Table A6, the "unexplained positiveness in evaluator assessment," defined as the residual obtained from regressing "evaluator assessment" on "non-evaluator assessment" and "colleague assessments," is significantly lower in the masked scheme. This is also consistent with our interpretation that the CGCS engages less in influence activities under the masked scheme. For this result to be confounded, an alternative interpretation would need to explain why the evaluating supervisor can appreciate the achievements of the CGCS better than either the non-evaluator or any colleague does, and why such appreciation only exists in the revealed scheme.

each CGCS and interact this measure with the dummy for the masked scheme.[18] As can be seen, the "masking effect" is indeed significantly stronger for the more risk-averse CGCSs, consistent with the theoretical prediction.

Second, *Proposition 2.2* indicates that when the preferences of the two supervisors are more aligned, the masked scheme should have a more powerful effect. The intuition is that, when we switch from the revealed scheme to the masked scheme, the relative return to productive efforts (as compared to supervisor-specific influence activities) would increase only if some of the productive efforts can be appreciated by both supervisors. Therefore, the more aligned the two supervisors' preferences are, the more effective the masked scheme will be. In our survey, we separately elicited each supervisor's subjective ranking of the importance of each performance dimension. In Column (2) of Table 8, we find suggestive evidence consistent with this hypothesis, but the statistical test is slightly under-powered. Nevertheless, this pattern can help us rule out the alternative interpretation that the treatment effect of the masked scheme is driven by the CGCS responding to more diversified supervisor preferences.[19]

Third, in *Proposition 2.3*, we hypothesize that when the CGCS's productive work does not get full credit from both supervisors, an increase in "credit discrepancy" between the two supervisors improves CGCS performance in the revealed scheme, but not in the masked scheme. The reason is that a risk-neutral CGCS would choose his effort level based on the expected level of credit given to his productive performance. In the revealed scheme, he responds to either "high-credit" or "low-credit", while in the masked scheme, he responds to the average level of credit. Since the production function for performance is concave, the dispersion of credit levels generates better performance in the revealed scheme. To examine this prediction, in Column (3), we use the "difference in work assignment frequencies between the two supervisors" to measure their differential acknowledgments of CGCS performance. As can be seen, when the two supervisors have a larger gap in task assignment frequencies,

[18] Risk attitude is elicited through a hypothetical coin-flipping game. An individual preferring a certain 400 Yuan over a 50% chance 2000 Yuan is defined as highly risk averse. Our results are robust to alternative cutoffs for this definition.

[19] The alternative mechanism is that, under the revealed scheme, the CGCS responds to only one supervisor; under the masked scheme, he responds to both supervisors. When the two supervisors have different weights for different dimensions of productive performance, the masked scheme might cause the CGCS to work on a wider range of tasks, which might improve his overall performance. However, if this interpretation is true, the masked scheme should be even more effective when the two supervisors have more heterogeneous preferences along different dimensions, which is inconsistent with the results in Column (2).

the CGCS performs better in the revealed scheme, but not in the masked scheme, confirming our theoretical prediction.

Fourth, in *Proposition 2.4,* we predict that when the information gap (regarding CGCS performance) between the two supervisors is larger, the masked scheme becomes less effective in improving CGCS performance. The mechanism is similar to the "credit discrepancy" channel explained above: in the revealed scheme, the CGCS responds to either "high information" or "low information," while in the masked scheme, he responds to the average information level of the two supervisors. The concavity of the performance production function would thus indicate that information asymmetry generates better performance in the revealed scheme. This proposition is confirmed by Column (4), where we use the difference in the two supervisors' self-reported familiarity with the CGCS's performance to measure information asymmetry and find that higher information asymmetry leads to better CGCS performance in the revealed scheme, but not in the masked scheme.

V. Alternative Interpretations

Our model suggests that, under the revealed scheme, an agent engages in evaluator-specific influence activities; and under the masked scheme, the agent no longer knows who is evaluating, so would re-optimize efforts to work harder on the productive dimensions that would be appreciated by both supervisors. In Section IV, we presented evidence supporting our theoretical propositions. In this section, we discuss several alternative explanations for our empirical results and test their validity using additional data.

A. Alternative Interpretations of Asymmetric Supervisor Assessments under the Revealed Scheme

Our interpretation of the findings in Table 1 is based on *Proposition 1*: in the revealed scheme, the CGCS is able to perform evaluator-specific influence activities, and such behaviors can be observed by their co-workers. There are two potential confounding explanations.

1. Behavioral Differences between the Evaluating and Non-Evaluating Supervisors

In the revealed scheme, the evaluating and non-evaluating supervisors may act differently, simply because revealing the identity of the evaluator might directly affect the evaluator's behavior. For example, if a supervisor knows that he is the evaluator, he may follow the CGCS's work more closely throughout the year, assign more job tasks, or simply feel pressured

to give more positive assessments given the associated stakes. These concerns are partly alleviated by our experimental design, as supervisors were not informed by the research team about their roles in the evaluation. However, it is still possible that the CGCS might have delivered this information to his evaluator.

To investigate this possibility, in our end-line survey, we directly asked each evaluator whether he was aware of his responsibility in evaluating the CGCS. It turns out that the majority of them (more than 65%) did not know whether they were chosen as evaluators until after they had finished the evaluations. In Panel A of Table 9, we re-estimate the specification in Table 1 separately for two subsamples: the subsample in which supervisors did not know their (evaluator) roles, and the subsample in which supervisors knew their evaluator roles (most likely through the CGCS). We find that the asymmetry in supervisor assessments is almost identical in the two subsamples, suggesting that our results are not driven by supervisor behavioral changes due to being the evaluator.

Moreover, in Panel B of Table 9, we directly examine the existence of behavioral differences between the evaluating and non-evaluating supervisors. We focus on three outcomes: (1) the likelihood of a supervisor not responding to our endline survey; (2) whether one supervisor writes more words than the other supervisor in describing the CGCS's job tasks; and (3) whether one supervisor assigns more job tasks to the CGCS than the other. Our hypothesis is that, if the evaluating supervisor indeed paid more attention to the CGCS's performance, we should observe the evaluating supervisor being more likely to answer the survey, write more words in his assessments, and perhaps assign more tasks to the CGCS. Again, the data does not support this interpretation, as evidenced by regression results in Panel B.

2. More Information for Evaluating-Supervisor

Another confounding story is that, even without behavioral changes, the evaluating supervisor would receive more information regarding CGCS performance from various sources: the CGCS, the colleagues, and the other (non-evaluating) supervisor might all try to send signals to help him better evaluate. This increase in information might improve the evaluator's assessment and thus create the scoring asymmetry shown in Table 1.

To examine this interpretation, in our endline survey, we asked each supervisor "how frequently did the CGCS, the colleagues of the CGCS, or the other supervisor discuss the CGCS's performance with you?" We are interested in whether the evaluating supervisor would

receive more information than the non-evaluating supervisor from these three sources. In Panel C of Table 9, we find that the evaluator did not gain extra information from any of these sources, as compared to the non-evaluator.[20] Therefore, the asymmetry in supervisor assessments under the revealed scheme cannot be explained by the difference in information between the two supervisors.

B. Alternative Interpretations of Improved Assessments under the Masked Scheme

Our interpretation of the "improved colleague and supervisor assessments" under the masked scheme is based on *Proposition 2*: masking evaluator identity makes supervisor-specific influence activities less beneficial, which incentivizes the CGCSs to work harder on productive dimensions that are appreciated by both supervisors, resulting in better work performance. There are four potential confounding explanations.

