Influence Activities and Bureaucratic Performance: Evidence from a Large-Scale Field Experiment in China

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

Shaoda Wang[†], Qiong Zhang[▪]

[Preliminary and Incomplete]

Subjective performance evaluation is widely used by firms and governments to provide work incentives. However, delegating evaluation power to senior leadership could induce influence activities: agents might devote much efforts to please 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 state employees are able to impose evaluator-specific influence to affect evaluation outcomes, and that this process could be partly observed by their co-workers. Furthermore, introducing uncertainty in the identity of the evaluator, which discourages evaluator-specific influence activities, can significantly improve the work performance of state employees.

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

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 California, Berkeley. Email: shaoda@berkeley.edu. Corresponding Author.
[▪] Renmin University of China. Email: zhangqiong8@ruc.edu.cn.

I. Introduction

For a large share of jobs in a modern economy, objective performance measures are difficult to obtain, leading employers to rely heavily on supervisor's 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 multiplicity of the goals 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 decision 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 letting the supervisor manipulate decisions in his own interests, undermining the organization's ability to fully incentivize the agents to perform on their jobs. Despite being consistent with anecdotal examples and case studies, the long-standing theoretical discussion on the existence and implications of influence activities is yet to be backed up by rigorous empirical evidence (Oyer and Schaefer, 2011; Lazear and Oyer, 2012).

Empirically investigating influence activities is fundamentally challenging due to at least three reasons. First, behaviors such as buttering up supervisors or providing personal favors to them are difficult to observe by nature. Agents may 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 to improve evaluation outcomes (instead of simply being friendly), making us unable to exclusively classify them as influence activities. Third, even if the existence of influence activities is established, quantifying the implications of influence 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 to address these three challenges altogether, and provide the first rigorous empirical evidence on the existence and implications of influence activities in 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 state employees are referred to as College Graduates Civil Servants (CGCSs) in this paper.

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

Exploiting this unique setting, we collaborate with two provincial governments in China to randomize two different types of performance evaluation schemes among their 3,785 CGCSs. In both schemes, we randomly select one of the two supervisors to be the "evaluator" (whose opinion matters for future promotion), but also collect the assessment from the other non-evaluating supervisor. The only difference is that, in the first "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 career development; while in the second "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 be having 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 the colleagues of the CGCS to speculate who among the two supervisors would give higher assessments of CGCS performance, they are also more likely to (correctly) guess the randomly selected evaluator would do, of which they were not informed. These results are consistent with a scenario where the agent is able to impose supervisor-specific influence activities to improve evaluation outcomes, and such influence activities could partly be observed by his other co-workers.

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

In the masked scheme, since the CGCS no longer knows who the evaluating supervisor is during the evaluation cycle, the asymmetry in supervisor assessments disappears. Furthermore, the masked scheme improves CGCS performance, as measured by both average colleague assessment and average supervisor assessment. The effect of the masked scheme is stronger when the CGCS is more risk averse, when the two supervisors have more aligned preferences for performance, and when the two supervisors have little information asymmetry. These results can be rationalized in a conceptual framework, where introducing uncertainty in the identity of the evaluator incentivizes the agent to reallocate efforts from evaluator-specific influence activities toward productive dimensions that would be appreciated by both supervisors.

We conduct a battery of additional analyses to rule out alternative interpretations of our findings, such as supervisor behavioral change due to being selected to evaluate, agents responding differently to the evaluator's job assignments, evaluating supervisor\masked supervisors getting better information on CGCS performance, CGCSs trying to influence everyone under masked scheme, etc. We also find suggestive evidence that the improvement in performance is not simply "cheap talk" by supervisors and colleagues: in the masked scheme, the CGCSs are more likely to be assessed as "qualified for tenure" and also receive higher salaries in positions with established incentive pay schemes. We further provide various ways to benchmark the magnitude of our treatment effects.

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 workplace. As pointed out by Lazear and Oyer (2012), while a large theoretical literature has studied how agents try to influence the assessments of their supervisors (e.g. Milgrom and Roberts, 1988; Milgrom, 1988; Meyer et al., 1992; Schaefer, 1998), there lacks quality empirical evidence, aside from anecdotes and case studies, to verify these arguments. Our paper fills in this gap by providing field experimental evidence on influence activities among Chinese local state employees, as well as quantifying the causal impacts of alleviating influence activities on job performance. More broadly, subjective performance evaluation is ubiquitous across private and public sectors, and has been investigated extensively by a large body of theoretical works (Gibbons and Murphy, 1992; Baker et al., 1994; Prendergast and Topel, 1996; MacLeod, 2003). 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 could undermine the power of subjective performance evaluation.

Second, this paper adds to a growing empirical literature on the personnel economics of the state, specifically on incentivizing state employees (Finan et al., 2016). Most of the existing works on this topic focus on the role of financial incentives,[2] with only a few exceptions studying non-pecuniary incentives, such as transfers and postings (Banerjee et al., 2012), 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 incentive scheme in public sectors, but has rarely been studied in the literature. In addition, we show that holding the "final reward" fixed, slightly revising the performance evaluation practice can lead to substantial improvement in job performance of state employees, indicating a highly cost-effective way to strengthen 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 different performance indicators, such as GDP growth rate (Guo, 2009), environmental standard (He et al., 2018), and implementation of the one-child policy (Serrato et al., 2018), affect the incentives and behaviors of provincial and prefectural officials in China. However, the existing evidence has been focusing almost exclusively on relatively high-level government officials, leaving the vast majority of local bureaucrats under-researched, whose incentives and constraints could potentially be very different from their high-level leaders. For instance, a key distinction is that, the job tasks for the low-level bureaucrats are much more difficult to be quantified by objective measures such as GDP growth and environmental quality, so most of the grassroots bureaucrats are only rewarded based on subjective evaluations by their supervisors. Our paper intends to unravel the black box of the incentive schemes for grassroots bureaucrats in China, who are the building blocks of state capacity and play key roles in policy implementation and public service delivery.

The remainder of this paper is organized as follows. In Section II, we layout a simple conceptual framework to help guide the experimental design, and generate some testable hypotheses. In Section III, we introduce the institutional background, design, and implementation of our field experiment. In Section IV, we present experimental results to test

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

the main predictions of our conceptual framework. In Section V, we discuss the potential alternative interpretations of our findings. In Section VI, we benchmark the magnitude of our treatment effects. Section VII concludes.

