Institute for Economic Studies, Keio University

Keio-IES Discussion Paper Series

賃金の不平等と同僚との関係性：労働市場の選別に関するフィールドからの新たな事実

スバシシュ・ドゥガル、亀井 憲樹

2025 年 9 月 18 日
DP2025-020
https://ies.keio.ac.jp/publications/26817/

Institute for Economic Studies, Keio University
2-15-45 Mita, Minato-ku, Tokyo 108-8345, Japan
ies-office@adst.keio.ac.jp
18 September, 2025

賃金の不平等と同僚との関係性：労働市場の選別に関するフィールドからの新たな事実

スバシシュ・ドゥガル、亀井 憲樹

IES Keio DP2025-020
2025 年 9 月 18 日
JEL Classification: C93, J31, M52, D81
キーワード: フィールド実験、賃金の公平性、トーナメント、サボタージュ、ソーティング

【要旨】

競争に基づく成果給は生産性を高める一方で、同僚間の負の相互作用を招き、それが労働者の報酬制度に対する選好に影響を及ぼす可能性がある。本研究は、インドの野菜包装労働者を対象としたフィールド実験を通じて、サボタージュ（同僚間の妨害行為）のリスクが労働者の制度選択に与える影響を検証するものである。実験ではまず、報酬総額は同一だが分配の不平等度のみが異なるトーナメントに労働者を外生的に割り当てて作業させ、その後、経験した制度のいずれを好むか選択させることで、内生的ソーティングに関する独自のデータを収集した。公平な第三者による評価が行われる環境では、労働者は報酬格差の大きいトーナメントを選び、不平等や競争を忌避する傾向は見られなかった。これに対し、同僚同士が互いに評価するピア評価の環境では、報酬格差の拡大に伴い過小評価などのサボタージュが急増し、労働者はより公平な制度を選好するようになった。本研究は、労働市場におけるソーティングの根本要因としてサボタージュ・リスクを位置づけるとともに、同僚間の破壊的な相互作用が職場における賃金構造の圧縮を合理化しうることを示している。

スバシシュ・ドゥガル
ユタ大学経済学部
subhasish.dugar@economics.utah.edu

亀井 憲樹
慶應義塾大学経済学部
kenju.kamei@keio.jp

Pay Inequity and Peer Dynamics: New Field Evidence on Labor Market Sorting[1]

Subhasish Dugar[1] and Kenju Kamei [2]

[1] Department of Economics, University of Utah. Email: subhasish.dugar@economics.utah.edu.

[2] Faculty of Economics, Keio University, Japan. Email: kenju.kamei@gmail.com

September 2025

Abstract: Performance pay raises productivity but can also trigger costly peer dynamics, which can influence workers' preferences over pay schemes. We test whether sabotage risk drives compensation choices using a field experiment with Indian vegetable packers. Workers first perform under exogenously assigned tournaments that differ only in pay inequality but are equivalent in total payout, then choose between them, enabling endogenous sorting. Under impartial expert evaluation, workers select steeper tournaments, indicating no aversion to inequality or competition. Under peer evaluation, sabotage escalates sharply with pay dispersion, prompting workers to preemptively prefer more equitable schemes. Our study expands the literature on labor market sorting by identifying sabotage risk as a fundamental driver of sorting and shows how destructive peer dynamics can rationalize compressed wage structures in practice.

JEL classification codes: C93, J31, M52, D81

Keywords: Field experiment, Pay equity, Tournament, Sabotage, Sorting.

[1] This project was supported by a Grant-in-Aid from the Mitsubishi Foundation. The JSPS's "Invitational Fellowship for Research in Japan" program (2023) helped the authors launch the project. The authors thank Debasis Dugar, Prosenjit Mondal, Chiranjit Mondal, Priyanshu Das, Snigdha Roy, Dipanjan Sardar, and Jodipta Sarkar for helping the authors run the field experiment, and Yin Tang for excellent research assistance.

1. Introduction

Performance-based pay, a cornerstone of contract theory, is designed to align worker and firm interests to enhance productivity in the presence of contract enforcement problem (Holmström, 1979), thereby advancing a central goal of firms (e.g., Bolton and Dewatripont, 2004) and addressing core themes in organizational and personnel economics, namely, incentive design (Holmström and Milgrom, 1991; Lazear 2000), labor market sorting (Lazear and Rosen, 1981), and workplace cooperation (Kandel and Lazear, 1992; Ichniowski et al., 1997). Widely implemented through piece rates, tournaments, and revenue-sharing schemes, these performance incentives are pervasive in economic settings like sales, financial services, and manufacturing, leading to substantial intra-firm pay differentials. A rich body of empirical and experimental research confirms their efficacy in stimulating effort and raising productivity (Bull et al., 1987; Ehrenberg and Bognanno, 1990; Eriksson, 1999; Lazear, 2000; Lavy, 2004; Cadsby et al., 2007; Dohmen and Falk, 2011; Leuven et al., 2011; Dechenaux et al., 2015).

Beyond its effects on productivity, performance-based pay is also a key driver of labor market sorting. Prior research, rooted in theory (Lazear and Rosen, 1981; Lazear, 1989) and empirics (Lazear, 2000; Cadsby et al., 2007; Dohmen and Falk, 2011), has established that workers' self-selection into various variable-payment schemes is contingent upon traits such as productivity, ability, risk preferences, gender, and so on.[2] This existing research, however, does not fully account for how performance-based pay schemes, especially tournaments with steep pay differentials, may also shape sorting through the counterproductive peer dynamics it generates in the workplace, namely, sabotage. Despite its central role in theories of tournaments, sabotage has rarely been examined as a determinant of workers' pay scheme choices. Our work provides the first direct evidence on sabotage risk as a mechanism driving labor market sorting.

Tournament-style compensation structures, while designed to encourage effort, can also inadvertently foster sabotage. Workers may seek to improve their relative standing not only through increased effort but by undermining their colleagues – whether by deliberately obstructing coworkers' output (Bandiera et al., 2005, 2007), manipulating performance records or quality

[2] A strand of literature demonstrates that significant pay disparities resulting from performance-based compensation can produce several negative outcomes. These include diminished workplace morale (Breza et al., 2018), reduced effort among those on the losing end of tournaments (Card et al., 2012), elevated rates of employee turnover (Dube et al., 2022), and activities to secure own benefits through distorting information (Milgrom and Roberts, 1990). Chowdhury and Gürtler (2015) and Dechenaux et al. (2015) comprehensively review these effects.

ratings (Harbring and Irlenbusch, 2008, 2011; Carpenter et al., 2010; Kuhn and Villeval, 2015), or retaliating with slowdowns in manufacturing and service jobs (Mas and Moretti, 2009).[3] Evidence shows that the tendency for sabotage is amplified in environments with wider pay disparities (Harbring and Irlenbusch, 2008, 2011; Carpenter et al., 2010). Journalistic accounts also document the occurrence of sabotage in warehouses, garment factories, and gig platforms, where workers tamper with ratings or disrupt tasks to disadvantage their rivals (e.g., The Economist, 2018). Because both perpetrating and enduring sabotage impose significant pecuniary and non-pecuniary costs, workers may prefer tournaments with compressed or comparatively equitable pay differentials, which limit incentives for destructive peer behavior (Lazear, 1989). The central trade-off is thus between the benefits from stronger material incentives and the costs of peer-on-peer sabotage that can compromise fair performance evaluation. This raises a key, yet largely unexplored question: does the risk of sabotage influence workers' sorting into pay schemes?

Despite broad recognition of sabotage as a workplace threat, little is known about whether tournament incentives also drive sorting based on workers' willingness to engage in – or avoid – the risks of sabotage. Research on labor market sorting has identified competitiveness (Niederle and Vesterlund, 2007; Buser et al., 2014; Flory et al., 2015), risk preferences (Eriksson et al., 2009; Dohmen and Falk, 2011), ability (Cadsby et al., 2007; Eriksson and Villeval, 2008), prosociality (Bartling et al., 2009), and honesty (Faravelli et al., 2015) as key drivers of pay scheme choice, while the sabotage literature shows how destructive peer dynamics distort workplace behavior (Harbring and Irlenbusch, 2008, 2011; Carpenter et al., 2010). Yet these strands remain disconnected. We bring these two strands of literature together and provide, to our knowledge, the first direct experimental evidence that sabotage risk may shape sorting into tournament pay schemes. In doing so, we highlight an overlooked dimension of labor market dynamics and show how workers balance strong material incentives against the costs of destructive peer behavior.