1. CGCS Influencing Both Supervisors More

The first alternative interpretation is that, under the masked scheme, the CGCS does not work harder on productive dimensions. Instead, he simply extends more influence activities toward both supervisors, which is why we see improved average supervisor assessment. However, this interpretation is inconsistent with a series of empirical results.

First, it is inconsistent with the fact that colleague assessments improved substantially under the masked scheme. As explained in Section II, the CGCS has no incentive to influence his colleagues; every CGCS is clearly informed that only his evaluating supervisor's opinion will be taken into account by the provincial government, and colleague assessments will never enter into their promotion functions. Therefore, if the CGCS is simply extending more influence activities toward both supervisors, rather than working harder, there should not be a significant improvement in average colleague assessment.

Second, if the CGCS is engaging in more influence activities instead of working harder, we also should not observe objective performance improvements under the masked scheme. As discussed in Section IV, CGCSs under the masked scheme receive substantially higher performance bonuses, which are directly linked to objective performance indicators. This, again, supports our interpretation and contradicts the competing hypothesis.

[20] If anything, the evaluator is 3% less likely to receive information regarding CGCS performance from colleagues, although the coefficient is small in magnitude and only marginally significant.

Third, as documented in Table 7, under the masked scheme, the CGCSs are less worried about handling personal relationships with supervisors, as compared to their peers under the revealed scheme. This also suggests a reduction in total influence activities under the masked scheme, rather than an increase in influence activities targeting both supervisors.

Fourth, the alternative interpretation is also at odds with the heterogeneity results presented in Table 8. If the CGCS is indeed trying to undertake more influence activities instead of working more on productive dimensions, then whether the two supervisors have aligned preferences along the productive dimensions, and whether they have comparable information or similar weights regarding productive performance, should not affect the effectiveness of the masked scheme. The heterogeneity results with respect to these characteristics of the productive dimensions are therefore consistent with our theoretical predictions and inconsistent with the "increased influence activities" interpretation.

2. CGCS Influencing Colleagues under the Masked Scheme

Suppose that the CGCSs, for whatever reason, are trying to influence their colleagues, and they do so to a larger extent under the masked scheme. Could this be confounding our results on improved colleague assessments under the masked scheme?

This interpretation is inconsistent with the result in Table 7 Column (2), which shows that the proportion of CGCSs worrying about "handling personal relationships with colleagues" remains the same across both schemes. In this subsection, we conduct two additional placebo tests to further rule out this confounding interpretation.

In the first placebo test, we hypothesize that under the masked scheme, any additional influence activities toward colleagues will result in more interactions between the CGCSs and their colleagues, especially on non-professional occasions. However, in Panel A of Table 10, we find that, according to colleagues, the masked scheme did not make the CGCSs communicate more frequently with them (Columns (1) and (2)), nor did it make them more familiar with the CGCS's work or personal life (Columns (3) and (4)). These results are inconsistent with the alternative interpretation.

Moreover, we conduct a second placebo test, which indirectly tests the change in influence activities by comparing the magnitudes of "hometown bias" between the revealed and masked schemes, for supervisors and colleagues respectively. Specifically, "hometown bias," defined as the gap in assessments between a "same-hometown evaluator" and a "different-hometown

evaluator," consists of two parts: (1) top-down preference, meaning that an evaluator would assess a same-hometown CGCS more positively; and (2) bottom-up influence, meaning that a CGCS would find it easier to influence an evaluator from the same hometown. While "top-down preference" should remain the same across different evaluation schemes, "bottom-up influence" would change with respect to the amount of influence activities extended by the CGCS.

According to our model, the CGCS reduces influence activities toward his supervisors under the masked scheme, and does not influence his colleagues in either scheme. If this interpretation is correct, if influence activities were more effective toward a same-hometown supervisor/colleague, we should expect the hometown bias in evaluator assessment to be smaller in the masked scheme than the revealed scheme, while the hometown bias in colleague assessment would remain the same across both schemes.

In contrast, if the alternative interpretation of "influencing colleagues" is correct, and the treatment effect of the masked scheme is confounded by the CGCS increasing his influence activities toward both supervisors and colleagues under the masked scheme, we should expect to see both the "evaluator hometown bias" and the "colleague hometown bias" become larger under the masked scheme than the revealed scheme.

As shown in Table 11, the data strongly support our preferred interpretation over the alternative one: a "same hometown evaluator" is significantly more positive under the revealed scheme, while a "same hometown colleague" does not show differential positiveness across the two schemes.[21] Again, these results suggest that the CGCSs do not engage in additional influence activities toward supervisors or colleagues under the masked scheme; thus, increased influence activities under the masked scheme cannot confound the main findings in this paper.

3. Higher Information Quality in the Masked Scheme

Another possibility is that supervisors in the masked scheme tend to get better information on CGCS performance, which might explain the increase in average supervisor assessment.

To address this concern, in Panel B of Table 10, we examine whether supervisors get additional information on CGCS performance under the masked scheme, either from colleagues or from the CGCS himself. We find that, for both the evaluating supervisor and the non-evaluating supervisor, being in the masked scheme does not increase the frequency

[21] The sample size is smaller due to missing values for supervisors' hometown.

of CGCSs and other colleagues reporting to them regarding CGCS performance. This suggests that improved supervisor assessments in the masked scheme cannot be explained by changes in information quality.

4. CGCS Gets Discouraged when Matched to "Hostile Evaluator" under Revealed Scheme

A remaining possibility is that, under the revealed scheme, some CGCSs might be matched with an evaluator whom they perceive as hostile, in that, no matter how hard one works, efforts will not be appreciated by this evaluator. As a result, the CGCSs get discouraged and put little effort into the productive dimensions, which might explain why performance is higher under the masked scheme.

In our baseline survey, before the randomizations of schemes and evaluators were realized, we asked each CGCS "among the two supervisors, whom would you prefer to be your evaluator?" Due to randomization, half of the CGCSs under the revealed scheme would be evaluated by their "non-preferred" supervisor and the other half evaluated by their "preferred" supervisor. Since the "discouragement" mechanism should operate only through those evaluated by the non-preferred supervisor, we can compare performance differences between CGCSs facing preferred supervisor under the revealed scheme and those under the masked scheme. Were it true that the "discouragement" mechanism is driving the observed improvement in CGCS performance, we should expect the performance improvement under the masked scheme to disappear in such a restricted comparison. However, as shown in Appendix Table A8, the masking effect is actually slightly stronger, rather than weaker, in this subsample analysis, providing strong evidence against the "discouragement" interpretation.

VI. Conclusion

Subjective evaluations are widely used in both private and public sectors, especially in contexts where job tasks are inherently multi-dimensional and vaguely defined, making it impossible to obtain sharp measures of employee effort and performance. A key limitation to subjective evaluation is that it may distort the employee's incentives and make him more likely to cater to the evaluator's personal interest, rather than focusing on productive tasks that benefit the whole organization. However, rigorous empirical evidence on the existence and implications of influence activities remains scarce.

To shed light on this topic, we conducted a large-scale field experiment, where we randomized two subjective performance evaluation schemes among 3,785 junior state employees in China. In the "revealed" scheme, we randomly chose one of the two supervisors as the performance evaluator, and informed the subordinate *ex-ante* about the evaluator's identity. We find that under this scheme, subordinates are indeed induced to engage in evaluator-specific influence activities to improve their evaluation outcomes. We also find that other colleagues can correctly predict that the randomly selected evaluator would better appreciate the subordinate (as compared to the non-evaluating supervisor), suggesting that the subordinate's evaluator-specific influence activities can be observed by fellow co-workers.