II. Conceptual Framework

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

To mimic our "duality" empirical setting, we assume that there are two supervisors, $j \in \{a, b\}$. In addition to working on the productive dimension of the job, agent i also has the option to exert (supervisor-specific) influence activities u_{ij} to please supervisor j, in order to improve the assessment score:

$$Y_{ij} = x_i + u_{ij}$$

The agent maximizes his expected utility, which is a function of both evaluation outcome and work efforts, subject to a time constraint:

$$\max_{x_i, u_{ij}} \sum_{j \in \{a,b\}} s_{ij} \cdot V(Y_{ij},\ L_i)$$

s.t.

$$f(x_i) + \sum_{j \in \{a,b\}} g(u_{ij}) + L_i = T$$

where s_{ij} is the probability of each supervisor j's assessment being used to determine individual i's reward in the performance evaluation scheme ($\sum_{j \in \{a,b\}} S_{ij} = 1$); L_i is leisure; $V(Y, L)$ is the utility function of the agent, which is increasing and concave in both evaluation score and leisure. In the time constraint, $f(x)$ is a convex and increasing function measuring the time needed to obtain performance x, $g(u)$ is a convex and increasing function measuring the time needed to exert influence activities u, T is the total time budget for an individual.

Other colleagues also observe the 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 evaluations. Colleagues therefore receive no influence activities, and base their evaluations solely on the productive dimension:

$$Y_{ic} = x_i$$

Proposition 1: *when one of the two supervisors is randomly chosen to be the evaluator, and his identity is known to the agent, other things being equal, the evaluating supervisor gives higher assessment than the non-evaluating one.*

Without loss of generality, assume: $s_a = 1, s_b = 0$. Solving the model, we can show that:

$$u_{ia} > u_{ib} = 0$$

which leads to:

$$Y_{ia} > Y_{ib}$$

The intuition is that, when the agent knows the identity of the evaluator, he would exert more evaluator-specific efforts (i.e. $u_{ia} > u_{ib}$) that lead to more positive assessment from the evaluating supervisor ($Y_{ia} > Y_{ib}$).

Proposition 2: *when one of the two supervisors is randomly chosen to be the evaluator, but his identity is unknown to the agent, other things being equal, compared to the case where the evaluator's identity is known, the performance of the agent will improve, as measured by both supervisor assessment and colleague assessment.*

In this masked scheme, $s_a{}' = s_b{}' = \frac{1}{2}$. Since the identity of the evaluator is unknown, the agent would reallocate efforts from supervisor-specific influence activities (u_{ij}) toward productive tasks that would be appreciated by both supervisors (x_i). Therefore, compared to the revealed scheme, we can show:

$$u'_{ia} = u'_{ib} < u_{ia};\ \ x'_i > x_i$$

and:

$$Y'_{ia} + Y'_{ib} > Y_{ia} + Y_{ib}$$

$$Y'_{ic} > Y_{ic}$$

The intuition is that, since the identity of the evaluator is unknown, the marginal cost of improving expected evaluation outcome through influence activities doubles, while the marginal cost of improving evaluation outcome through the productive dimension remains unchanged. Therefore, the agent re-optimizes by switching from exerting influence activities towards being productive for the organization.

Proposition 2.1: *if the evaluating leader is more positive under the masked scheme than the revealed scheme, then the agent must enjoy less leisure and lower utility under the masked scheme than the revealed scheme.*

Formally, given that $(x_i, u_{ia}) \in \underset{x_i, u_{ia}}{\operatorname{argmax}} V(Y_{ia},\ L_i)$, we know that $V(Y'_{ia},\ L'_i) < V(Y_{ia},\ L_i)$. If $Y'_{ia} > Y_{ia}$, it must be that $L'_i < L_i$.

The intuition is that, in the revealed scheme, the agent only has incentives to please the evaluating leader. If the evaluating leader is happier in the masked scheme, then the agent must has lower leisure and utility, otherwise his inputs under the revealed scheme would be strictly dominated by the inputs under the masked scheme.

Proposition 2.2: *more risk-averse agents are more strongly incentivized when the identity of the evaluating supervisor is unknown.*

When the agent is more risk averse (i.e. $V(Y, L)''$ being pointwise more concave than $V(Y, L)'$), the reallocation from uncertain dimension (u_{ij}) to certain dimensions (x_i and L_i) would be more salient. So we can show:

$$(Y''_{ia} + Y''_{ib}) - (Y_{ia} + Y_{ib}) > (Y'_{ia} + Y'_{ib}) - (Y_{ia} + Y_{ib})$$

Proposition 2.3: *when at least one of the supervisors has imperfect information about job performance, introducing ex post randomization in evaluator identity is less effective as the information gap enlarges.*

Without loss of generality, assume:

$$Y_{ia} = x_i + u_{ia};\ \ Y_{ib} = \alpha \cdot x_i + u_{ib}$$

where $\boldsymbol{\alpha < 1}$. Solving the model, it is straightforward to show that:

$$\frac{\partial (Y'_{ic} - Y_{ic})}{\partial \alpha} > 0$$

The intuition is that, when $\alpha \to 0$, x_i becomes more and more specific to supervisor a, therefore similar to the role of u_{ia}. As a result, the effect of "reallocating towards commonly appreciated dimension" is attenuated.

Proposition 2.4: *when job tasks involve multiplicity, introducing ex post randomization in evaluator identity is more effective when the two supervisors' subjective weights for different dimensions of performance are more consistent.*

Without loss of generality, we extend the previous setting by allowing for two dimensions of performance: x_{i1}, x_{i2}, for which a supervisor imposes weights w_{j1} and w_{j2} respectively ($w_{j1} + w_{j2} = 1$). So the evaluation function becomes:

$$Y_{ij} = w_{j1} \cdot x_{i1} + w_{j2} \cdot x_{i2} + u_{ij}$$

and the time constraint becomes:

$$f(x_{i1}) + f(x_{i2}) + \sum_{j \in \{a,b\}} g(u_{ij}) + L_i = T$$

Solving the model, we prove that:

$$\frac{\partial[(Y'_{ia} + Y'_{ib}) - (Y_{ia} + Y_{ib})]}{\partial|w_{a1} - w_{b1}|} < 0$$

and:

$$\frac{\partial(Y'_{ic} - Y_{ic})}{\partial|w_{a1} - w_{b1}|} < 0$$

The intuition is that when both supervisors have more consistent subjective weights, it is more desirable to reallocate efforts from supervisor-specific influence activities (u_{ij}) to productive dimensions that would be appreciated by both supervisors (x_{i1}, x_{i2}), which improves both average supervisor evaluation and average colleague evaluation.

III. Experiment

Institutional Background

Since the early 2000s, the Chinese government has launched several large-scale human capital transfer programs, which altogether hired more than one million college graduates to work in rural areas, in the hope that their modern human capital and independence from local interest groups could help improve state effectiveness at the local levels. For example, in the College Graduate Village Officials (CGVOs) program, college graduates were hired as village officials on a contractual basis, and it is found that they were able to significantly improve policy implementation and reduce leakage in poverty subsidies (He and Wang, 2017).