To examine how sabotage risk shapes sorting, we conducted a field experiment with professional vegetable packers in rural West Bengal, India, in collaboration with local contractors. Subjects performed packing tasks that closely mirrored their daily work, ensuring high external validity.[4] Workers were randomly assigned to small groups and asked to maximize both the

[3] Sabotage can also stem from status-seeking motives: even if pay is unaffected, performance rankings alone can incentivize workers to undermine their peers (Charness et al., 2014).

[4] Levitt and List (2007) question the external validity of laboratory findings, emphasizing that behavior is shaped not only by financial incentives but also by contextual features such as tasks, subject pools, and role selection. Natural

quantity and quality of produce packed within a fixed time. We implemented three tournament schemes – Small (S), Medium (M), and Large (L) – which differed only in the degree of wage dispersion between the top performer and non-winners, while holding total prize money constant at the group level. Each session unfolded in three stages: in the first two, groups worked under two exogenously assigned schemes with perfect stranger matching; in the third, workers chose between those schemes, sorting endogenously into new groups before repeating the task. Depending on the treatment, individual performance in each stage was evaluated either by peers – allowing for both quantity sabotage (e.g., manipulating output counts) and quality sabotage (e.g., distorting quality ratings) – or by an impartial expert, which eliminated the scope for sabotage. A composite score based on both quantity and quality determined winners and non-winners in each group.

Our design allows for clean identification of the effects of sabotage risk on pay scheme choice. First, by holding total prize money constant across schemes, we isolate preferences over intra-group wage dispersion and associated sabotage risks, ruling out confounds from aggregate earnings potential or group-level efficiency. Second, by varying the evaluation method – peer versus impartial expert – we directly manipulate the scope for sabotage, allowing a clean comparison of how workers weigh the benefits of pay inequality against sabotage risks, using the expert evaluation as a no-sabotage benchmark. Third, rather than relying on stylized mechanisms such as point deductions in computerized lab games (e.g., Harbring and Irlenbusch, 2011), we embed sabotage opportunities in a real production task, enhancing external validity. Fourth, by enabling both quantity and quality sabotage, our design captures how workers navigate distinct forms of destructive peer behavior under different incentive structures that substantially vary pay differential across workers. Together, our design generates rich sabotage and sorting data, providing firms with actionable guidance on how to design tournament schemes that limit sabotage while respecting workers' preferences over pay schemes – all without increasing the wage bill.

Our experiment yields three key findings. First, consistent with standard tournament theory, steeper pay differentials significantly increase effort, as evidenced by expert-assessed output, particularly through higher packing quantity. This productivity gain, however, comes at the expense of reduced packing quality, as workers accelerate their pace under steeper incentives.

---

field experiments, such as ours, that incorporate these dimensions may therefore generate more generalizable insights. Camerer (2011), however, contends that concerns about external validity in lab settings may be overstated, highlighting the unparalleled control they afford, while noting that field experiments often trade off precision and control for realism. For a taxonomy of field experiments and discussion of this trade-off, see Harrison and List (2004).

Under expert assessment, nearly all workers prefer pay schemes with larger intra-group wage dispersion, suggesting no inherent aversion to competition or inequality. Second, under peer evaluation, both quantity and quality sabotage rise sharply with pay differentials, with quality manipulation especially pronounced. These behaviors offset the productivity gains observed under expert assessment, and when given a choice, most workers select more equitable pay schemes to shield themselves from sabotage. Third, in a robustness check treatment that systematically increases the total prize money condition while holding relative pay disparities constant, workers still avoid steep tournaments despite the promise of higher aggregate and individual earnings. This finding demonstrates that workers will forgo even substantial surplus to avoid sabotage-prone environments, underscoring sabotage risk as a decisive factor in pay scheme selection. Together, these findings reveal sabotage risk as a critical driver of workers' pay preferences, with important implications for designing incentives that balance productivity with workplace cohesion.

Our findings carry important implications for both theory and practice. Lazear (1989) argued that while steep pay differentials can raise effort, they also magnify sabotage incentives, implying that compressed pay structures may paradoxically maximize firm profits when peer interference is salient. Our study provides rare field-based confirmation of this logic. When sabotage is infeasible, workers embraced steep pay gradients and delivered higher output. By contrast, when sabotage opportunities were real, workers strategically sorted into more compressed pay schemes – a rational response to heightened sabotage risks, as corroborated by observed sabotage behavior, which eroded the productivity gains seen under expert evaluation. Thus, sabotage risk is demonstrably capable of hurting firm profitability. Importantly, workers' preference for compressed schemes does not reflect aversion to inequality or competition, but pragmatic attempts to shield oneself from destructive peer dynamics. For firms, our findings underscore that the effectiveness of incentive schemes hinges on the monitoring environment: steep tournaments may succeed under reliable external assessment, but in contexts reliant on peer evaluation – or where sabotage is otherwise hard to monitor – compressed pay structures may better sustain cooperation, morale, and productivity. More broadly, our study expands the literature on labor market sorting by identifying sabotage risk as a fundamental determinant of labor market sorting and provides a rationale for why firms and workers themselves may gravitate toward pay compression despite its apparent efficiency costs.

Section 2 reviews related literature. Section 3 describes the experiment and design. Section

4 presents a simple model to generate hypotheses. Section 5 reports empirical results. Section 6 concludes.

2. Related Literature

A large body of work has examined how individual traits and preferences shape compensation scheme choices. Our study advances this literature by examining how workplace environments shape compensation scheme preferences. Prior research highlights diverse determinants of sorting. Niederle and Vesterlund (2007) and Buser et al. (2014) offer laboratory evidence that men, despite comparable ability, are more likely than women to select competitive pay schemes (e.g., tournament over piece rate), often at the cost of expected earnings, with implications for long-run career trajectories. Flory et al. (2015) provide field evidence that women disproportionately avoid competitive workplaces. Other studies highlight the role of risk attitudes. Dohmen and Falk (2011) and Eriksson et al. (2009) offer evidence from laboratory experiments that risk-averse individuals shun winner-take-all tournaments, while risk-tolerant individuals embrace them. Ability and confidence also shape sorting in laboratory experiments: Cadsby et al. (2007) and Eriksson and Villeval (2008) find that high-ability workers disproportionately select variable pay over fixed wages; and Eriksson et al. (2009) document that overconfident workers self-select into tournaments. Sorting has also been linked to moral traits. Faravelli et al. (2015) find in a laboratory experiment that dishonest workers are more likely to select tournaments, while Bartling et al. (2009) show that prosocial individuals opt for cooperative contracts. Collectively, these studies underscore competitiveness, risk tolerance, ability, confidence, honesty, and prosociality as key sorting determinants. However, no prior experimental work – lab or field – has isolated *sabotage risk* as a causal factor shaping pay scheme choice. Our study is the first to address this dimension of sorting, examining whether workers account for the potential of destructive peer behavior when selecting between equitable and unequal pay structures.

A parallel literature has documented how tournaments invite sabotage. We also contribute to this literature. Harbring and Irlenbusch (2008, 2011) demonstrate in laboratory settings that subjects frequently exploit stylized sabotage technologies – such as point deductions – especially as prize spreads widen. Carpenter et al. (2010) move closer to workplace conditions, using a real-effort envelope-stuffing task in which subjects could underreport peers' output or downgrade quality assessments. They find that sabotage erodes tournament incentives: rank-contingent pay schemes improve effort only when sabotage is impossible; once "office politics" are possible, subjects anticipate being sabotaged and scale back their own effort, eroding the incentive effects of tournaments. Reviews by Chowdhury and Gürtler (2015) and Dechenaux et al. (2015) corroborate these findings, identifying sabotage as a persistent and robust feature of tournaments across a range of institutional settings varying crucial elements of decision environments such as prize spreads, feedback policies, sabotage technologies, and cost structures. Yet, aside from Carpenter et al. (2010), most studies operationalize sabotage in abstract forms divorced from production, and they typically capture either *quantity* sabotage (directly undermining output) or *quality* sabotage (biasing evaluations), but not both. Our study advances this literature on three fronts: (i) embedding sabotage opportunities in a natural production activity carried out by professional vegetable packers; (ii) capturing both objective and subjective interference within the same setting; and (iii) varying the mode of evaluation – peer versus impartial expert – to manipulate whether sabotage is feasible at all. This design allows us to identify the causal effects of sabotage opportunities on productivity and, crucially, on workers' sorting into more or less unequal pay schemes, an outcome neglected in prior work.