In the "masked" scheme, we also randomly chose one of the two supervisors as the performance evaluator, but the identity of the evaluator was not disclosed to the subordinate, which reduces the expected return to supervisor-specific influence activities. Therefore, masking evaluator identity should encourage the subordinate to reallocate his efforts from influence activities toward common productive dimensions that could be appreciated by both supervisors. We find that this evaluation scheme indeed improves the subordinate's work performance, as measured by average colleague assessments, average supervisor assessments, likelihood of being recommended for "tenure," and monthly bonus payments determined by objective performance indicators.

In addition to providing the first rigorous empirical evidence on the existence and implications of influence activities, our findings have important policy implications. We find that by randomizing the evaluator's identity, which has minimal implementation cost, the government can significantly improve the job performance of its employees. These findings not only have direct policy implications for the more than 50 million state employees in China, but might also be relevant for many other contexts where high-stakes rewards are linked to the subjective opinions of designated evaluators.

REFERENCES

Ashraf, Nava, and Oriana Bandiera. "Social incentives in organizations." *Annual Review of Economics* 10 (2018): 439-463.

Ashraf, Nava, Oriana Bandiera, Edward Davenport, and Scott S. Lee. "Losing prosociality in the quest for talent? Sorting, selection, and productivity in the delivery of public services." *American Economic Review* 110, no. 5 (2020): 1355-94.

Ashraf, Nava, Oriana Bandiera, and B. Kelsey Jack. "No margin, no mission? A field experiment on incentives for public service delivery." *Journal of Public Economics* 120 (2014): 1-17.

Baker, George, Robert Gibbons, and Kevin J. Murphy. "Subjective performance measures in optimal incentive contracts." *The Quarterly Journal of Economics* 109, no. 4 (1994): 1125-1156.

Banerjee, Abhijit V., Raghabendra Chattopadhyay, Esther Duflo, Daniel Keniston, and Nina Singh. "Can institutions be reformed from within? Evidence from a randomized experiment with the Rajasthan police." (2012).

Bertrand, Marianne, Robin Burgess, Arunish Chawla, and Guo Xu. "The glittering prizes: Career incentives and bureaucrat performance." *The Review of Economic Studies* (2018).

Chevalier, Judith, and Glenn Ellison. "Career concerns of mutual fund managers." *The Quarterly Journal of Economics* 114, no. 2 (1999): 389-432.

Deb, Joyee, Jin Li, and Arijit Mukherjee. "Relational contracts with subjective peer evaluations." *The RAND Journal of Economics* 47, no. 1 (2016): 3-28.

Finan, Frederico, Benjamin A. Olken, and Rohini Pande. "The personnel economics of the state." No. w21825. National Bureau of Economic Research, 2015.

Gibbons, Robert, and Kevin J. Murphy. "Optimal incentive contracts in the presence of career concerns: Theory and evidence." *Journal of Political Economy* 100, no. 3 (1992): 468-505.

Hayes, Rachel M., and Scott Schaefer. "Implicit contracts and the explanatory power of top executive compensation for future performance." *The RAND Journal of Economics* (2000): 273-293.

He, Guojun, Shaoda Wang, and Bing Zhang. "Watering down environmental regulation in China." *The Quarterly Journal of Economics* 135, no. 4 (2020): 2135-2185.

He, Guojun, and Shaoda Wang. "Do college graduates serving as village officials help rural China?" *American Economic Journal: Applied Economics* 9, no. 4 (2017): 186-215.

Hetts, John J., Michiko Sakuma, and Brett W. Pelham. "Two roads to positive regard: Implicit and explicit self-evaluation and culture." *Journal of Experimental Social Psychology* 35, no. 6 (1999): 512-559.

Lazear, Edward, and Paul Oyer. *The Handbook of Organizational Economics*. Gibbons, Robert, and John Roberts, eds. Princeton University Press, 2012.

Li, Hongbin, and Li-An Zhou. "Political turnover and economic performance: the incentive role of personnel control in China." *Journal of Public Economics* 89, no. 9-10 (2005): 1743-1762.

Lü, Xiaobo, and Pierre F. Landry. "Show me the money: Interjurisdiction political competition and fiscal extraction in China." *American Political Science Review* (2014): 706-722.

MacLeod, W. Bentley. "Optimal contracting with subjective evaluation." *American Economic Review* 93, no. 1 (2003): 216-240.

Maestri, Lucas. "Bonus payments versus efficiency wages in the repeated principal-agent model with subjective evaluations." *American Economic Journal: Microeconomics* 4, no. 3 (2012): 34-56.

Meyer, Margaret, Paul Milgrom, and John Roberts. "Organizational prospects, influence costs, and ownership changes." *Journal of Economics & Management Strategy* 1, no. 1 (1992): 9-35.

Milgrom, Paul R. "Employment contracts, influence activities, and efficient organization design." *Journal of Political Economy* 96, no. 1 (1988): 42-60.

Milgrom, Paul, and John Roberts. "An economic approach to influence activities in organizations." *American Journal of Sociology* 94 (1988): S154-S179.

Olken, Benjamin A., Rohini Pande, and Raluca Dragusanu. "Governance review paper: J-PAL governance initiative." *Rep., Abdul Latif Jameel Poverty Action Lab, Cambridge, MA*(2011).

Oyer, P., and S. Schaefer. "Personnel Economics: Hiring and Incentives." *Handbook of Labor Economics*. (2011): 1769-1823.

Powell, Michael. "An influence-cost model of organizational practices and firm boundaries." *The Journal of Law, Economics, and Organization* 31, no. suppl_1 (2015): i104-i142.

Prendergast, Canice. "The provision of incentives in firms." *Journal of Economic Literature* 37, no. 1 (1999): 7-63.

Rasul, Imran, and Daniel Rogger. "Management of bureaucrats and public service delivery: Evidence from the Nigerian civil service." *The Economic Journal* 128, no. 608 (2018): 413-446.

Schaefer, Scott. "The Dependence of Pay-Performance Sensitivity on the Size of the Firm." *Review of Economics and Statistics* 80, no. 3 (1998): 436-443.

Serrato, Juan Carlos Suárez, Xiao Yu Wang, and Shuang Zhang. "The limits of meritocracy: screening bureaucrats under imperfect verifiability." *Journal of Development Economics* 140 (2019): 223-241.

Shirk, Susan L. *The Political Logic of Economic Reform in China*. Vol. 24. Univ of California Press, 1993.