In this paper, we focus on one of such human capital transfer programs, known as the "3+1 Supports" initiative, which was launched in 2006 by the Ministry of Human Resources and Social Security.[3] Several types of civil service positions are offered to the college graduates

[3] In Chinese, the initiative corresponds to the "*San Zhi Yi Fu* (三支一扶)" program. Six other major 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.

through the "3+1 Supports" initiative, including township administrative clerics who work on poverty alleviation, agricultural support, water resource management, and social security provision; teachers in township primary schools; and pharmacists and nurses in township clinics. By the end of 2018, more than 350,000 college graduates had been hired to work as grassroots civil servants through this program. We refer to these individuals as "College Graduates Civil Servants" (CGCSs) in this paper.

The CGCSs are recruited from colleges on a yearly basis. Before each job market season in May, the provincial government announces the total number of vacancies on its website and invites college students to apply. In most provinces, the procedure of CGCS recruitment is similar to that of formal (permanent) state employee recruitment: the applicants need to first take a comprehensive written exam, which is similar to the Administrative Aptitude Test and Essay Writing Test used in the National Civil Service Exam. The high-score applicants are then interviewed and the top-ranked candidates will be recruited. Some provinces also recruit CGCSs by reviewing their application materials, where priorities are often given to (Chinese Communist) Party members, those who have local *Hukou* (household registration status), those who graduated with honors, and those who have more advanced and relevant degrees.

The admission of CGCSs is highly competitive. For examples, in Shandong province, the total number of available positions in 2017 was around 1,500, but the program attracted more than 31,000 applicants (admission rate < 5%); in Guangxi province, the government planned to hire 800 CGCSs in 2017; but the total number of applicants exceeded 13,600 (less than 6% applicants were recruited). In most provinces and in most years, the admission rate for the "3+1 Supports" program is below 10%. Such intense competition ensures the quality of the selected CGCSs.

A CGCS is similar to a formal entry-level township civil servant in many ways. For CGCSs in clerical positions, for example, like most formal rural civil servants, the job tasks tend to be a combination of routine paper works, interactions with villagers, and case-based assignments from supervisors, making job performance hardly quantifiable, and difficult to compare across individuals. For CGCSs in more specialized positions like township clinic nurses or primary school teachers, the job tasks are also similar to their formal state employee colleagues. While certain dimensions of these jobs are better defined than cleric jobs, the grassroots nature of these jobs still make objective performance evaluation largely infeasible, e.g., student score cannot be used to incentivize teachers due to the lack of unified written exams at the primary

school level. As a result, the evaluation of a CGCS relies almost solely on the evaluating supervisor's subjective assessment, which is also the *status quo* for the vast majority of civil service jobs across the world.

The only major difference between a CGCS position and a formal civil servant position is that the former is contract-based, while the latter is permanent ("tenured").[4] The majority of CGCSs are eager to be promoted to formal state employees upon finishing the two-year term, which would only be approved by the government subject to satisfactory job performance. As a result, aspiring CGCSs have strong incentives to impress their supervisors and improve the evaluation outcomes. While the strong incentives could potentially lead to higher work efforts, they might also cause influence activities to please the evaluating supervisor in ways that are inconsistent with the government's interest.[5]

A CGCS is supervised by two leaders: a party leader and an administrative leader. This is due to the unique dual-leadership governance structure in China: in principal, the administrative leader is in charge of the daily operation of the government entity, while the party leader oversees the process; the administrative leader often has larger economic power, while the party leader can exert larger political power. These two leaders have the same official ranking, but normally the party leader is perceived to have an edge in authority. This dual-leadership provides *de facto* checks and balances in China's local governance system (Li, 2018) and is adopted by different levels of administrative units (ranging from the central government to the small village committees). 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, this dual-leadership rule needs to be enforced. The dual-leadership setting is important for our experiment, as we will randomly choose one from the two to evaluate the performance of a CGCS.

Experimental Design and Data

In collaboration with two provincial governments in China, we randomize two subjective performance evaluation schemes, a "revealed scheme" vs. a "masked scheme," among all their 3785 CGCSs. In both schemes, one of the two supervisors is randomly selected to be the

[4] In Chinese, it corresponds to the "*Bianzhi* (编制)" status.

[5] Simple examples of such influence activities include picking up supervisor's kid from school, making coffee/tea for supervisor, doing chores for supervisor, etc.

evaluating supervisor for CGCS performance, meaning that his assessment will be given a 100% weight in the evaluation outcome, which in turn affects future career development of the CGCS. For the other randomized non-evaluating supervisor, we also collect his assessment of CGCS performance, but this assessment is given 0% weight in determining the evaluation outcome. In both schemes, we never directly inform a supervisor whether or not he is chosen as the evaluating supervisor, nor do we inform the colleagues of the CGCS.

In the "revealed" scheme, we inform (and only inform) the CGCS about the identity of his evaluating supervisor, at the beginning of the evaluation cycle. This mimics the *status quo* of CGCS performance evaluation, where the agent is informed about the evaluating supervisor *ex ante*. The difference is that, in the *status quo*, 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, we eliminate the endogeneity in evaluator selection, and directly pin down the causal impact of knowing the identity of the evaluator. Two thirds of the CGCSs are assigned to this scheme.

In the "masked" scheme, we do not inform the CGCS about the identity of the evaluator until after the end of the evaluation cycle. Compared to the CGCSs in the revealed scheme, those in the masked scheme do not know the identity of their evaluator throughout the year. As a result, it becomes less beneficial to exercise supervisor-specific influence activities, the CGCSs are expected to reallocate their efforts toward productive dimensions that would be appreciated by both supervisors. One third of the CGCSs are assigned to this scheme.

We focus primarily on two sets of outcomes: supervisors' assessments and colleagues' assessments of CGCS performance. We define "colleagues" as co-workers in the same office as the CGCS that are not hired through the "3+1 Supports" program. We consider the colleagues' assessments as the ideal performance measure in this context for three reasons. First, the colleagues were randomly chosen from the same office where the CGCSs worked. They worked 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 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 senior colleagues do not directly complete with each other in career development. Finally, the CGCSs do not have strong incentive to influence their colleagues for evaluation reasons

throughout the study. At the baseline, the provincial governments explicitly told the CGCSs that only the evaluating supervisor's opinions would be used for evaluating their final performance. Most CGCSs in fact did not even expect that we would survey their colleagues until the surveyors were sent to their offices at the end line.