Finally, our work connects to research on pay compression. Lazear (1989) showed theoretically that when sabotage is possible, compressed pay structures may maximize profits by reducing destructive peer behavior. Breza et al. (2018) provide field evidence from Indian manufacturing that wage compression boosts morale and cooperation, even if it weakens effort incentives. Yet this literature has largely focused on productivity consequences of compression imposed by firms, rather than on whether *workers themselves* prefer compressed pay in the face of sabotage risk. Our study provides the first direct evidence on this point. By allowing professional workers to choose among tournament schemes while holding efficiency constant, we show that workers anticipate sabotage and adjust their sorting accordingly. Thus, we offer a new perspective on why firms may rationally flatten pay not only because compression mitigates sabotage, but also because workers themselves demand it to shield against destructive peer dynamics.

Taken together, our study bridges these three strands of research by integrating sorting, sabotage, and pay compression in a unified framework. Prior work on sorting has emphasized individual traits; research on sabotage has highlighted how destructive peer dynamics erode incentive effects; and studies on pay compression have focused on firm-imposed wage structures. We connect these literatures by showing that sabotage risk is not merely a theoretical concern for firms, but a salient determinant of workers' own compensation preferences. In doing so, we provide the first field-based evidence that workers endogenously sort into compressed or unequal pay schemes depending on the institutional environment that governs sabotage opportunities, thereby uncovering a novel mechanism through which workplace dynamics shape both individual choices and the design of incentive systems.

3. Experimental Design and Implementation

3.1. Location and Subject Recruitment

a. Location

The field experiment was conducted in a cluster of villages within the Basirhat subdivision of North 24 Parganas district, state of West Bengal, India. Spanning approximately 1,777 square kilometers, Basirhat had a population of 2,271,810 according to the last Indian census conducted in 2011. Per the latest census, the population is nearly evenly split between Hindus (51.37%) and Muslims (48.37%), reflecting a diverse cultural landscape. The average literacy rate in Basirhat stands at 86.88%.[5] Bengali is the predominant local language, which we used in our experiment.

The subdivision consists of 617 inhabited villages. We recruited subjects from nine villages randomly chosen from the pool: Gandharbapur, Bajitpur, Katiahut, Aturia, Sayestanagar, Piyara, Tegharia, Basirhat, and Pura. Along with pisciculture, agriculture is the primary occupation in the subdivision, with villagers cultivating a variety of crops such as potatoes, rice, tomatoes, legumes, and seasonal fruits and vegetables, which are exported to other parts of the state, typically in large wooden and jute bags and baskets. A significant portion of the local population earns their primary livelihood from year-round, relatively low-skilled agricultural activities such as planting and harvesting, and fruit and vegetable packaging – a task that subjects in our experiment perform.[6] Further details on the task are provided in Section 3.2.

b. Subject Recruitment

We collaborated with multiple local vegetable packing contractors who provided expertise

---

[5] See https://basirhat.westbengalonline.in/guide/about-basirhat#:~:text=Apart%20from%20agriculture%2C%20other%20businesses,upon%20the%20trading%20done%20.

[6] Although we are unaware of any government data on the incidence of such activities in the area (e.g., picking vegetables, transporting them to local crop collection centers, packing them in wooden baskets, and finally transporting them to markets in the city of Kolkata) or on the percentage of local population that depends on these activities, anecdotal evidence suggests that a substantial portion of the local population heavily relies on these jobs for their daily livelihood. Most contracts in the vegetable and fruit packing sectors are informal in nature (i.e., not based on written signed contracts) and payments are typically made in cash on a daily, weekly, or monthly basis.

in crop selection and operational support for our experiment. As subjects, we selected adults from nearby villages (aged 18 or older) who had at least six months of prior experience in vegetable or fruit packing.[7] Since our experiment involves a tomato-packing task (Section 3.1.a), using experienced vegetable-packers from local villages strengthens the external validity of the study. Consistent with local pay practices, workers are typically compensated with a flat daily wage for each day they report to work or sometimes with a monthly salary. Accordingly, all subjects in our experiment were already familiar with flat-wage payment systems. While there are no formal daily production targets, workers risk termination for failing to fulfill their assigned daily responsibilities, repeated absences, or disruptive conduct. Performance-based incentives such as bonuses or additional pay for meeting seasonal targets are sometimes used in the region. Moreover, factories in the area sometimes offer flat wages that vary according to workers' experience and skill levels. Because performance-based pay is a central element of our design, we ensured that all subjects had prior knowledge of such payment structures. Specifically, during each recruitment cycle, we included as subjects only individuals who answered "yes" to the following question: "Are you familiar with bonuses (i.e., a fixed lump-sum amount paid for meeting a target) in the workplace?" Although this question does not guarantee that subjects were indeed familiar with rank-order tournaments, which are rarely used locally, it still provides a minimal screening mechanism for ensuring subjects understood the link between performance and monetary rewards.

To sign up subjects, we recruited six research assistants (RAs).[8] The research objectives and hypotheses of our study were not disclosed to the RAs throughout the study. Before experimental sessions, the RAs randomly approached households in the selected villages, reading a standardized Bengali script and inviting individuals to participate in an economics research project, which would last no more than 90 minutes, as a part-time job opportunity.[9] They also provided villagers with a specific date and an array of time slots available on that day, from which each villager could choose time slots convenient for them (Each villager was asked to choose

---

[7] In our experiment, 68.9% of the workers were male, and they had on average 3.4 years of experience in packing jobs. According to the 2011 Indian Census, 51.3% of the population in Basirhat was male and 48.7% was female.

[8] Each RA received a fixed sum of 7,000 Indian Rupees (approximately USD 81) as compensation for assisting the researchers with each recruitment and experiment cycle (further details are provided below).

[9] Our study was approved by the University of Utah Institutional Review Board (#IRB_00172037). It was pre-registered on AsPredicted.org with the project title "Sabotage and Endogenous Tournament Selection: A Field Experiment" (AsPredicted #:154075). See Appendix Section A.6 for the subject recruitment letter and Appendix Sections A.1 to A.3 for experimental instructions.

multiple time slots). The research team later allocated a slot to them such that each session had exactly 45 subjects (Section 3.2.e).

We ran a total of three recruitment (experiment) cycles, described in Section 3.2 below. Potential subjects were promised a 200 Rupee show-up fee, with additional earnings dependent on their and other individuals' performance on a task unbeknownst to them at the time of recruitment, and some elements of chance.[10] A rough estimate of the daily wage in the area hovered between Rs. 400 (approximately USD 4.7) and Rs. 600 (approximately USD 7.0) around the experiment dates, as revealed by a pre-experiment survey carried out by us in the Spring of 2023; thus, the compensation subjects received in the experiment was financially salient. We did not provide any estimates of expected additional earnings at the time of recruitment. Upon agreeing to participate, subjects signed a consent form in Bengali. If a subject was found to be illiterate, we did not sign up that subject.

3.2. Experimental Design and Logistics

a. The Task

In the experiment, subjects packed tomatoes within a predetermined time frame. They repeated this packing task three times. We chose tomatoes in the task because they are harvested year-round and transported to Kolkata and nearby regions for sale. Moreover, unlike firmer vegetables such as potatoes, onions, pointed gourds, or bitter gourds, tomatoes possess a relatively softer texture, making them more susceptible to damage from improper packing. According to the contractors, this delicate nature necessitates a higher level of skill to optimize space efficiency while packing, as careful placement is crucial to prevent bruising or spoilage.