Figure1. Experimental Design

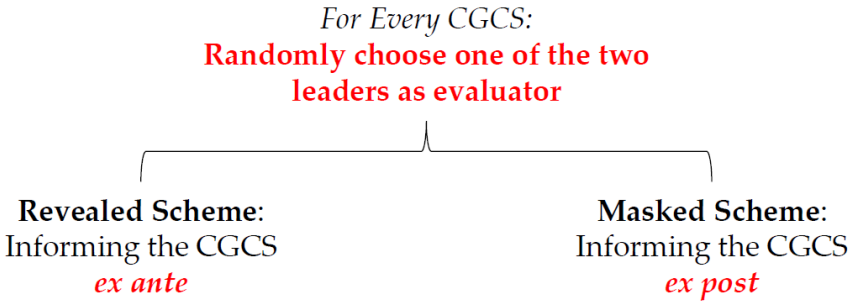

Table 1. Scoring Asymmetry When the Evaluator's Identity is Revealed

	Supervisor 1's Score Minus Supervisor 2's Score		Colleagues Speculate Sup. 1 Gives a Higher Score	
	(1)	(2)	(3)	(4)
Supervisor 1 Evaluating (Revealed)	0.310***	0.310***	0.058***	0.058***
	(0.082)	(0.082)	(0.020)	(0.020)
Control Mean	-0.142	-0.142	0.541	0.541
Control S.D.	1.273	1.273	0.498	0.498
Controls	N	Y	N	Y
County FE	Y	Y	Y	Y
Type FE	Y	Y	Y	Y
Cohort FE	Y	Y	Y	Y
Obs.	1,301	1,301	5,582	5,582
R-Squared	0.160	0.163	0.108	0.111

Notes: Each column represents a separate regression. The sample is for the *revealed* scheme only. Standard errors clustered at the work unit level are reported below the coefficients. Control variables include CGCS and colleague characteristics listed in Appendix Tables A1-A2. * significant at 10% ** significant at 5% *** significant at 1%.

Table 2. Personality Traits and Influence Activities in the Revealed Scheme

	Supervisor 1's Score Minus Supervisor 2's Score		Colleagues Speculate Sup. 1 Gives a Higher Score	
	(1)	(2)	(3)	(4)
Panel A. Careerist vs. Altruistic				
Supervisor 1 Evaluating	0.555***	0.261***	0.098*	0.051**
(Revealed)	(0.194)	(0.095)	(0.053)	(0.022)
Sample	Careerist	Altruistic	Careerist	Altruistic
Controls	Y	Y	Y	Y
County FE	Y	Y	Y	Y
Type FE	Y	Y	Y	Y
Cohort FE	Y	Y	Y	Y
Obs.	229	1,006	1,244	4,322
R-Squared	0.383	0.195	0.281	0.136
Panel B. Slick vs. Candid				
Supervisor 1 Evaluating	0.420***	0.322***	0.086***	0.045
(Revealed)	(0.121)	(0.122)	(0.028)	(0.030)
Sample	Slick	Candid	Slick	Candid
Controls	Y	Y	Y	Y
County FE	Y	Y	Y	Y
Type FE	Y	Y	Y	Y
Cohort FE	Y	Y	Y	Y
Obs.	681	547	2,971	2,595
R-Squared	0.231	0.281	0.164	0.202

Notes: Each column represents a separate regression. The sample is for the *revealed* scheme only. Standard errors clustered at the work unit level are reported below the coefficients. Control variables include CGCS and colleague characteristics listed in Appendix Tables A1-A2. * significant at 10% ** significant at 5% *** significant at 1%.

Table 3. Masking Evaluator's Identity Eliminates Scoring Asymmetry

	Supervisor 1's Score Minus Supervisor 2's Score		Colleagues Speculate Sup.1 Gives a Higher Score	
	(1)	(2)	(3)	(4)
Supervisor 1 Evaluating	-0.003	0.001	0.007	0.012
(Masked)	(0.113)	(0.116)	(0.019)	(0.019)
Control Mean	0.162	0.162	0.523	0.523
Control S.D.	1.193	1.193	0.499	0.499
Controls	N	Y	N	Y
County FE	Y	Y	Y	Y
Type FE	Y	Y	Y	Y
Cohort FE	Y	Y	Y	Y
Obs.	580	580	2,615	2,615
R-Squared	0.242	0.247	0.175	0.183

Notes: Each column represents a separate regression. The sample is for the *masked* scheme only. Standard errors clustered at the work unit level are reported below the coefficients. Control variables include CGCS and colleague characteristics listed in Appendix Tables A1-A2. * significant at 10% ** significant at 5% *** significant at 1%.

Table 4. Treatment Effects on Colleague Evaluations

	Performance (1-7)		Top 10%	
	(1)	(2)	(3)	(4)
Masked	0.220***	0.186***	0.081***	0.070***
	(0.033)	(0.027)	(0.012)	(0.011)
Control Mean	5.127	5.127	0.670	0.670
Control S.D.	1.231	1.231	0.470	0.470
Controls	N	Y	N	Y
County FE	Y	Y	Y	Y
Type FE	Y	Y	Y	Y
Cohort FE	Y	Y	Y	Y
Obs.	9,256	9,167	9,256	9,167
R-Squared	0.130	0.371	0.084	0.268

Notes: Each column represents a separate regression. Standard errors clustered at the work unit level are reported below the coefficients. Control variables include CGCS and colleague characteristics listed in Appendix Tables A1-A2. * significant at 10% ** significant at 5% *** significant at 1%.

Table 5. Treatment Effects on Supervisor Assessments

	Mean (Supervisor Assessment)	Evaluator Assessment	Non-Evaluator Assessment	Supervisors' Assessment Deviation
	(1)	(2)	(3)	(4)
Masked	0.140***	0.069	0.210***	-0.097*
	(0.046)	(0.057)	(0.057)	(0.050)
Control Mean	5.098	5.155	5.041	0.937
Control S.D.	0.900	1.145	1.091	0.947
Controls	Y	Y	Y	Y
County FE	Y	Y	Y	Y
Type FE	Y	Y	Y	Y
Cohort FE	Y	Y	Y	Y
Obs.	1,945	1,940	1,940	1,940
R-Squared	0.257	0.214	0.216	0.134

Notes: Each column represents a separate regression. Control variables include CGCS characteristics listed in Appendix Table A1. Standard errors clustered at the work unit level are reported below the coefficients. * significant at 10% ** significant at 5% *** significant at 1%.

Table 6. Treatment Effects on "Revealed Preference" and Objective Measures

	Qualify for Tenure		Wage		Wage (Medical Support)	
	(1)	(2)	(3)	(4)	(5)	(6)
Masked	0.032***	0.024***	48.81**	50.35**	115.54*	110.42*
	(0.009)	(0.009)	(22.41)	(22.91)	(61.94)	(59.13)
Control Mean	0.864	0.864	2103	2103	1852	1852
Control S.D.	0.343	0.343	645	645	349	349
Controls	N	Y	N	Y	N	Y
County FE	Y	Y	Y	Y	Y	Y
Type FE	Y	Y	Y	Y	Y	Y
Cohort FE	Y	Y	Y	Y	Y	Y
Obs.	9,349	9,171	2,750	2,750	193	193
R-Square	0.099	0.131	0.64	0.64	0.74	0.75

Notes: Each column represents a separate regression. Control variables include CGCS and colleague characteristics listed in Appendix Tables A1-A2. Standard errors clustered at the work unit level are reported below the coefficients. * significant at 10% ** significant at 5% *** significant at 1%.

Table 7. Treatment Effects on Influence Activities and Work Efforts

	CGCS Challenge: *Supervisor* Relationship	CGCS Challenge: *Colleague* Relationship	Colleagues: CGCS is *Hardworking*	CGCS Belief: Civil Service is *Meritocratic*
	(1)	(2)	(3)	(4)
Masked	-0.030**	-0.003	0.023**	0.024**
	(0.014)	(0.009)	(0.012)	(0.012)
Control Mean	0.160	0.052	0.446	0.896
Control S.D.	0.367	0.221	0.497	0.306
Controls	Y	Y	Y	Y
County FE	Y	Y	Y	Y
Type FE	Y	Y	Y	Y
Cohort FE	Y	Y	Y	Y
Obs.	2,839	2,839	9,349	2,839
R-Squared	0.110	0.092	0.491	0.075

Notes: Each column represents a separate regression. Standard errors clustered at work unit level are reported below the coefficients. Control variables include CGCS and colleague characteristics listed in Appendix Tables A1-A2. * significant at 10% ** significant at 5% *** significant at 1%.