The combination of our experimental design and data collection allows us to test the main propositions of our conceptual framework in powerful ways. First, Proposition 1 suggests that in the revealed scheme, the randomly chosen evaluator should give more positive assessments of CGCS performance, due to evaluator-specific influence activities imposed by the CGCS. Since the evaluator is randomly selected from the two supervisors, we could directly test this prediction by estimating the causal effect of "being the evaluator" on "giving more positive assessment." In addition, for each CGCS, we also elicit his colleagues' beliefs regarding which of the two supervisors would give the CGCS more positive assessments. While we did not inform the colleagues about which supervisor of the two was chosen to evaluate, to the extent that supervisor-specific influence activities indeed happened and could at least be partly observed by others throughout the year, the colleagues should be able to correctly predict that the (randomly selected) evaluator would favor the CGCS.

Second, Proposition 2 asserts that when the identity of the evaluator is masked, evaluator-specific influence activities would be alleviated, causing the agent to reallocate efforts toward productive dimensions that would be appreciated by both supervisors, which results in better overall performance. The first part of this statement could be verified by seeing whether "supervisor assessment asymmetry" disappears as we switch from the revealed scheme to the masked scheme; the second part of this statement could be verified by testing whether CGCSs in the masked scheme have better performance as measured by average supervisor assessment and colleague assessment. Furthermore, our conceptual framework predicts that the masked scheme is most effective when the agent is more risk averse, when the two supervisors have more aligned preferences, and when the two supervisors have smaller information asymmetry. These additional predictions can be tested by exploring the heterogeneous treatment effects of adopting the masked scheme.

Implementation

Our experiment is conducted in collaboration with two of the largest provinces in China, with a combined population of nearly 150 million. Province A is coastal and more developed, while

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

The baseline survey was carried out in September 2017, one month after the 2017-cohort CGCSs finished job training and were assigned to their positions. All the CGCSs were then randomized into the two evaluation schemes based on their work units, and informed each CGCS about the evaluation scheme he was assigned to. The randomization was conducted at the work unit level instead of the individual level.[6] Different CGCSs who worked in the same unit (i.e., an office or an organization branch led by the same set of supervisors) were assigned to the same scheme. This is per the request of our government partner to ensure that the evaluation results of CGCSs working in the same unit can be fairly compared to each other. In this setting, because 83.9% of the work units only had one CGCS allocated, randomizing at the work unit level did not hurt our statistical power in any substantial way, as compared to randomizing at the individual level.

To ensure the credibility of our intervention, the provincial governments sent a formal letter with official sealed stamps to every CGCS. The letter emphasized the importance of the performance evaluation, confirmed the evaluation scheme (including the name of the relevant supervisor) specific to the CGCS, and the detailed timeline. 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 surveyors visited the office of a CGCS, if there were less than 5 colleagues in the office, all of them were invited to fill in the colleague questionnaire; if there were more than 5 colleagues, the surveyor randomly

[6] In Chinese, a work unit corresponds to a "*Gong Zuo Dan Wei* (工作单位)."

sampled 5 of them to fill in the colleague questionnaire, using a random number generator.[7] To protect the privacy of the colleagues and encourage truth-telling, colleague questionnaires were strictly anonymous, and CGCSs were not allowed to communicate with the 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 completed online, with an individual-specific link for each supervisor, listing all the CGCSs in his unit.

In the colleague and supervisor surveys, we ask questions on basic information of the colleague\supervisor, his interactions and familiarity with the CGCS, the job tasks of the CGCS, and his assessment of the CGCS along various dimensions. Specifically, we asked for an overall assessment of CGCS performance, as well as a "revealed preference" measure asking the colleague\supervisor whether they think the GCGS deserves a tenured position in this unit upon finishing his 2-year contract.

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

Balance and Attrition Tests

To ensure the validity of our experiment, we conduct a battery of relevant balance tests: in the revealed scheme, supervisor characteristics should be balanced between the evaluating and non-evaluating supervisors, which supervisor is selected to evaluate should also be orthogonal to both CGCS and colleague characteristics; between the two evaluation schemes, CGCS characteristics, supervisor characteristics, as well as colleague characteristics should all be balanced. As shown in Appendix Tables A1-A3, these balance tests all go through, suggesting that our randomization was well executed.

Between the baseline and the end-line surveys, we lost 918 CGCSs (24.3%) in the sample. The main reason for the attrition is that the CGCSs or their supervisors changed their positions during our study period. For example, a CGCS could be relocated from one township to another because of changes in policy targets. The supervisors could be promoted,

[7] If a colleague was not in the office when the surveyor visited, his contact information was collected and he would be surveyed over the phone in the following day. To ensure data accuracy, the team leader randomly called some colleagues in the following days to verify the sampling procedure and the answers collected.

retired, or move to other institutions. Such job changes would break the supervisor-subordinate relationship and thus invalidate the experimental design. Besides, some CGCSs passed the civil service exam or got admitted by graduate schools (participation in the "3+1 Supports" program gives them extra points in the exam). These people tended to quit their current CGCS jobs and left our experiment. To test whether our experiment caused such attrition, we estimate the relationship between attrition status and the treatment status in Appendix Table A4. We find that our treatment does not lead to selective attrition.

IV. Main Results

In this section, we present a series of experimental results consistent with our theoretical predictions. Specifically, we show that: (1) when the identity of the evaluator is revealed to the agent, the assessment given by the evaluating supervisor is higher than that given by the non-evaluating supervisor, and that this asymmetry in assessments is correctly predicted by the colleagues of the agent; (2) when the identity of the evaluator is masked, the asymmetry in supervisor assessments disappears, and agent performance improves significantly as measured by average colleague assessment and average supervisor assessment; (3) the evaluating supervisor being more positive under the masked scheme indicates increased effort and decreased utility for the agent; (4) masking evaluator identity is more effective when the agent is more risk-averse, when the two supervisors have similar opinions on what constitute good performance, and when there is little information asymmetry between the two supervisors.

These results are highly compatible with an interpretation of the agent undertaking influence activities (that target the evaluator and are at the expense of the assigned job) in the revealed scheme, and reallocate efforts from influence activities toward productive dimensions in the masked scheme, as formalized in our conceptual framework. The combination of these findings could already rule out some obvious confounding stories, and we will further confront each of those remaining alternative interpretations in Section V.

Asymmetry in Leader Assessments under the Revealed Scheme

In Table 1, the outcome variable is "Supervisor 1's evaluation score minus Supervisor 2's score," measuring the extra positivity of Supervisor 1 than Supervisor 2 towards the same CGCS. The explanatory variable is a dummy variable indicating whether Supervisor 1 was

(randomly) chosen to be the evaluating supervisor. We focus on the revealed scheme, in which every CGCS is informed about the identity of his evaluator at the beginning of the evaluation cycle, who is randomly selected from the two co-supervisors.

In Column (1), we find that in the revealed scheme, if a supervisor was chosen as the evaluator at the baseline, he indeed gave more positive assessment at the end line. In Column (2), we include a rich set of control variables in the regression and the estimated coefficient remains highly stable, confirming that the "Supervisor 1 Evaluating" treatment is randomly assigned. This asymmetry in supervisor assessments is consistent with the agent imposing 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 be partially 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 noisy signals of u_{ij}, even without knowing who is the evaluator, they would be more likely to correctly identify the more positive supervisor.