To implement the experiment, we procured tomatoes from local markets at wholesale rates. The subjects (experienced vegetable packers) were instructed to pack as many tomatoes as possible into a wooden basket, while also aiming for the highest possible quality within a fixed time frame. While these subjects are accustomed to maximizing quantity in their daily jobs, their wages in their

[10] The State Government mandates a minimum daily wage of approximately 400 Rupees for skilled workers, including those engaged in vegetable and fruit packing industries. Source: https://wblc.gov.in/sites/default/files/upload/min_wages/january-2025/Agril%20&%2015%20others_1.pdf. However, our pre-experiment survey found that wages for fruit and vegetable packaging workers fluctuate between 200 and 500 Rupees per day in the subjects' residence, depending on factors such as factory type, employer-worker relationships, and labor demand.

regular jobs are not contingent on the *quality* of packing, as emphasized in our experiment. Subjects were truthfully informed that the packed tomatoes would be transported for sale to nearby towns and the city of Kolkata, the capital of the state of West Bengal.

b. Treatment Design

This study implemented six main treatments in a 3 × 2 between-subjects design (Table 1.a), such that each subject was assigned to only one treatment condition. The first dimension of the design involves three remuneration schemes: Small (S), Medium (M), and Large (L), which differ in the size of the gap between the payment awarded to the top performer ("winner") and the payments to the other two workers ("non-winners") within a group of three (see Table 1.b and Subsection g for details). In each scheme, the non-winner's payoff is considered fixed pay, while the payoff difference between the winner and a non-winner serves as the bonus for winning. We design three schemes by systematically varying the size of the payoff difference in increments of 300 INR between adjacent schemes: the difference is 200 INR in Scheme S, 500 INR (= 200 + 300) in Scheme M, and 800 INR (= 200 + 2 × 300) in Scheme L. This is achieved by systematically increasing the winner's payoff by 200 INR and decreasing the non-winner's payoff by 100 INR between each pair of adjacent schemes. Importantly, while the payment distribution differs across Schemes S, M, and L, the total size of the monetary pie available to each group remains constant. Each group was assigned a pair of these schemes, resulting in three scheme combinations: SM, SL, and ML. The second dimension of the design concerns who evaluates the workers' performance: either their peers (P) or a third-party expert (E). This yields six treatment conditions in total: "SM-P," "SL-P," "ML-P," "SM-E," "SL-E," and "ML-E" (Table 1.b). In addition to these six, we introduced a seventh treatment, SL-P-EG, where "EG" denotes "efficiency growth," indicating an increase in the total monetary pie size available to the group when the pay differential grows. This treatment allows us to study the effects of increasing the size of the pie while keeping the relative payment structure constant. We provide further details about SL-P-EG in Section 5.3.

c. The Makeshift Laboratory & the Expert

We set up a makeshift experimental laboratory in a temporarily rented four-story building in Gandharbapur. All seven treatments were carried out in this space. The top three floors housed small, private workstations where individual workers completed the tasks (see Appendix A for the floor map and a picture of the makeshift laboratory). Each floor contained 15 workstations.

Table 1: *Experimental Design*

(a) Summary of Treatments

Treatment	Remuneration schemes used	Assessor of performance	# of workers
[Main treatments:]			
SM-P	Scheme S, Scheme M	Peers	90
SL-P	Scheme S, Scheme L	Peers	90
ML-P	Scheme M, Scheme L	Peers	90
SM-E	Scheme S, Scheme M	Expert (unrelated to the group)	90
SL-E	Scheme S, Scheme L	Expert (unrelated to the group)	90
ML-E	Scheme M, Scheme L	Expert (unrelated to the group)	90
[Additional treatment[#1]:]			
SL-P-EG	Scheme S, Scheme L	Peers	90
Total			630

(b) Remuneration Scheme (rank-order tournament)

Remuneration Scheme	Payoff[#2]	
	Winner	Each of the two non-winners
Scheme S (Small pay differentials)	600 INR	400 INR
Scheme M (Medium pay differentials)	800 INR	300 INR
Scheme L (Large pay differentials)	1,000 INR	200 INR

Note: [#1] In addition to the six main treatments pre-registered, one additional treatment, SL-P-EG (SL-P with Efficiency Growth), was conducted for discussion purposes (see Section 5.3 for the details). [#2] Other than the earnings from tomato-packing, each worker received 200 INR as a show-up fee.

Every workstation was equipped with a chair, a wooden basket, a pile of (a fixed number of) tomatoes on the floor, experimental instructions in Bengali, and a pencil. In the four treatments with peer assessments, each subject was also provided with three record sheets.

The ground floor served as a reception and payment area. It also featured a separate room for the expert responsible for evaluating each worker's output along two dimensions discussed below. The expert was a 43-year-old male with 14 years of experience in the vegetable packing industry. The expert was paid ten Rupees to rate and count a worker's output. To ensure the anonymity and privacy of the expert, the expert's room was isolated from the reception area. To prevent any interaction between the subjects and the expert, on each experiment day, the expert arrived much before the subjects and left after all the subjects retreated from the building. We used the same expert for all experimental sessions.

d. The Experimental Cycles

The seven experimental treatments were conducted across three cycles over the course of a year, with approximately six months between each cycle to minimize local dissemination and settling of information about our experiments. The first cycle, held in December 2023, included three treatments in which workers' performance was evaluated by their peers: SM-P, SL-P, and ML-P. The remaining two cycles, June 2024 and November 2024, featured three treatments where performance was assessed by an uninvolved expert: SM-E, SL-E, and ML-E, and the SL-P-EG treatment (peer assessment with an expanded pie). The SL-E and SL-P-EG treatments (the SM-E and ML-E treatments) were implemented in the June (November) cycle.

To minimize information spillovers across cycles, we recruited subjects from different villages for each round of the experiment. The December 2023 cycle drew subjects from the first four villages listed in Section 2.1.a, the June 2024 cycle from the fifth village, and the November 2024 cycle from the last four villages, with each set corresponding to specific treatments.

To preserve experimental integrity, only the authors were aware in advance of the possibility of additional cycles; others involved in the implementation were informed only at the time of their participation. Given the geographic proximity of the villages, some degree of word-of-mouth dissemination was possible. To address this, we rigorously screened subjects in the latter two cycles, excluding individuals who showed any prior knowledge of the study. While it is possible that some research assistants and subjects were hopeful that further cycles might occur, none had access to the specific treatment assignments or design features.

Although elements of the task may have circulated after each cycle, the six-month interval between cycles likely helped to mitigate substantial information spillovers. Furthermore, while the basic task structure and overall compensation format were held constant, key elements such as the method of performance assessment (peer vs. expert) and the magnitude of performance-based bonuses varied across treatments. Through deliberate scheduling, selective disclosure, and recruitment protocols, we aimed to minimize information contamination across treatment groups. No subject took part in more than one treatment condition.

e. The Experiment Structure and Grouping Details

Each treatment consisted of two sessions. Every session involved 45 workers (15 three-person groups). Thus, on the first cycle (December 2023), we ran a total of six sessions (2 sessions/treatment × 3 treatments) involving a total of 270 workers. As intended, the sessions lasted approximately 90 minutes on average. The second cycle (June 2024) and the third cycle (November 2023) differ from the first cycle in one aspect in that the number of treatments was two per cycle. Each of these two cycles involved a total of 180 subjects.

Each session consisted of three tomato-packing stages (Stage One, Stage Two, and Stage Three). For a given treatment, the two 45-subject sessions differed from each other by the sequence of two remuneration schemes in the first two stages. Specifically, the ordering was switched for the two sessions to control for any potential order effects of the schemes. For example, for the SL-P treatment, the first session had 45 subjects facing the remuneration scheme S first (in Stage One), followed by the remuneration scheme L (in Stage Two). The second session of the same treatment had 45 subjects facing the remuneration scheme L first, followed by the remuneration scheme S. The other design elements were identical for the two sessions.

Each experimental session followed a pre-planned sequence of activities, assisted by the six research assistants, with two stationed on each floor. Due to varying literacy rates among the subjects, all instructions were also verbally communicated in Bengali.