Table 8. Mechanisms: Risk Aversion and Relative Importance of Supervisors

	Colleague Evaluation (1-7)			
	(1)	(2)	(3)	(4)
Risk Aversion	-0.031			
	(0.032)			
Masked*Risk Aversion	0.093*			
	(0.055)			
Supervisors' Weights Similarity		-0.049		
		(0.071)		
Masked*Weights Similarity		0.117		
		(0.108)		
Δ in Superiors' Work Assign. Freq.			0.003***	
			(0.001)	
Masked*Δ in Work Assign. Freq.			-0.002**	
			(0.001)	
Supervisors' Info. Gap				0.217**
				(0.103)
Masked*Supervisors' Info. Gap				-0.331*
				(0.187)
Masked	0.148***	0.174***	0.204***	0.502***
	(0.039)	(0.031)	(0.030)	(0.175)
Controls	Y	Y	Y	Y
County FE	Y	Y	Y	Y
Type FE	Y	Y	Y	Y
Cohort FE	Y	Y	Y	Y
Obs.	8,859	8,859	8,859	8,859
R-Squared	0.33	0.33	0.33	0.33

Notes: Each column represents a separate regression. Control variables include CGCSs' and colleagues' characteristics listed in Tables A1-A2. Standard errors clustered at the work unit level are reported below the coefficients. * significant at 10% ** significant at 5% *** significant at 1%.

Table 9. Alternative Explanations to Influence Activities in the Revealed Scheme

	(1)	(2)	(3)
Panel A. Does Supervisor Evaluation Depend on their Awareness of their Roles?			
	Supervisor 1 Score Minus Supervisor 2 Score		
Supervisor 1 Evaluating	0.310***	0.334***	0.320*
(*Revealed)*	(0.082)	(0.099)	(0.166)
Sample	Full Sample	Supervisor 1 Unaware of being the Evaluator	Supervisor 1 Aware of Being the Evaluator
Obs.	1,301	888	333
R-Squared	0.160	0.206	0.270
Panel B. Is there Behavioral Changes of the Evaluating Supervisor?			
	Supervisor 1 Not Responding to the Survey	Sup.1 Writes More Words in Describing CGCS's Job	Sup. 1 Assigns More Tasks to the CGCS
Supervisor 1 Evaluating	-0.010	0.649	0.236
(*Revealed)*	(0.019)	(0.431)	(0.181)
Obs.	1,910	1,910	1,910
R-Squared	0.144	0.147	0.144
Panel C. Does the Evaluating Supervisor Receive More Information?			
	Supervisor 1 Gets More Information from CGCS than Supervisor 2 Does	Supervisor 1 Gets More Information from Colleagues than Supervisor 2 Does	Supervisor 1 Gets More Information from Opposing Supervisor than Supervisor 2 Does
Supervisor 1 Evaluating	0.000	-0.031*	0.022
(*Revealed)*	(0.019)	(0.017)	(0.020)
Obs.	1,910	1,910	1,910
R-Squared	0.146	0.162	0.158
County FE	Y	Y	Y
Type FE	Y	Y	Y
Enrol Year FE	Y	Y	Y

Notes: Each column represents a separate regression. Standard errors clustered at the work unit level are reported below the coefficients. * significant at 10% ** significant at 5% *** significant at 1%.

Table 10. Alternative Explanations for Improved Performance under the Masked Scheme

	(1)	(2)	(3)	(4)
Panel A. Do CGCGs Influence All Their Colleagues?				
	Communication with Colleagues	Meeting with Colleagues	Colleagues Familiar with CGCS Work	Colleagues Familiar with CGCS Life
Masked	-0.008	0.013	0.020	0.066
	(0.013)	(0.020)	(0.034)	(0.059)
Obs.	9,272	9,349	9,252	9,244
R-Squared	0.055	0.066	0.066	0.083
Panel B. Does Masking Identity Lead to Information Difference				
	Evaluator Information Across two Schemes		Non-Evaluator Information Across two Schemes	
Masked	0.014	0.010	0.001	-0.021
	(0.020)	(0.016)	(0.019)	(0.016)
Information from	CGCSs	Colleagues	CGCSs	Colleagues
Obs.	2,839	2,839	2,839	2,839
R-Squared	0.134	0.123	0.123	0.121
County FE	Y	Y	Y	Y
Type FE	Y	Y	Y	Y
Cohort FE	Y	Y	Y	Y

Notes: Each column represents a separate regression. Standard errors clustered at the work unit level are reported below the coefficients. * significant at 10% ** significant at 5% *** significant at 1%.

Table 11. Hometown Favoritism and Influence Activities

	Evaluating Supervisor's Score			Colleague Evaluation		
	Full Sample	Revealed Sample	Masked Sample	Full Sample	Revealed Sample	Masked Sample
	(1)	(2)	(3)	(4)	(5)	(6)
Same Home Town	0.117**	0.192***	0.029	0.051*	0.054	0.056
	(0.053)	(0.068)	(0.099)	(0.028)	(0.034)	(0.051)
Control Mean	5.11	5.21	5.06	5.18	5.33	5.11
Control S.D.	1.13	1.10	1.14	1.22	1.17	1.23
County FE	Y	Y	Y	Y	Y	Y
Type FE	Y	Y	Y	Y	Y	Y
Cohort FE	Y	Y	Y	Y	Y	Y
Obs.	2,291	1,549	686	9,252	6,286	2,954
R-Squared	0.189	0.235	0.263	0.326	0.350	0.340

Notes: Each column represents a separate regression. Standard errors clustered at the work unit level are reported below the coefficients. * significant at 10% ** significant at 5% *** significant at 1%.

Appendix to "Performance Evaluation, Influence Activities, and Bureaucratic Work Behavior: Evidence from China"

APPENDIX A

Table A1. Characteristics of CGCSs and Balance Checks

	Mean (Std. Dev.)	Difference between Masked Group and Revealed Group
	(1)	(2)
Age	25.01	0.04
	(1.56)	(0.06)
Gender	0.62	0.01
(=1 if Female)	(0.49)	(0.02)
Social Science Major	0.54	-0.02
(=1 if Yes)	(0.50)	(0.02)
4-Year College or Above	0.76	-0.00
(=1 if Yes)	(0.43)	(0.02)
STEM Students in High School	0.35	-0.01
(=1 if Yes)	(0.48)	(0.02)
Party Member	0.22	-0.00
(=1 if Yes)	(0.41)	(0.02)
Parent Completing College	0.29	-0.00
(=1 if Yes)	(0.45)	(0.02)
Work in Village	0.15	-0.01
(=1 if Yes)	(0.36)	(0.02)
CEE Score	483.50	4.05*
(100 Points)	(59.88)	(2.34)
Risk Averse	0.47	-0.00
(=1 if Yes)	(0.50)	(0.02)
Obs.		2839

Notes: Column (1) summarizes the mean and standard deviation of CGCS characteristics. Column (2) checks the covariate balances between the revealed group and the masked group by regressing each covariate on a dummy variable for "masked scheme," while controlling for CGCS type FE, cohort FE, and county FE. Standard errors clustered at the work unit level are reported in parentheses.