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." 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 indeed correctly predict the direction of supervisor assessment asymmetry.

No Asymmetry under the Masked Scheme

If the findings in Table 1 are indeed caused by evaluator-specific influence activities as we interpreted, then these results should only exist when the agent 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 leader assessments either.

Therefore, as a placebo test, in Table 2, we focus on the masked scheme, where the identity of the randomly chosen evaluator is not announced until the end of the evaluation cycle. As we can see in the first two columns, being selected as the evaluator no longer has any impact on the positivity of assessment, again confirming our theoretical prediction.

In Columns (3)–(4), we find that when the identity of the evaluating supervisor is unknown to the agent, colleagues are no longer able to identify the evaluator as the more positive supervisor.

Combined together, these results are consistent with a case where agents impose evaluator-specific influence activities when and only when knowing who the evaluator is.

Improved Performance under Masked Scheme

We exploit the random assignment of CGCSs between the two evaluation schemes to estimate the causal effect of masking evaluator identity on job performance.

First, in Table 3, we examine colleague assessments, which we consider as the ideal measure of CGCS performance, for reasons explained in Section III. Specifically, the dependent variable is average colleague assessment of the CGCS's performance, which is framed relative to other civil servants working in the same work unit. The assessment score ranges from 1 to 7, representing different categories from "worse than all the colleagues" to "better than all the colleagues" in the questionnaire. Relatedly, we have an outcome variable indicating whether the colleagues think the CGCS's performance ranks in the Top 10% of the organization.

The results in Columns (1)–(2) show that masking the identity of the evaluator significantly improved colleague assessments of CGCS performance. In Columns (3)–(4), we use the "Top 10%" as the outcome variable, all the results remain similar: CGCSs in the second treatment arm are more likely to be recognized as top performers by their colleagues.

These results are consistent with our theoretical prediction that masking evaluator identity would reduce influence activities, thus improving job performance. To corroborate this finding and better understand the driving-force behind performance improvement, we also investigate the impacts on supervisor assessments.

Table 4 reports the treatment effects (of being assigned to the masked scheme rather than the revealed scheme) on supervisor assessments. In Columns (1), the outcome variable is the mean assessment of the two supervisors. We find that the "masking the identity of the evaluator" significantly improves average supervisor assessment. To understand how the CGCSs manage to achieve overall higher assessments, we look at the treatment effects on evaluator supervisor and non-evaluating supervisor separately. Columns (2) and (3) reveal that both the evaluating and non-evaluating supervisor's assessment improved under the masked scheme, with a larger improvement in the non-evaluating supervisor's assessment. The

difference between evaluating and non-evaluating supervisors' treatment effects motivate Column (4): deviation between supervisor assessments is smaller under the masked scheme.

Rationalizing Results on Supervisor Evaluation

In Table 4, the result for the non-evaluating supervisor's assessment is intuitive: the CGCS now needs to impress both supervisors, instead of one, making the non-evaluating supervisor more pleased. The result for the evaluating supervisor's assessment, however, is less intuitive: the evaluating supervisor's assessment score is also improved, despite no longer receiving influence activities catered to his private utility. This result is rationalized in our model: if in scheme 2, the agent sacrifices enough of his leisure to work on productive dimensions, this "labor participation effect" could dominate the "reduced influence activity" effect, leading to an increase in the evaluation from the more favorable supervisor.

We test this intuition from our model in Table 5. In our survey, besides asking the colleagues to give an overall assessment of the CGCS's performance, we also asked them to specifically tell us whether they think the CGCS work harder and overtime relative to other colleagues. In Panel A of Table 5, we find that masking the evaluator's identity indeed makes the CGCS work harder and overtime (as observed by their colleagues), suggesting the CGCS allocate more efforts in the productive dimensions.

In addition, we report the findings for the CGCSs' self-evaluations. Note that results from self-evaluations are less straightforward to interpret, because the incentives behind the self-reporting are complicated.[8] In our setting, for example, almost all of the CGCSs thought that they performed better than their colleagues and the majority of them were in the Top 10% category. They generally exaggerated their work performances, which resulted in smaller variations in self-evaluated performance scores. Acknowledging various problems of self-evaluated outcomes, here, however, we still find that the "masking the identity of the evaluator" positively affected the CGCSs' self-evaluations.

An important prediction of our model is that under the uncertainty treatment, as the CGCS needs to allocate more efforts to the productive dimension, his overall welfare would decrease. In our survey, we asked the CGCS whether they would be willing to find a job in the private sector, and if so, what the expected wage is. In Panel B of Table 5, we find that masking the

[8] A vast psychology literature has looked at this issue but the validity of self-evaluation remains controversial. See Hetts (1999) for a review of the psychology literature.

identity of the evaluator makes the CGCS more willing to accept a job in the private sector, and the "reserve wage" is reduced by about 5%. Our interpretation is that because the CGCSs had to work harder under uncertainty, it made them more willing to quit and accept a low-pay job. These results help rationalize our previous finding that both the evaluating and non-evaluating supervisors' assessment improved.

Heterogeneous Effects of the Masked Scheme

In Table 6, we explore heterogeneous treatment effects under the guidance of our theoretical predictions. Combined together, these results shed light on the channels through which "masking evaluator identity" improves CGCS performance.

In the conceptual framework, we predict that more risk-averse agents will respond more strongly to the introduction of uncertainty. This is confirmed by the results in Column (1), where the interaction term of risk averse dummy and the introduction of the masked scheme is positive and significant.[9]

Our model also predicts that when the information gap of the two supervisors gets larger, the "productive dimension" effectively becomes more and more supervisor-specific. As a result, larger information asymmetry between the two supervisors weakens the treatment effect of masking evaluator identity. This is confirmed by both Column (2), where we use "difference in two supervisors' self-reported information" to measure information asymmetry, and Column (4), where we use "two supervisors' difference in work assignment frequency" to proxy for information asymmetry.

Another prediction of the model is that when the preferences of the two supervisors are highly aligned, masking evaluator identity should be most powerful. In our survey, we separately elicited each supervisor's subjective ranking on the importance of each performance dimension. In Column (3), we find that the treatment effects are larger when the two supervisors give the exact same ranking of importance, confirming the prediction. This result suggests that the treatment effect of masking evaluator identity is not driven by "CGCS responding to more diversified supervisor preferences," because if that is the case, we should see a weaker treatment effect when supervisors have more aligned preferences, not stronger.