Upon arrival, each of the 45 workers was randomly assigned a unique registration ID for identification and payment purposes. They were then escorted to individual workstations. While the experimental session consisted of three stages, one of the three stages would be randomly selected for payment at the end of the experiment. Workers were informed about this random incentive rule before the experiment began. Furthermore, they were told in advance that their earnings would be determined by their performance, their group members' performance, and mutual evaluations of the tomato-packing task in the SM-P, SL-P, ML-P, and SL-P-EG treatments (evaluations made by the uninvolved expert for the tomato-packing task in the SM-E, SL-E, ML-E treatments). They were also made aware that their group members could be present on the same floor as them or on different floors of the building.

In the first two stages, workers were randomly grouped into three-person groups under perfect stranger matching. This means that each worker is grouped with the same person, not more than once. To streamline this matching process, we pre-assigned 45 workers into 15 three-person groups before the experiment using an alphanumeric coding system that was known only to the research team, and not to the experimental subjects. Specifically, each of the 45 workstations was marked with an alpha-numeric code, denoted by F_iJ, where F_i represented the floor number (1, 2, or 3) and where J represented the workstation number (1, 2, …, 15). The numbering system on a given floor followed a counterclockwise pattern: for any given floor F_i, the first workstation on the right was labeled *Fi1*, the second *Fi2*, and so on, continuing counterclockwise until the last workstation on the left, labeled *Fi15* (refer to the floor map for details). This arrangement formed a semicircle, progressing from right to left.

Our pre-experiment matching scheme in the first two stages allowed workers from different floors to be grouped into three-person groups. For example, if workers assigned to workstations F_12, F_113, and F_22 formed a group in a given stage according to the pre-experimental matching, we could identify their corresponding registration IDs. In Stage Three, however, group composition was determined based on the remuneration scheme choices workers made at the beginning of that stage (see Subsection *f*). Still, since we knew the registration IDs of the subjects forming a group in Stage Three, we could conveniently locate which workstations they were working in and identify them by their workstation codes.

f. Subject Decisions per Session

In each stage, the workers' task was to pack as many tomatoes as possible into the basket and maintain recommended levels of quality within a *three-minute* time limit.[11] A pile of 150 tomatoes was placed in each workstation for this task. After a given stage ended, each workstation was restocked with 150 tomatoes before the next stage began. In the treatments with peer assessments, after completing the task, each group member evaluated the other two members' performance based on the following two criteria. In the treatments with expert assessments, the same evaluation was made by the uninvolved expert (Subsection *c*).

[11] A thorough pre-experimental analysis of real-world packing tasks in India revealed that a skilled worker could pack up to roughly 500 tomatoes within a ten-minute time frame. In natural settings, workers typically use much larger baskets, made of wood or jute, capable of holding around more than 1,000 tomatoes, which takes approximately 25 minutes for a skilled packer to fill and comfortably seal the basket. However, to balance efficiency with potential skill variations across workers and to keep the experiment within a reasonable timeframe – we opted for a shorter, three-minute task. We also used a significantly smaller wooden basket designed to accommodate up to 150 tomatoes. Additionally, we coordinated with multiple local contractors to ensure that the packed tomatoes from our experiment would be sent directly to nearby markets for sale.

1. A count of the tomatoes packed by each of the other two group members.
2. A quality rating to each member's work on a scale of 0, 0.1, 0.2, …, 0.9, 1, where 0 indicates a completely unacceptable quality of packing and 1 signifies an outstanding quality of packing the tomatoes (see Appendix A2).

Before each session began, a research team member demonstrated how to pack tomatoes into the basket. Additionally, each worker was shown three reference images on a research assistant's phone, depicting packed baskets rated at 0, 0.5, and 1 (see Appendix A5). Before starting the task, each worker was instructed to sit facing the wall and wait for further directions from the research team. All workers started the task simultaneously and were instructed to stop as soon as the research team announced the end of the three-minute time limit. During this period, communication was strictly prohibited, and research staff members did not intervene. Since each worker faced a wall while performing the task, they could not see others' work.

The evaluation process (by peers and/or the expert) began after the three minutes had passed. In the treatments with peer assessments, the subjects started the peer evaluations along two dimensions: quantity and quality. Since counting would require unpacking a packed basket, we started with the quality dimension. Following our alphanumeric workstation matching scheme, a research assistant carried each worker's packed basket to the workstations of their two peers. Each group member rated the packing quality of the other two members on a scale of 0 to 1. Group members' baskets were identified by the registration ID to maintain anonymity. All baskets were also brought to the room of the expert to provide an objective rating for each basket using the same scale (the subjects were not aware of this rating process). This was used as an unbiased quality measure in the data analysis. In the three treatments with expert assessments, only the expert evaluated the quality of each basket, whose procedure was clearly explained to the subjects.

Once the quality ratings ended for all 15 groups, we repeated the same process as above, whereby each worker (and/or the expert) counted the number of tomatoes packed by their group members. In peer-assessment treatments, each worker recorded the counts and ratings on one of the three identical record sheets provided, sheets designated for each of the three stages. A research team member collected the record sheet immediately after recording at each stage.

The collected record sheets contained information about a worker's registration ID, their quality ratings, and counts for their group members identified by their registration IDs. In the four peer-assessment treatments, once the research team collected the record sheets for all 45 workers, they calculated a composite performance score for each worker using the following formula, which was explained to the subjects before each session began:

The performance score $P = C \times Q$, where,

C (Quantity Score): The average of two counts, each representing the number of tomatoes recorded by a group member for a peer's packing performance in a given stage.

Q (Quality Score): The average of two quality ratings, each provided by a group member to evaluate a peer's packing quality on a scale from 0 to 1, in 0.1 increments, for a given stage.

For example, suppose a worker's two peers recorded that the given worker packed 100 and 120 tomatoes. Then, $C = (100 + 120)/2 = 110$ for that worker in question. If the same worker received quality ratings of 0.6 and 0.8 from the two peers, then Q was equal to $(0.6 + 0.8)/2 = 0.7$ for that worker in question. The performance (P) of that worker was then calculated as $C \times Q = 110 \times 0.7 = 77$. This score captured a merged measure of both the quantity and perceived quality of the given worker's packing performance in our experiment.

In the treatments with expert assessments, P is simply calculated by the product of the expert-assessed C and Q.

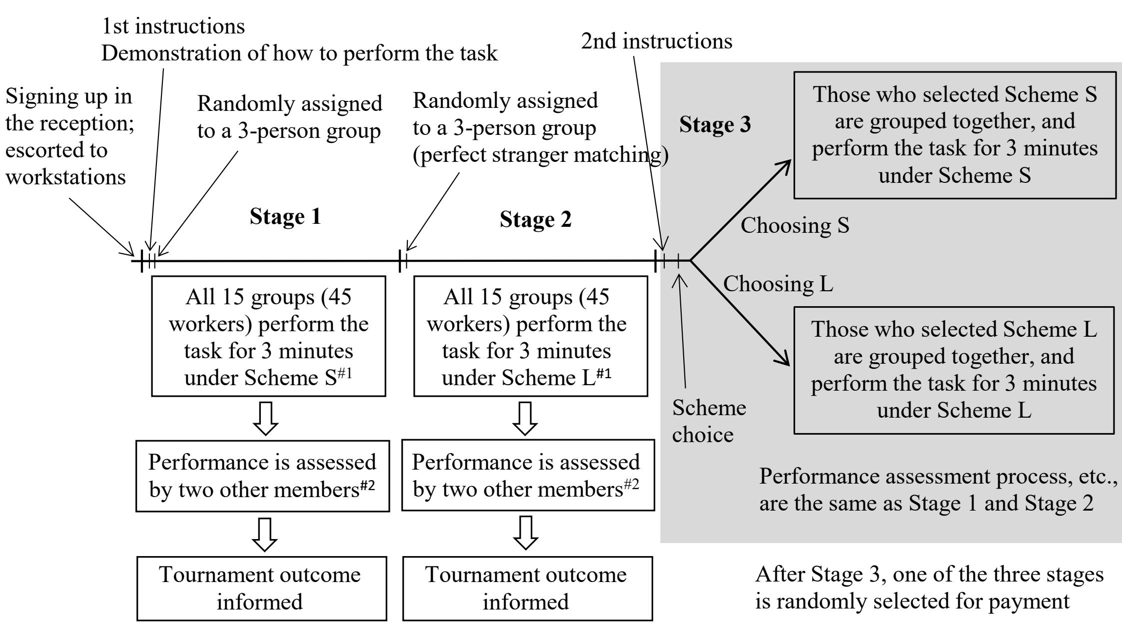
We determined the winner and the two non-winners in each group by comparing the three P measures. The worker with the highest P value in the group was the winner, and the other two were the non-winners. If two or more P measures were identical to each other, a random tie-breaking mechanism determined the winner (see the instructions for details). Finally, we informed each group member of their standing in their group, that is, whether they became the winner in their group at that stage or not. The earnings of the winner and non-winners differed according to which remuneration scheme was used in a given stage (Subsection *g*).