Table A2. Characteristics of CGCSs' Colleagues and Balance Checks

	Mean (Std. Dev.)	Difference between Masked Group and Revealed Group
	(1)	(2)
Colleague Age	34.50	-0.28
	(8.92)	(0.26)
Colleague Gender	0.57	-0.01
(=1 if Female)	(0.50)	(0.01)
Colleague Education	3.46	-0.02
	(0.71)	(0.02)
Colleague Tenured	0.74	0.00
	(0.44)	(0.01)
Meet Frequency with CGCS	4.75	0.01
Weekly	(0.72)	(0.02)
Know CGCS Well (Work)	9.28	0.02
(0-10)	(1.25)	(0.03)
Know CGCS Well (Life)	8.33	0.07
(0-10)	(2.03)	(0.06)
Colleague Self Assessment	4.46	0.05*
(1-7)	(1.21)	(0.03)
Obs.		9349

Notes: Column (1) summarizes the mean and standard deviation of colleagues' characteristics. Column (2) checks the covariate balances between the revealed group and the masked group group by regressing each covariate on a dummy variable for "masked scheme," while controlling for CGCS type FE, cohort FE, and county FE. Education is measured by a categorical variable (primary school =1, junior high =2, senior high=3, 3-year college =4, 4-year college =5, graduate school=6). Standard errors clustered at work unit level are reported in parentheses.

Table A3. Characteristics of Supervisors and Balance Checks

	Mean (S.D.)	Difference between Masked Group and Revealed Group
	(1)	(2)
Evaluator Gender	0.21	-0.02
(=1 if Female)	(0.41)	(0.02)
Evaluator Age	45.11	-0.29
	(7.15)	(0.35)
Evaluator Work Experience	7.60	0.00
(Years)	(3.25)	(0.17)
Evaluator Education	4.68	0.03
	(0.61)	(0.03)
Evaluator Title	0.42	-0.04
(=0 if Party, =1 if Admin)	(0.49)	(0.03)
Non-Evaluator Gender	0.25	-0.01
(=1 if Female)	(0.43)	(0.02)
Non-Evaluator Age	43.63	-0.09
	(7.83)	(0.36)
Non-Evaluator Work Experience	7.25	-0.09
(Years)	(3.39)	(0.36)
Non-Evaluator Education	4.60	-0.02
	(0.63)	(0.03)
Non-Evaluator Title	0.58	-0.01
(=0 if Party, =1 if Admin)	(0.49)	(0.03)
Obs.		2227

Notes: Column (1) summarizes the mean and standard deviation of supervisors' characteristics. Column (2) checks the covariate balances between the revealed group and the masked group. In the revealed group, the identity of the randomly-chosen evaluator is disclosed to the CGCS *ex-ante*, while in the masked group, the evaluator is chosen randomly *ex-post*. Education is measured by a categorical variable (primary school =1, junior high =2, senior high=3, 3-year college =4, 4-year college =5, graduate school=6). Standard errors clustered at the work unit level are reported in parentheses. Data are collected by the research team.

Table A4. Evaluator and Non-Evaluator Characteristics under the Revealed Scheme

	Evaluator (Revealed)	Non-Evaluator (Revealed)	Difference
	(1)	(2)	(3)
Gender	0.240	0.234	-0.017
(=1 if Female)	(0.427)	(0.423)	(0.024)
Age	43.327	43.438	0.220
	(7.990)	(7.544)	(0.415)
Work Experience	6.822	7.085	0.340*
(Years)	(3.396)	(3.350)	(0.205)
Education	4.690	4.656	-0.035
	(0.601)	(0.601)	(0.034)
Title	0.583	0.550	-0.176
(=0 if Party, =1 if Admin)	(0.493)	(0.498)	(0.225)
Obs.	1,935	1,935	3870

Notes: We keep the subsample of all CGCS supervisors under the revealed scheme. Column (1) summarizes the mean and standard deviation of evaluating supervisors' characteristics. Column (2) summarizes the mean and standard deviation of non-evaluating supervisors' characteristics. Column (3) checks the covariate balances between the two groups controlling for CGCS FE. Education is measured by a categorical variable (primary school =1, junior high =2, senior high=3, 3-year college =4, 4-year college =5, graduate school=6). Standard errors clustered at the work unit level are reported in parentheses.

Table A5. Test for Attrition Selection

	Attrition			
	(1)	(2)	(3)	(4)
Masking Evaluator's Identity	-0.015	-0.010	-0.010	-0.010
	(0.016)	(0.015)	(0.015)	(0.015)
County FE	N	Y	Y	Y
Type FE	N	N	Y	Y
Enroll Year FE	N	N	N	Y
Obs.	3,785	3,773	3,773	3,773
R-Squared	0.000	0.141	0.145	0.145

Notes: Each column represents a separate regression. Standard errors clustered at the work unit level are reported below the coefficients. * significant at 10% ** significant at 5% *** significant at 1%.

Table A6. CGCSs Increase Efforts under the Masked Scheme

	Residualized Supervisor Positiveness	
	(1)	(2)
Masked	-0.104**	-0.098**
	(0.042)	(0.043)
Controls	N	Y
County FE	Y	Y
Type FE	Y	Y
Cohort FE	Y	Y
Obs.	2,037	1,935
R-Squared	0.145	0.150

Notes: Each column represents a separate regression. "Residualized Supervisor Positiveness" is the residual obtained from regressing evaluator assessment on non-evaluator assessment and colleague assessments. Control variables include CGCS characteristics listed in Appendix Table A1. Standard errors clustered at the work unit level are reported below the coefficients. * significant at 10% ** significant at 5% *** significant at 1%.

Table A7. Partial Correlations between CGCS Characteristics and Performance

	Performance (1-7)	
	by Colleague	Supervisor
	(1)	(2)
Age	0.074***	0.074***
	(0.010)	(0.016)
Gender	-0.055	-0.085*
	(0.040)	(0.049)
Social Science	-0.018	-0.028
	(0.036)	(0.043)
4-Year College	0.222***	0.260***
	(0.041)	(0.048)
STEM Students	-0.028	0.051
	(0.037)	(0.044)
Party Member	0.256***	0.250***
	(0.042)	(0.053)
Parent High Sch.	0.038	0.102**
	(0.035)	(0.044)
Parent College	-0.037	0.059
	(0.040)	(0.048)
Work in Village	0.042	0.154**
	(0.059)	(0.061)
CEE Score	0.039	0.100***
	(0.027)	(0.034)
Risk Averse	-0.033	-0.036
	(0.031)	(0.046)

Notes: Each cell represents a separate regression between the outcome variable and the CGCS's certain characteristics. No control is included in any of these partial regressions. In Column (1), the outcome variable is the colleague assessments of CGCS performance; in Column (2), the outcome variable is the supervisor assessments of CGCS performance. Standard errors clustered at the work unit level are reported in parentheses. * significant at 10% ** significant at 5% *** significant at 1%.

Table A8: Discouragement Effect

	Colleague Assessment Score	
	Full Sample	Masking vs. Being Evaluated by Preferred Leader
	(1)	(2)
Masking	0.220***	0.247***
	(0.033)	(0.038)
County FE	Y	Y
Type FE	Y	Y
Enrol Year FE	Y	Y
Obs.	9,256	6,206
R-Squared	0.130	0.158

Notes: Each column represents a separate regression. Standard errors clustered at the work unit level are reported below the coefficients. * significant at 10% ** significant at 5% *** significant at 1%.

APPENDIX B

In this appendix section, we discuss the additional model predictions on the heterogeneous treatment effects of masking evaluator identity.