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

V. Alternative Interpretations

Our model suggests that under the revealed scheme, an agent imposes evaluator-specific influence activities; and under the masked scheme, the agent no longer knows who is evaluating, so would re-optimize his efforts by reducing influence activities to work on productive dimensions that would be appreciated by both supervisors. In Section IV, we documented a series of evidence supporting the main predictions of our model.

While some confounding interpretation might be consistent with one of the baseline findings, it could hardly rationalize the combination of all these patterns at the same time. Nevertheless, in this section, we treat each of the baseline results separately, and examine the alternative explanations specific to it.

A. Confounders for "Supervisor Assessment Asymmetry"

Confounder 1: Behavioral Differences between the Evaluating and Non-Evaluating Supervisors

Our interpretation of the findings in Table 1 is that the subordinate is able to perform evaluator-specific influence activities, and such behaviors can be observed by their co-workers. An alternative explanation is that the evaluating and non-evaluating supervisors may act differently, simply because that "revealing the identity of the evaluator" might directly affect the evaluator's behaviors. For example, if the evaluating supervisor feels more responsible after knowing that he is the evaluator, he may give the CGCS more positive assessments to avoid potential psychological burdens. Although this concern is alleviated by our experimental design that the supervisors were not informed by the research team about their roles in the evaluation, it is still possible that the CGCSs might deliver this information to them.

To investigate this probability, in our end-line survey, we directly asked the supervisors about whether they were aware of their responsibility in evaluating the CGCS. It turns out that the majority of them (more than 65%) did not know that they were chosen as the evaluators until after the end of the evaluation cycle. In Panel A of Table 7, we estimate the impact of "Supervisor 1 Evaluating" on "Supervisor 1 Being More Positive" separately for different subsamples: the subsample in which the supervisors did not know their (evaluator) roles, and the subsample in which the supervisors knew their roles (likely through the CGCS). We find that in both subsamples, "Supervisor 1 Evaluating" increases the likelihood that "Supervisor

1 Being More Positive," suggesting that our results are not driven by supervisor behavioral change due to being evaluator.

Moreover, in Panel B of Table 7, 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 the survey, (2) whether one supervisor writes more words in describing the subordinated CGCS's tasks than the other supervisor, 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, we should observe the evaluating supervisor to be more likely to answer the survey, write more words in their assessment, and even assign more tasks to the CGCS. Again, our analyses do not support this hypothesis. "Supervisor 1 Evaluating" does not affect any of these outcomes, as evidenced by the regression results in Panel B.

Confounder 2: Higher Information Quality 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 try to send signals to help him better evaluate. This improvement in information quality might improve supervisor assessment, while the CGCS did not actually exhibit any "influence activities" at the cost of the organization.

To examine this interpretation, in our end-line supervisor survey, we asked each supervisor how frequently did the subordinated CGCS, the colleagues of the CGCS, or the other supervisor discuss with him about CGCS performance. We are interested in whether the evaluating supervisor would receive additional information than the non-evaluating supervisor from these sources. In panel C of Table 7, we find that "Supervisor 1 Evaluating" does not make supervisor 1 obtain extra information from any of these sources, as compared to supervisor 2. In other words, revealing the identity of the evaluator to the CGCS does not differentially change the "information quality" of the evaluating supervisor.

B. Confounders for "Treatment Effects of Masking Evaluator Identity"

Confounder 1: CGCS Influencing Everyone under Masked Scheme

We find that CGCSs in the masked scheme on average receive more positive assessments from colleagues and both supervisors, which we interpret as indicating better job performance.

A potential alternative interpretation is that under the masked scheme, CGCSs do not work harder on the productive dimensions. Instead, they simply try to "influence" everyone, i.e., providing personal favors to all their colleagues, which leads to more positive assessments from everyone. This is unlikely to be the case, because the CGCSs have no incentive to do so: we clearly indicated at the beginning of the experiment that only the evaluating supervisor's assessment will determine the evaluation outcome. Nevertheless, we examine this possibility more formally.

If the CGCSs are indeed trying to influence their colleagues more often, we would expect that they interact more with these individuals, and the colleagues and supervisors become more familiar with the CGCSs, in both work and life. However, in Panel A of Table 8, we find that "masking the identity of the evaluator" did not make the CGCSs to communicate more or meet more frequently with their colleagues (Columns (1) and (2)), nor did it make the colleagues become more familiar with the CGCS's work or life (Columns (3) and (4)). This is again inconsistent with the alternative interpretation.

Confounder 2: Higher Information Quality in Masked Scheme

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

To address this concern, in Panel B of Table 8, we examine whether supervisors get additional information on CGCS performance under the masked scheme, either from colleagues or 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 of colleagues and CGCSs reporting to them regarding CGCS performance. This suggests that the improved supervisor evaluations in the masked scheme cannot be explained by changes in information quality.

VI. Benchmarking the Effect Size

Usually, subjective evaluations are adopted when there lack clearly objective performance measures. Due to this nature, it is often challenging to benchmark the improvement in subjective evaluations using objective indicators. To better interpret the economic significance of our "masked scheme" intervention, we use two different ways to provide some more intuitive benchmarks for the magnitude of our estimated coefficients.

Comparing Magnitudes with Correlative Coefficients

In Table 9, we report the correlations between subjective evaluations and a rich set of CGCSs' characteristics.[10] In Column (1), we see that age (or more experience), education and party member have the strongest predictive power in explaining colleagues' assessments. Being one year older is associated with a 0.05-point increase in colleague assessment, graduating from a 4-year college instead of a 3-year community college is correlated with a 0.15-point increase in colleague assessment, and being a party member is associated with 1 0.17-point increase in colleague assessment. In Column (2), we estimate the correlations between supervisors' evaluations and CGCSs' personal characteristics. We find similar patterns that CGCSs' age, education and party membership are positively associated with better evaluation outcomes. The magnitudes of the coefficients are similar to those in Column (1).

Recall in Table 3, "masking the identity of the evaluator" can increase colleagues' assessments by 0.22. The treatment effect is almost "equivalent" to having 4 more years of (life and work) experience, and is substantially larger than the effect of "upgrading" 3-year college graduates to 4-year college graduates, or replacing non-CCP members with CCP-members. Given the substantial edge in prestige and ability associated with 4-year colleges and CCP member status, the treatment effect does seem to be economically significant.

"Revealed Preference" Measures and Objective Benchmark

Relatedly, another concern is that if there is any reason that the incentives to "cheap talk" would vary systematically with our random assignment of evaluation schemes, the economic significance of our findings can be compromised. To address this potential issue, in Table 10,

[10] In Appendix Table A5, we report the partial correlations between evaluations and CGCSs' personal characteristics separately for each variable. The estimates are largely the same.

we examine two "revealed preference" outcomes, which involve real stakes and cannot be explained by cheap talk.