Three stages in each session followed the same procedure, except for the remuneration schemes and grouping. While grouping was randomly constructed for Stage One and Stage Two, it was based on the subjects' choices in the last stage. At the onset of Stage Three, each of the 45 workers selected one of the two previous remuneration schemes to work under in Stage Three, on condition that two other workers who had the same scheme preference would be assigned to the same group.[12] Based on these choices, we formed as many three-person homogeneous groups as possible, ensuring that all members of a group had the same scheme preference. If the number of workers selecting a particular remuneration scheme was not a multiple of three, we rounded to the nearest multiple of three and randomly dismissed the remaining workers for that payment scheme. However, worker dismissals in Stage Three were found to be negligible across all cycles.

After completing all three stages, workers were asked to assemble on the ground floor for payment. A research team member then randomly selected a chip from a box containing three chips labeled 1, 2, and 3. The number drawn determined which stage would be used to calculate actual payments. Based on this, we computed each worker's earnings and placed the corresponding amount of cash into individual envelopes. We arranged all 45 envelopes on a table and asked workers to collect the one corresponding to their private registration ID. Upon exit, each worker completed a brief questionnaire and signed a payment form. Once all workers had departed, a group of independent packers re-packed baskets that did not meet the minimum quality standards for dispatch. The finalized baskets were then loaded onto a truck for delivery to nearby markets.

A schematic diagram for the SL-P treatment, as an example, is provided in Figure 1.

Figure 1: *Schematic Diagram for SL-P Treatment*



Notes: The SL-P treatment is shown above as an example. #1 Another 45 workers experienced Scheme L in Stage One and Scheme S in Stage Two. #2 Performance is assessed by the uninvolved expert in the SL-E treatment.

[12] As years of schooling are low on average among the villagers, it is crucial to avoid cognitive overload to happen in the experiment. We took a gradual learning approach in that we explained the instructions for the first two stages at the onset of the experiment. We gave the instructions for Stage Three only after Stage Two was over.

g. Remuneration schemes

As discussed, three remuneration schemes were used in the experiment. All three schemes are rank-order tournaments in that the best performer receives the highest earnings than the other two in each group, and the two non-winners receive the same earnings. The remuneration schemes differed only by the payoffs of the winner and the non-winners (Table 1.b). In Scheme S (Small pay differentials), Scheme M (Medium pay differentials), and Scheme L (Large pay differentials), the winner's payoffs were Rs. 600, 800, and 1,000, respectively; and each of the non-winners' payoffs was Rs. 400, 300, and 200, respectively. Thus, Scheme S is the most equal scheme, and Scheme L is the most unequal scheme. Notice that the sum of three members' payoffs per group was Rs. 1,400 under all remuneration schemes (= 600 + 400 × 2 = 800 + 300 × 2 = 1,000 + 200 × 2). Hence, economic efficiency was identical for the three remuneration schemes in terms of monetary surplus, but the degree of pay equality differed by the scheme.

4. Theoretical Considerations and Hypothesis

This section presents a simple theoretical model to generate testable hypotheses for our experiment. Specifically, we illustrate how steeper pay differentials, when coupled with opportunities to sabotage co-workers' performance, can lead workers to favor remuneration schemes that promote equity and reduce competition in the workplace.

4.1. Negative Effects of Competitive Workplace

We begin with the assumption that workers are motivated by self-regarding preferences and seek to maximize their own payoffs, a fact that is common knowledge among them. Our theoretical framework and hypotheses build on prior work on rank-order tournaments by Orrison et al. (2004) and Harbring and Irlenbusch (2008, 2011), while incorporating key features specific to our experimental design. In the previous work, each worker's output y_i is modeled as a function of their work effort and a random noise component: $y_i = C_i Q_i + \varepsilon_i$, where noise ε_i is uniformly distributed between $-\eta$ and η and η taking on a fixed value. In our setting, each worker exerts two types of effort: one aimed at maximizing the number of tomatoes packed (quantity), and another aimed at maintaining the quality of packing. C_i refers to the quantity measure, and Q_i to a quality score expressed as a percentage. For notational simplicity, we denote effort as $e_i = (C_i, Q_i)$. While we do not explicitly introduce a noise term in our experiment, workers' outputs are plausibly subject to random variation due to factors such as differences in mood, the quality of baskets, individual ability, time of the day, or other potential measurement errors.

In the peer-assessment treatments (SM-P, SL-P, and ML-P treatments), a worker can influence their peers' performance through underreporting or sabotage. Let $S_{C,i \to k}$ and $S_{Q,i \to k}$ denote the magnitude of worker i's under-evaluation of worker k's count and quality rating, respectively. For simplicity, assume that worker i sabotages both of their peers, j and k, symmetrically: $S_{C,i \to j} = S_{C,i \to k} \equiv S_{C,i}$, and $S_{Q,i \to j} = S_{Q,i \to k} \equiv S_{Q,i}$ (where j and k are the two other workers in the group to which worker i belongs). Let us denote the sabotage vector as $S_i = (S_{C,i}, S_{Q,i})$ for notational convenience. In the peer-assessment treatments, each worker's output is defined as shown in Equation (1):

$$y_i = (C_i - \tilde{S}_{C,i})(Q_i - \tilde{S}_{Q,i}) + \varepsilon_i, \tag{1}$$

$$\text{where } \tilde{S}_{C,i} = \frac{1}{2}(S_{C,j} + S_{C,k}) \text{ and } \tilde{S}_{Q,i} = \frac{1}{2}(S_{Q,j} + S_{Q,k}).$$

Here, $\tilde{S}_{C,i}$ and $\tilde{S}_{Q,i}$ represent the average under-evaluations worker i received from their peers for the quantity (or count) and quality ratings of the tomatoes packed, respectively.

Under this setup, each worker's payoff in SM-E, SL-E, and ML-E treatments is expressed as follows:

$$u_i(e_i, e_{-i}) = M \cdot f(e_i, e_{-i}) + m \cdot \{1 - f(e_i, e_{-i})\} - \alpha g(C_i) - \beta h(Q_i) \tag{2}$$

Each worker's payoff in SM-P, SL-P, and ML-P treatments is expressed as follows:

$$u_i(e_i, S_i, e_{-i}, S_{-i}) = M \cdot f(e_i, S_i, e_{-i}, S_{-i}) + m \cdot \{1 - f(e_i, S_i, e_{-i}, S_{-i})\}$$
$$-\alpha g(C_i) - \beta h(Q_i) - \rho \xi(S_{C,i}) - \gamma \varphi(S_{Q,i}) - \kappa_1 \textstyle\sum_{j \neq i} S_{C,j} - \kappa_2 \sum_{j \neq i} S_{Q,j} \tag{3}$$

In Equations (2) and (3), M and m, where $M > m > 0$, represent the monetary prizes awarded to the winner and the two non-winners in the group, respectively. The values of M and m vary across treatments, as shown in Table 1.b. The term f denotes the probability that worker i wins the tournament, which is a function of worker i's own effort choice e_i, the effort profile of their competitors e_{-i}, and the sabotage decisions of both worker i (S_i) and their competitors' (S_{-i}), relevant in the three peer-assessment treatments. The expressions $\kappa_1 \sum_{j \neq i} S_{C,j}$ and $\kappa_2 \sum_{j \neq i} S_{Q,j}$ (where $\kappa_1 > 0$ and $\kappa_2 > 0$) capture non-material disutilities that worker i incurs if their work is sabotaged by other workers in their group, reflecting frustration, demoralization, or related psychological costs.[13] We