Proposition 2.1: *When the CGCS is risk-averse, masking evaluator identity is more effective in improving performance.*

In the baseline model setup, all results are driven by the "price effect," and masking does not introduce any real uncertainty: costs are deterministic, and the stochastic return to influence activities is linear and symmetric. Therefore, the CGCS's risk attitude does not matter in his relative allocation of efforts between u and x. We need a model in which the benefits of the influence activities are not linear and there exists some asymmetry for risk aversion to affect the CGCS decision. We propose one model here that pushes the asymmetry to the extreme in that it leads the CGCS to pursue influence activities with only one of the leaders.[22]

In a modified setup where each leader cares about the relative rather than absolute influence that he receives, uncertainty and risk-aversion start to play a role in addition to the baseline "price" mechanism. The intuition is that when each leader cares about how much influence he receives relative to the other person, the CGCS can only effectively influence one leader, and has to bet on whom to influence in the masked scheme. Specifically, assume that the CGCS's payoff in each state can be written as:

$$V_j = v(x) + M(u_j, u_{-j}) - ax - b_1 u_1 - b_2 u_2,$$

where $M(u_j, u_{-j}) = u_j$ if $u_j > u_{-j}$, and 0 otherwise. Without loss of generality, assume that $b_1 \leq b_2$, which implies $u_1 > u_2$.

Assume that the CGCS is risk-averse, with a mean-variance utility function, and risk aversion ϕ. In the masked scheme, the CGCS chooses efforts to maximize:

$$U = E[V] - \phi \cdot Var[V] = v(x) + \left(\frac{1}{2} - b_1\right) \cdot u_1 - b_2 u_2 - ax - \frac{\phi}{4} \cdot u_1^2$$

[22] The result that risk aversion reduces influence activities would also be obtained with alternative setup, for instance, a model where the production funcition of influence activities is concave and asymmetric, or a model where there is a fixed cost in influencing each supervisor.

As long as the marginal return to productive performance is above a certain lower bound ($v'(T) > a - b_2$), we have:

$$\frac{\partial \mathrm{U}}{\partial u_2} = -v'(x) + a - b_2 < 0$$

This leads to a corner solution: $u_2 = 0$. The CGCS maximizes utility by choosing u_1:

$$\frac{\partial \mathrm{U}}{\partial u_1} = -v'(x) + \frac{1}{2} + a - b_1 - \frac{\phi}{2}u_1 = 0$$

Applying the implicit function theorem to this FOC equation, we have:

$$\frac{du_1}{d\phi} < 0, \frac{dx}{d\phi} > 0$$

Therefore, when the agent is more risk-averse, he responds more strongly to the masked scheme, by relocating more efforts from influence activities to the common productive dimensions.

The intuition is that, when masking evaluator identity, we are introducing uncertainty in the return to influence activities. As a result, risk-averse CGCSs have more incentives to reallocate efforts from the risky investment (influence activities) to the safe investment (productive dimension).

Proposition 2.2: *When job tasks are multi-dimensional, masking evaluator identity is more effective when the two supervisors' subjective weights for different dimensions of performance are more consistent.*

To capture the fact that productive performance x could be multi-dimensional, and the two leaders could have different weights when aggregating different dimensions of performance, assume that there are two dimensions of performance: x_a and x_b, which are perfect substitutes for the organization (thus avoiding the additional issue related to possible complementarity of tasks in order to concentrate on the pure preference heterogeneity effect).

To illustrate how the masked scheme affects performance relative to the revealed scheme when leaders have more aligned preferences, we derive the equilibrium CGCS performance under four different conditions: with vs. without heterogeneity in supervisor preferences under the revealed vs. masked scheme. For tractability, we model "heterogeneous preferences" as the extreme case of two supervisors appreciating orthogonal dimensions of performance (supervisor 1 considering x_a only, and supervisor 2 considering x_b only).

Specifically, under the revealed scheme without heterogeneity, we have:

$$V = v(x_a + x_b) + \sum_j s_j u_j - a(x_a + x_b) - \sum_j b u_j$$

which leads to:

$$x_r = x_a + x_b = v'^{-1}(1 + a - b)$$

Under the revealed scheme with heterogeneity, we have:

$$V_1 = v(x_a) + u_1 - a x_a - b u_1$$
$$V_2 = v(x_b) + u_2 - a x_b - b u_2$$

which leads to:

$$x_a = v'^{-1}(1 + a - b); \quad x_b = 0 \text{ under leader 1}$$
$$x_a = 0; \quad x_b = v'^{-1}(1 + a - b) \text{ under leader 2}$$

Under both leaders, we thus have:

$$x_r = v'^{-1}(1 + a - b)$$

Therefore, under the revealed scheme, overall performance is the same with or without heterogeneity; this stark result is due to the perfect substitutability of tasks in performance.

Under the masked scheme without heterogeneity, we have:

$$V = v(x_a + x_b) + \frac{1}{2}(u_1 + u_2) - a(x_a + x_b) - b(u_1 + u_2)$$

which leads to:

$$x_m^{hom} = x_a + x_b = v'^{-1}\left(\frac{1}{2} + a - b\right).$$

Under the masked scheme with heterogeneity, we have:

$$V = \frac{1}{2}(v(x_a) + u_1 + v(x_b) + u_2) - a(x_a + x_b) - b(u_1 + u_2),$$

which leads to:

$$x_m^{het} = x_a + x_b = 2v'^{-1}(1 + 2a - 2b)$$

The effect of heterogeneous supervisor preferences is:

$$\Delta X = x_m^{het} - x_m^{hom} = 2v'^{-1}(1 + 2a - 2b) - v'^{-1}(\frac{1}{2} + a - b)$$
$$\approx v'^{-1}(\frac{1}{2} + a - b) + (1 + 2a - 2b)(v'^{-1})'(\frac{1}{2} + a - b).$$

The first term is positive and the second term is negative. If we note $z = \frac{1}{2} + a - b$ and $w = v'^{-1}$, then if $-2z\frac{w'(z)}{w(z)} > 1$, ΔX is negative. Hence if the elasticity of w is smaller than

$-1/2$, which means that v is not "too concave," productive effort under the masked scheme is higher when the two supervisors have homogeneous preferences. This implies that the return to the masked scheme is larger when the two supervisors have homogeneous preferences.

Intuitively, as illustrated in the baseline model, the masked scheme improves CGCS performance because it increases the relative return to productive efforts, as compared to supervisor-specific influence activities. When the two supervisors have heterogeneous preferences about different dimensions of productive performance, from the CGCS's perspective, the return to productive performance is more supervisor-specific, which weakens the reason that CGCS perceived productive efforts as offering a better payoff than influence activities. Therefore, when the two supervisors have more aligned preferences regarding different dimensions of productive performance, the return to the masked scheme would be larger.