First, we directly ask the colleagues whether they think the CGCSs should be offered a tenured position in the office after the 2-year term. The question was posited at the end of the survey, and many of the colleagues asked our surveyors whether their answers to this question would be referenced by upper level government in the decision-making process. While we told them that all the information in the survey would be useful, their particular cautiousness on this question suggests that they were more aware of the stakes involved. Using this as an alternative outcome, we find that masking the evaluator's identity indeed leads more colleagues to think the CGCS should get tenured, as reported in Column (1) of Table 10.

Second, we asked each CGCS to report their total monthly remuneration, which includes basic wages and bonuses (if there is any), which we in turn verified using administrative information provided by the provincial government. The basic wage is set by upper-level government and should be exactly the same for all the CGCSs, conditional on the county of residence, enrollment year, and position. Each working unit in turn has some discretion to reward the best performing employees with a modest amount of bonus or allowance (usually very small in the government system). In Columns (2) to (3) of Table 10, we observe that the CGCSs in the masked scheme earn 2% (roughly 49 RMB per month, 0.3 S.D.) 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 2% gap completely reflects the difference in bonus or allowance.

During our field interviews, we were informed that mainly those CGCSs who work in township hospitals and village clinics can earn additional allowance. The reason is that hospitals and clinics have a "business" feature and can keep some profits to reward the most hard-working staffs. For nurses, the most important factor determining their wage differentials is the number of night shifts. In our field works, we learnt that a nurse working in a township hospital could earn additional 5–20 RMB (0.7 to 2.9 USD) for each night shift, with the most common compensation being 10 RMB (1.45 USD) per night. When we restrict the sample to nurses, we find the coefficient gets substantially larger (116 RMB). The compensation differential between the "revealed" group and "masked" group is equivalent to 11 more night shifts. This finding suggests that the improved subjective evaluation is attached to real stakes and could indeed reflect improvement in objective productivity.

VII. Conclusion

Subjective evaluations are widely used in both public and private sectors, especially in contexts lacking sharp measures of employee efforts and performances. A key limitation of subjective evaluation is that it may distort employees' incentives and make them more likely to cater to the preferences of the evaluator, rather than focusing on productive tasks that benefit the organization. However, rigorous empirical evidence on the existence of influence activities is rare and the implications of such activities remain under-explored.

To shed light on these issues, we conduct a large-scale field experiment, where we randomize two performance evaluation schemes among 3,785 junior state employees in China. In the first scheme, we randomly choose one supervisor (from two) as the performance evaluator, and inform the subordinate about the identity of the evaluator *ex ante*. We find that doing so indeed increases evaluator-specific influence activities, as the evaluating supervisor on average gives more positive assessment of the subordinate's performance than the non-evaluating supervisor. Furthermore, other colleagues could correctly predict that the randomly selected evaluator would favor the subordinate (as compared to the non-evaluating supervisor), which indicates that the subordinate's supervisor-specific influence activities could be partly observed by their fellow co-workers.

In the second intervention, the identity of the evaluator is not disclosed to the subordinate, which in theory should encourage the subordinate to reallocate supervisor-specific efforts toward productive dimensions that could be appreciated by both supervisors. We find that this intervention indeed improves the work performance of the subordinate, as measured both by both colleague and supervisor assessments. These results provide causal evidence that reducing influence activities leads to better work performance in the government.

In addition to providing the first rigorous empirical evidence on the existence and implications of influence activities, our findings also have important policy implications. We find that by randomizing the identity of the evaluating supervisor, which has minimal implementation cost, the government could significantly improve the job performance of its employees. This intervention could be adopted to incentivize the 39.5 million state employees in China, and might also apply to broader labor markets given the prevalence of influence activities in workplaces across countries and sectors.

REFERENCES

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).

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.

Guo, Gang. "China's local political budget cycles." *American Journal of Political Science* 53, no. 3 (2009): 621-632.

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. "Environmental Regulation and Firm Productivity in China: Estimates from a Regression Discontinuity Design." (2018).

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 and 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.

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

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.

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

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.

Shirk, Susan L. *The political logic of economic reform in China*. Vol. 24. Univ of California Press, 1993.

Suárez Serrato, Juan Carlos, Xiao Yu Wang, and Shuang Zhang. *The limits of meritocracy: screening bureaucrats under imperfect verifiability*. No. w21963. National Bureau of Economic Research, 2016.

Table 1. Revealing Evaluator's Identity Causes Scoring Asymmetry

	Supervisor 1 Score Minus Supervisor 2 Score		Colleagues Speculate: Supervisor 1 more Positive	
	(1)	(2)	(3)	(4)
Supervisor 1 Evaluating (Revealed)	0.310***	0.310***	0.058***	0.057***
	(0.082)	(0.082)	(0.020)	(0.020)
Outcome Mean	-0.019	-0.019	0.555	0.555
Outcome S.D.	1.299	1.299	0.497	0.497
Controls	N	Y	N	Y
County FE	Y	Y	Y	Y
Type FE	Y	Y	Y	Y
Enrol Year FE	Y	Y	Y	Y
R-Squared	0.160	0.162	0.108	0.110
Observations	1,301	1,301	5,582	5,582

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

Table 2. Masking Evaluator's Identity Eliminates Scoring Asymmetry

	Supervisor 1 Score Minus Supervisor 2 Score		Colleagues Speculate: Supervisor 1 more Positive	
	(1)	(2)	(3)	(4)
Supervisor 1 Evaluating (Masked)	0.003	0.033	0.008	0.009
	(0.113)	(0.114)	(0.019)	(0.019)
Outcome Mean	-0.005	-0.005	0.546	0.546
Outcome S.D.	1.217	1.217	0.498	0.498
Controls	N	Y	N	Y
County FE	Y	Y	Y	Y
Type FE	Y	Y	Y	Y
Enrol Year FE	Y	Y	Y	Y
R-Squared	0.242	0.254	0.175	0.182
Observations	580	580	2615	2615

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

Table 3. Treatment Effects on Colleague Evaluations

	Performance (1-7)		Top 10%	
	(1)	(2)	(3)	(4)
Masking Evaluator's Identity	0.220***	0.185***	0.081***	0.070***
	(0.033)	(0.027)	(0.012)	(0.011)
Colleague Controls	N	Y	N	Y
County FE	Y	Y	Y	Y
Type FE	Y	Y	Y	Y
Enrol Year FE	Y	Y	Y	Y
Obs.	9,256	9,167	9,256	9,167
R-Squared	0.130	0.371	0.084	0.267