[13] These two terms are not included in Harbring and Irlenbusch (2008, 2011), but their inclusion does not affect predictions, as the size of disutility is determined solely by the competitors' strategy choices.

assume that the cost of exerting effort increases monotonically with the level of effort, and that sabotage is also personally costly. We also assume that the psychological costs of committing sabotage increase with the extent of the two sabotage activities. Formally, we assume increasing marginal costs of quantity and quality of their tomato-packing work, $g'(C_i) > 0$, $h'(Q_i) > 0$, as well as increasing marginal costs of sabotaging others' performance along those two dimensions, $\xi'(S_{C,i}) > 0$, and $\varphi'(S_{Q,i})$, while $\alpha, \beta, \rho, \gamma > 0$.[14] These assumptions align with standard formulations in the behavioral economics literature. The subsections that follow derive the symmetric equilibrium behavior for expert- and peer-assessment treatments.

a. Expert-Assessment Treatments

The optimality conditions that must be satisfied for interior solutions can be derived using the first-order conditions $\frac{\partial u_i}{\partial C_i} = 0$ and $\frac{\partial u_i}{\partial Q_i} = 0$ for Equation (2) as follows:

$$\Delta \frac{\partial f}{\partial C_i} = \alpha \frac{\partial g(C_i)}{\partial C_i}. \tag{4}$$

$$\Delta \frac{\partial f}{\partial Q_i} = \beta \frac{\partial h(Q_i)}{\partial C_i}. \tag{5}$$

Here, $\Delta = M - m$ represents the payoff differential between the winner and each of the two non-winners. In a symmetric equilibrium ($C_i = C_j$ and $Q_i = Q_j$ for all $i \neq j$), Condition (6) must hold for the two marginal probabilities of winning.

$$\frac{\partial f}{\partial C_i} = \frac{Q_i}{2\eta} \text{ and } \frac{\partial f}{\partial Q_i} = \frac{C_i}{2\eta}. \tag{6}$$

The expression $\frac{\partial f}{\partial C_i} = \frac{Q_i}{2\eta}$ holds from the assumption that the noise term ε_i is uniformly distributed between $-\eta$ and η for all i. To see this, consider a small deviation in C_i by worker i. While the density of the uniform distribution is constant at $\frac{1}{2\eta}$, an infinitesimal increase in C_i increases y_i by $\frac{\partial y_i}{\partial C_i} = Q_i$. Holding the effort profiles of the other two workers fixed, this translates into a marginal increase in the probability of winning equal to $\frac{Q_i}{2\eta}$. See Orrison et al. (2004) for the details. By the same reasoning, $\frac{\partial f}{\partial Q_i} = \frac{C_i}{2\eta}$.

As an illustration, assume that the cost functions take the form $g(C_i) = C_i^3$ and $h(Q_i) = Q_i^3$. Given these functional forms, each worker's optimal effort choices can be derived using Equations (4) through (6) as follows:

[14] For simplicity, assume that functional forms and parameters satisfy the second-order condition for a local maximum. Appendix B summarizes the second-order conditions.

$$C_i^* = \frac{\Delta}{6\eta\sqrt[3]{\alpha^2\beta}}. \tag{7}$$

$$Q_i^* = \frac{\Delta}{6\eta\sqrt[3]{\alpha\beta^2}}. \tag{8}$$

It follows that both C_i^* and Q_i^*, and therefore $P_i^* = C_i^* Q_i^*$, increase with the size of the pay differential Δ. However, the extent to which a worker adjusts each component of effort in response to a change in Δ depends on the relative costs of increasing C_i or Q_i.

$$\frac{Q_i^*}{C_i^*} = \sqrt[3]{\frac{\alpha}{\beta}}. \tag{9}$$

Given the assumed cost functions, condition (9) implies that if $\beta > \alpha$, then improving quality is more costly than increasing quantity. Thus, a worker will respond to a larger payoff differential by increasing the number of counts rather than improving the quality. Conversely, if $\beta < \alpha$, improving quality is relatively less costly, and the worker will place greater emphasis on enhancing quality. Which of the two outcome measures, count or quality, responds more strongly to changes in pay differentials in the context of Indian vegetable packers is ultimately an empirical question. As such, no directional prediction can be made ex ante.

Since an uninvolved expert evaluates all workers' outputs, the expert-assessed values of C, Q, and P (= $C \times Q$) serve as proxies for objective performance. The theoretical framework outlined above yields two testable hypotheses. First, the greater the pay differential Δ, the higher the expert-assessed values of both C and Q in the tomato-packing task. Second, despite this performance response, workers are expected to prefer remuneration schemes with smaller pay differentials – regardless of the pair of schemes offered (e.g., they will prefer Scheme S over Scheme L in the SL-E treatment). This preference arises because the expected monetary payoff is constant in equilibrium across all schemes: $M \cdot f^* + m \cdot (1 - f^*) = 1400/3$, where f^* is he equilibrium probability of winning. In contrast, the effort costs increase with the size of the pay differential Δ, which implies that the utility $u_i(e_i^*, e_{-i}^*)$ decreases as Δ increases.

b. Peer-Assessment Treatments

The method for deriving the optimal solution in the peer-assessment treatments is the same as that in the expert-assessment treatments, with one key difference that there are now four types of decision variables: C_i, Q_i, $S_{C,i}$, and $S_{Q,i}$. By applying the first-order conditions $\frac{\partial u_i}{\partial C_i} = 0$, $\frac{\partial u_i}{\partial Q_i} = 0$, $\frac{\partial u_i}{\partial S_{C,i}} = 0$ and $\frac{\partial u_i}{\partial S_{Q,i}} = 0$ to Equation (3), we obtain the following four optimality conditions:

$$\Delta \frac{\partial f}{\partial C_i} = \alpha \frac{\partial g(C_i)}{\partial C_i}. \tag{10}$$

$$\Delta \frac{\partial f}{\partial Q_i} = \beta \frac{\partial h(Q_i)}{\partial C_i}. \tag{11}$$

$$\Delta \frac{\partial f}{\partial S_{C,i}} = \rho \frac{\partial \xi(C_i)}{\partial S_{C,i}}. \tag{12}$$

$$\Delta \frac{\partial f}{\partial S_{Q,i}} = \gamma \frac{\partial \varphi(Q_i)}{\partial S_{Q,i}}. \tag{13}$$

Under a symmetric equilibrium ($C_i = C_j = C$, $Q_i = Q_j = Q$, $S_{C,i} = S_{C,j} = S_C$, and $S_{Q,i} = S_{Q,j} = S_Q$ for all $i \neq j$), the marginal probabilities of winning are given by Equation (14).

$$\frac{\partial f}{\partial C_i} = \frac{Q-\tilde{S}_Q}{2\eta};\ \frac{\partial f}{\partial Q_i} = \frac{C-\tilde{S}_C}{2\eta};\ \frac{\partial f}{\partial S_{C,i}} = \frac{Q-\tilde{S}_Q}{4\eta};\text{ and } \frac{\partial f}{\partial S_{Q,i}} = \frac{C-\tilde{S}_C}{4\eta}. \tag{14}$$

Each of these marginal probabilities can be derived analogously to Equation (6). First, $\frac{\partial f}{\partial C_i} = \frac{Q-\tilde{S}_Q}{2\eta}$ follows from the fact that an infinitesimal increase in C_i increases y_i by $\frac{\partial y_i}{\partial C_i} = Q_i - \tilde{S}_{Q,i}$ while keeping the other two workers' y-values unchanged. Given the uniform noise distribution with density $\frac{1}{2\eta}$, this raises worker i's probability of winning by $\frac{Q_i-\tilde{S}_{Q,i}}{2\eta}$. Likewise, $\frac{\partial f}{\partial Q_i} = \frac{C-\tilde{S}_C}{2\eta}$, since $\frac{\partial y_i}{\partial Q_i} = C_i - \tilde{S}_{C,i}$. Second, an infinitesimal increase in sabotage decreases the y-values of two other peers equally, thereby having an effect similar to an increase in work effort. For example, while $\frac{\partial y_i}{\partial S_{C,i}} = 0$, $\frac{\partial y_k}{\partial S_{C,i}} = -\frac{Q_k-\tilde{S}_{Q,k}}{2}$ for $k \neq i$. Thus, $\frac{\partial f_i}{\partial S_{C,i}} = \frac{Q-\tilde{S}_Q}{2}\frac{1}{2\eta} = \frac{Q-\tilde{S}_Q}{4\eta}$. Similarly, $\frac{\partial f}{\partial S_{Q,i}} = \frac{C-\tilde{S}_C}{4\eta}$. See the supplementary material of Harbring and Irlenbusch (2008) for the formal derivation in the context of rank-order tournaments when both work effort and sabotage are possible.