Proposition 2.3: *If the two supervisors give different weights to the same productive task, when the asymmetry is larger, the revealed scheme is more effective, and the return to adopting the masked scheme is lower.*

The baseline model assumes that both supervisors apply the same rate of substitution between productive performance (x) and influence activities (u_j). In some cases, a supervisor could assign more job tasks to the CGCS than his counterpart does, which causes the two supervisors to have different rates of substitution between x and u. In this section, we extend the baseline model to incorporate the evaluator-specific rate of substitution between x and u:

$$Y_j = \alpha_j v(x) + u_j$$

Let $\alpha = \frac{1}{2}(\alpha_1 + \alpha_2)$, and $d = \frac{1}{2}(\alpha_1 - \alpha_2)$. Without loss of generality, assume $\alpha_1 > \alpha_2$. Solving the maximization problem, we have:

$$E_r[x] = \frac{1}{2} v'^{-1}\left(\frac{1+a-b}{\alpha+d}\right) + \frac{1}{2} v'^{-1}\left(\frac{1+a-b}{\alpha-d}\right)$$

$$E_m[x] = v'^{-1}\left(\frac{1+2a-2b}{2\alpha}\right)$$

Since v'^{-1} is decreasing and convex, we can show that:

$$\frac{\partial(E_m[x] - E_r[x])}{\partial d} = -\frac{\partial E_r[x]}{\partial d} < 0$$

Therefore, the larger the asymmetry between the two supervisors in frequency of job task assignment , the higher will be the level of productive efforts under the revealed scheme, and the less effective the masked scheme is in improving CGCS performance. The intuition for this result is that under the revealed scheme, the CGVS will choose levels of x apart from each other, corresponding with the weight given to $v(x)$ by the supervisors. The supervisor with the higher weight on $v(x)$ induces a higher level of productive effort, while the supervisor with a lower weight induces lower effort. Given that the production function is concave, an increase in weight (similar to a reduction in the marginal cost of effort) induces an increase in effort greater than the reduction in effort induced by a similar decrease in weight. Hence the farther apart the weights, the higher the average of the two values of the effort on productive activities. In contrast, in the masked scheme, optimal x only depends on the average weight given to productive activities, not on the heterogeneity across supervisors.

Intuitively, suppose there are two supervisors; one gives a larger weight to productive performance (x), the other gives a smaller weight. In the revealed scheme, if the former (latter) is chosen as the evaluator, the CGCS will exert more (less) effort on productive performance. Since the production function of productive performance is concave, such dispersion of weights leads to better performance on average than when the two supervisors have more similar weights, holding the average weight constant. In the masked scheme, the CGCS no longer knows which supervisor will evaluate, so he no longer knows what the weight for productive performance will be. A risk-neutral CGCS then decides efforts based on only average weight, ignoring the difference in weights.

Proposition 2.4: *If the two supervisors have imperfect information about job performance, the larger the information asymmetry is between the two supervisors, the less effective the masked scheme is.*

The baseline model assumes that the return to productive efforts is concave ($v'' < 0$), and the cost of productive efforts is linear ($c(x) = ax$). For tractability of the proofs, in this section we assume instead that the return to productive efforts is linear, while the cost of productive efforts is convex. The intuition behind the two setups is the same, while the derivations will be simplified.

Assume that $V = \theta x + \sum_j s_j u_j - \sum_j b_j u_j - c(x)$, where $c(x)$ is increasing and convex, and $\theta \sim f(1, \sigma)$. The CGCS is risk-averse: $U = E[V] - \phi \cdot Var[V]$. Replacing u with $T - x$, we get:

$$V = \theta x + (T - x) - b(T - x) - c(x),$$

and

$$U = bx - c(x) - \phi x^2 \sigma^2.$$

Assume that the two leaders have different levels of noise in their observation of performance: $\sigma_1^2 = \sigma^2 + \frac{1}{2}\delta$, $\sigma_2^2 = \sigma^2 - \frac{1}{2}\delta$. For simplicity, assume that $c(x)$ is quadratic.

Under the revealed scheme, we can derive the following first-order conditions:

$$c'(x_1) + 2\,\phi\,\left(\sigma^2 + \frac{\delta}{2}\right) x_1 = b; c'(x_2) + 2\,\phi\,\left(\sigma^2 - \frac{\delta}{2}\right) x_2 = b.$$

Applying the Implicit Function Theorem, we get:

$$\frac{dE_r[x]}{d\delta} = \frac{1}{2}\,\frac{-\phi\, x_1}{c'' + 2\,\phi\,(\sigma^2 + \frac{\delta}{2})} + \frac{1}{2}\,\frac{\phi\, x_2}{c'' + 2\,\phi\,(\sigma^2 - \frac{\delta}{2})} > 0$$

Under the masked scheme, we have:

$$V = \frac{1}{2}(\,\theta_1 x + u_1) + \frac{1}{2}(\theta_2 x + u_2) - b(u_1 + u_2) - c(x)$$

$$U = x + u - 2bu - c(x) - \frac{1}{4}\phi\,(\sigma_1^2 + \sigma_2^2)x^2$$

Solving the maximization problem, we get the first order condition:

$$\frac{1}{2} + b - c'(x)\; - \phi\,\sigma^2\, x = 0$$

Applying the Implicit Function Theorem, we get:

$$\frac{dx_m}{d\delta} = 0$$

Therefore, we have:

$$\frac{d(E_m[x] - E_r[x])}{d\delta} = -\,\frac{dE_r[x]}{d\delta} < 0$$

The intuition is as follows. Under the revealed scheme, each CGVS will define the optimal x such that the marginal cost is equal to the marginal benefit of his action (discounted by the noise on the information received by the supervisor). This leads to a higher level of activity when the supervisor is less well informed. Because the cost function is convex, a given increase in noise reduces the level of x by less than the same reduction in noise increases the level of x with the other supervisor. Hence, the larger the discrepancy between the noise in

the signals received by the supervisors, the higher the average level of productive activities will be under the revealed scheme, and the lower the return to switching from the revealed to the masked scheme.

The intuition is similar to that of *Propositions 2.3*. Incompelete information about performance discourages productive efforts. In the revealed scheme, an information asymmetry between the two supervisors would increase the average effort from the CGCS due to the concavity of the production function. In the masked scheme, a risk-neutral CGCS would act on the average information for the two supervisors, ignoring the level of asymmetry between them.

APPENDIX C

Sample notification letter (Revealed Scheme):

Dear Mr. [redacted]:

Greetings!

Per the request of the provincial human resources department, we, a research team based at Renmin University in China, will be conducting a "third-party evaluation" of CGCS performance in this fiscal year. The results of this third-party evaluation will be used by the provincial human resources department for decision making.

In June 2018, we will send a team of enumerators to your work unit ([redacted] department in [redacted] township), to collect information about your work performance in the past year. **Specifically, among your two supervisors, Mr. [redacted] and Mr. [redacted], we have randomly selected Mr. [redacted] to be the evaluator. We will collect his assessments of your work performance by the end of the evaluation cycle, and provide that information to the provincial human resources department.**

The performance information will be used only by the research team and the provincial human resources department. Under no circumstance will we provide your personal information to other irrelevant parties. If you have any questions, please contact us at:

Email: [redacted]

WeChat: [redacted]

Phone: [redacted]

Regards,

Renmin University, School of Public Administration

Sample notification letter (Masked Scheme):

Dear Mr. [redacted]:

Greetings!

Per the request of the provincial human resources department, we, a research team based at Renmin University in China, will be conducting a "third-party evaluation" of CGCS performance in this fiscal year. The results of this third-party evaluation will be used by the provincial human resources department for decision making.

In June 2018, we will send a team of enumerators to your work unit ([redacted] department in [redacted] township), to collect information about your work performance in the past year. **Specifically, among your two supervisors, Mr. [redacted] and Mr. [redacted], we will randomly select one of them to be the evaluator. We will collect this evaluator's assessments of your work performance by the end of the evaluation cycle, and provide that information to the provincial human resources department.**

The performance information will be used only by the research team and the provincial human resources department. Under no circumstance will we provide your personal information to other irrelevant parties. If you have any questions, please contact us at:

Email: [redacted]

WeChat: [redacted]

Phone: [redacted]

Regards,

Renmin University, School of Public Administration