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

Table 4. Treatment Effects on Supervisor Evaluations

	Mean (Superior Assessments)	Evaluating Sup's Assessment	Non-Evaluating Sup's Assessment	Supervisor Assessments Deviation
	(1)	(2)	(3)	(4)
Masking Evaluator's Identity	0.138***	0.095*	0.184***	-0.100**
	(0.046)	(0.056)	(0.058)	(0.050)
CGCS Characteristics	Y	Y	Y	Y
County FE	Y	Y	Y	Y
Type FE	Y	Y	Y	Y
Enrol Year FE	Y	Y	Y	Y
Obs.	1,945	1,940	1,940	1,945
R-Squared	0.255	0.210	0.204	0.132

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

Table 5. CGCSs Work Harder and Become More Willing to Leave

	(1)	(2)	(3)	(4)
Panel A. CGCSs Work Harder				
	Work Hard and Overtime		Self Assessment (1-7)	
Masking Evaluator's Identity	0.023**	0.023**	0.081*	0.080*
	(0.012)	(0.012)	(0.048)	(0.048)
Obs.	9,349	9,349	2,771	2,771
R-Squared	0.491	0.491	0.117	0.125
Panel B. CGCSs' Welfare Revealed by Reserved Wage				
	Willingness to Accept Private Sector Job		Reserved Wage (log)	
Masking Evaluator's Identity	0.028*	0.028*	-0.047*	-0.047*
	(0.016)	(0.016)	(0.026)	(0.026)
Obs.	2,735	2,735	2,736	2,736
R-Squared	0.234	0.235	0.233	0.233
Controls	N	Y	N	Y
County FE	Y	Y	Y	Y
Type FE	Y	Y	Y	Y
Enrol Year FE	Y	Y	Y	Y

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

Table 6. Mechanism: Risk Aversion and Relative Importance of Supervisors

	Colleague Evaluation (1-7)			
	(1)	(2)	(3)	(4)
Risk Aversion	-0.044			
	(0.031)			
Masking*Risk Aversion	0.105**			
	(0.053)			
Supervisors' Info. Gap		0.195*		
		(0.101)		
Masking*Supervisors' Info. Gap		-0.365*		
		(0.189)		
Supervisors' Weights Similarity			-0.088	
			(0.069)	
Masking*Weights Similarity			0.212*	
			(0.112)	
Δ in Superiors' Work Assign. Freq.				0.003***
				(0.001)
Masking*Δ in Work Assign. Freq.				-0.002***
				(0.001)
Masking Evaluator's Identity	0.140***	0.526***	0.173***	0.208***
	(0.037)	(0.176)	(0.030)	(0.029)
Controls	Y	Y	Y	Y
County FE	Y	Y	Y	Y
Type FE	Y	Y	Y	Y
Enroll Year FE	Y	Y	Y	Y
Obs.	9,221	8,788	8,770	9,243
R-Squared	0.355	0.361	0.397	0.335

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

Table 7. Ruling Out Alternative Explanations to Influence Activities

	(1)	(2)	(3)
Panel A. Does Supervisor Evaluation Depend on their Awareness of their Roles?			
		Supervisor 1 is More Positive	
Supervisor 1 Eva. (Revealed)	0.068***	0.063**	0.075
	(0.024)	(0.029)	(0.051)
Sample	Full Sample	Supervisor 1 Unaware of being the Evaluator	Supervisor 1 Aware of Being the Evaluator
Obs.	1,502	1,046	380
R-Squared	0.140	0.172	0.281
Panel B. 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 Eva. (Revealed)	-0.010	0.649	0.236
	(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 Eva. (Revealed)	-0.012	-0.010	0.017
	(0.016)	(0.017)	(0.020)
Obs.	1,910	1,910	1,910
R-Squared	0.123	0.146	0.157
County FE	Y	Y	Y
Type FE	Y	Y	Y
Enrol Year FE	Y	Y	Y

Notes: Each column represents a separate OLS regression. Standard errors clustered at work unit level are reported below the coefficients. In Columns (1) and (2), control variables include CGCSs' personal characteristics listed in Table 1. In Columns (3) and (4), control variables include colleagues' personal characteristics listed in Table A2. * significant at 10% ** significant at 5% *** significant at 1%.

Table 8. Alternative Explanations to the Treatment Effects of Uncertainty

	(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
Masking Evaluator Identity	-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. Masking the Identity of the Supervisor Does Not Lead to Information Difference				
	Information Diff. between the Evaluating Supervisors in T1 and T2		Information Diff. between the Non-Evaluating Supervisors in T1 and T2	
Masking Evaluator Identity	0.022	0.016	-0.007	-0.001
	(0.016)	(0.017)	(0.016)	(0.017)
Information Provided by	CGCSs	Colleagues	CGCSs	Colleagues
Obs.	2,839	2,839	2,839	2,839
R-Squared	0.125	0.109	0.114	0.113
County FE	Y	Y	Y	Y
Type FE	Y	Y	Y	Y
Enrol Year FE	Y	Y	Y	Y

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

Table 9. Correlations between CGCS Characteristics and Colleague Evaluation

	Performance (1-7)	
	Colleague	Supervisor
	(1)	(2)
Age	0.049***	0.058***
	(0.011)	(0.015)
Gender	-0.069*	-0.055
	(0.038)	(0.040)
Social Science	-0.005	-0.004
	(0.038)	(0.040)
4-Year College	0.153***	0.129***
	(0.048)	(0.048)
STEM Students	-0.065	-0.009
	(0.041)	(0.042)
Party Member	0.173***	0.143***
	(0.041)	(0.051)
Parent High Sch.	0.033	0.044
	(0.043)	(0.047)
Parent College	-0.030	0.096*
	(0.047)	(0.051)
Work in Village	0.004	0.090*
	(0.060)	(0.051)
CEE Score	-0.038	0.005
	(0.032)	(0.035)
Risk Averse	-0.009	-0.009
	(0.030)	(0.040)
Obs.	8,871	2,556
R-Squared	0.017	0.032

Notes: Each column represents a separate regression. Column (1) reports the correlation between CGCSs performances evaluated by their colleagues and the CGCSs' personal characteristics. Column (2) reports the correlation between CGCSs' performances evaluated by their supervisors and the CGCSs' personal characteristics. Column report the correlation between CGCS's performances evaluated by their colleagues and the colleagues personal characteristics. Standard errors clustered at work unit level are reported in the parentheses. * significant at 10% ** significant at 5% *** significant at 1%.

Table 10. Treatment Effects on "Revealed Preference" Measures

	Qualify for Tenure	log(Wage)	Wage	Wage (Medical Support)
	(1)	(2)	(3)	(4)
Masking Evaluator's Identity (T2)	0.032***	0.020**	48.812**	115.537*
	(0.009)	(0.008)	(22.411)	(61.936)
County FE	Y	Y	Y	Y
Type FE	Y	Y	Y	Y
Enroll Year FE	Y	Y	Y	Y
Obs.	9,349	2,750	2,750	193
R-Square	0.099	0.665	0.641	0.737

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