As an illustration, assume the following cost functions: $g(C_i) = C_i^3$, $h(Q_i) = Q_i^3$, $\xi(S_{C,i}) = S_{C,i}^3$, $\varphi(S_{C,i}) = S_{Q,i}^3$. Under these functional forms, each worker's optimal effort and sabotage levels can be simultaneously derived using Equations (10) through (14), yielding the following expressions:[15]

$$C_i^{**} = \Delta \cdot \Gamma(\alpha,\beta,\rho,\gamma,\eta) \cdot \sqrt{2\rho/\alpha}. \tag{15}$$

$$S_{C,i}^{**} = \Delta \cdot \Gamma(\alpha,\beta,\rho,\gamma,\eta). \tag{16}$$

$$Q_i^{**} = \Delta \cdot H(\alpha,\beta,\rho,\gamma,\eta) \cdot \sqrt{2\gamma/\beta}. \tag{17}$$

$$S_{Q,i}^{**} = \Delta \cdot H(\alpha,\beta,\rho,\gamma,\eta). \tag{18}$$

---

[15] $\Gamma(\alpha,\beta,\rho,\gamma,\eta) = \sqrt[3]{\frac{\left(\sqrt{\frac{2\gamma}{\beta}}-1\right)^2\left(\sqrt{\frac{2\rho}{\gamma}}-1\right)}{1728\eta^3\gamma\rho^2}}$ and $H(\alpha,\beta,\rho,\gamma,\eta) = \sqrt[3]{\frac{\left(\sqrt{\frac{2\rho}{\gamma}}-1\right)^2\left(\sqrt{\frac{2\gamma}{\beta}}-1\right)}{1728\eta^3\gamma^2\rho}}$.

These expressions imply that both effort and sabotage increase with the size of the pay differential Δ, even though the expected monetary payoff remains constant across all remuneration schemes. Therefore, as in the expert-assessment treatments, workers are predicted to prefer remuneration schemes with smaller pay differentials, regardless of the specific pair of schemes presented in the choice set.

Note that in the three peer-assessment treatments, an uninvolved expert also evaluates each worker's performance. This allows us to measure sabotage in count and quality ratings by computing the difference between the expert-assessed and peer-assessed values of C (quantity) and Q (quality). These differences serve as proxies for sabotage and enable us to empirically examine the relationship between pay differentials and sabotage activities.

In sum, the theoretical predictions discussed above are summarized in Hypothesis 1 below. Because all workers complete the same task under two different remuneration schemes in Stage One and Stage Two, Hypotheses 1(a) and 1(b) can be cleanly tested using data from these two exogenously assigned phases.

Hypothesis 1:
(a) In all six treatments, larger within-group pay differentials lead to higher task performance, as measured by the expert-assessed output.
(b) In the three peer-assessment treatments, larger pay differentials result in greater sabotage activities.
(c) Regardless of the specific pair of remuneration schemes offered, workers prefer to perform the tomato-packing task under the scheme with more equitable pay. This preference holds irrespective of whether sabotage is possible or not.

4.2. Alternative Behavioral Hypothesis

The theoretical analysis in Section 4.1 implies that workers' preferences over remuneration schemes are unaffected by the possibility of sabotage. This prediction remains unchanged even after incorporating standard risk preferences or social preferences into the model. However, if workers care about winning for reasons beyond material payoff, such as pride or recognition, then we can have different predictions about scheme preferences. This subsection provides details.

a. Risk and Social Preferences

Hypothesis 1 is based on the assumption of worker risk neutrality, a common assumption in the theoretical literature on tournaments. Importantly, relaxing this assumption to incorporate risk aversion does not change the prediction, as people are generally known to be risk averse (e.g., Zhou and Hey, 2018). For instance, suppose workers have Constant Relative Risk Aversion (CRRA) utility functions of the form: $u(x) = (x^{1-\theta} - 1)/(1 - \theta)$. In this function, a larger degree of risk aversion can be represented by a larger value of θ. Since the expected monetary payoff is constant under symmetric equilibrium across all remuneration schemes in our design, a risk-averse worker's expected utility declines monotonically with the size of the pay differential, Δ. Thus, increasing Δ makes more unequal schemes less attractive.[16]

Second, incorporating social preferences, such as those modelled by Fehr and Schmidt (1999) or as surveyed in Fehr and Schmidt (2006) and Sobel (2005), does not alter the predicted preference ordering either. Because the total group earnings are fixed across all schemes, but inequality differs, workers with inequality aversion will strictly prefer more equitable schemes. Under a symmetric equilibrium, social preference components only strengthen the preference for equity. For example, in the Fehr-Schmidt model, aversion to both advantageous and disadvantageous inequality implies that a worker would rank the schemes as follows: $S \succ M \succ L$. In sum, neither risk aversion nor social preferences undermines the core theoretical prediction: workers prefer more equal pay schemes when expected earnings are held constant.

b. Non-Material Motives for Winning

A growing body of research suggests that individuals care about winning not only for material rewards but also for reasons beyond material gain. Several studies, for instance, propose that workers derive non-material utility simply from winning a tournament or outperforming peers, while others emphasize the importance of relative payoff ranking within a group (see Dechenaux et al., 2015, for a review). In the context of our experiment, such non-material motives for winning can be incorporated into the model by augmenting the utility function. Specifically, we modify the probability of winning by multiplying a factor of $(1 + \mu)$, where $\mu > 0$ represents the non-material motives for winning.[17] Thus, we can rewrite the utility function for the SM-E, SL-E, and ML-E treatments as follows:

[16] A risk-averse preference implies a positive value of θ. For example, suppose that $\theta = 0.5$. The expected utilities from receiving monetary compensation, assuming neither positive effort provision nor sabotage, are calculated as 41.0, 40.0, and 37.9 under Schemes S, M, and L, respectively. This is because each worker has a 1/3 probability of winning the large prize. Thus, the individual consistently prefers a lottery with smaller pay differentials. The same logic summarized in Section 4.1 suggests that greater pay differentials lead to higher values of C_i, Q_i, $S_{C,i}$, and $S_{Q,i}$. Consequently, the total expected utility of the worker is the highest in Scheme S and the lowest in Scheme L. This qualitative implication holds for any positive value of the constant coefficient of relative risk aversion, θ.

[17] Judgment bias is also discussed as one reason (see again Dechenaux et al., 2015 for a survey). That is, people's decisions may reflect non-linear probably weighting such that they overreact to a small probability event of winning

$$u_i(e_i, e_{-i}) = M \cdot (1+\mu) \cdot f(e_i, e_{-i}) + m \cdot \{1 - (1+\mu) \cdot f(e_i, e_{-i})\} - \alpha g(C_i) - \beta h(Q_i)$$
$$= (1+\mu)(M-m) \cdot f(e_i, e_{-i}) + m - \alpha g(C_i) - \beta h(Q_i) \quad (19)$$

Similarly, we can rewrite the utility function for the treatments with peer assessments, SM-P, SL-P, and ML-P, as follows:

$$u_i(e_i, S_i, e_{-i}, S_{-i}) = M \cdot (1+\mu) \cdot f(e_i, S_i, e_{-i}, S_{-i}) + m \cdot \{1 - (1+\mu) \cdot f(e_i, S_i, e_{-i}, S_{-i})\}$$
$$-\alpha g(C_i) - \beta h(Q_i) - \rho \xi(S_{C,i}) - \gamma \varphi(S_{Q,i}) - \kappa_1 \sum_{j \neq i} S_{C,j} - \kappa_2 \sum_{j \neq i} S_{Q,j}$$
$$= (1+\mu)(M-m) \cdot f(e_i, S_i, e_{-i}, S_{-i}) + m$$
$$-\alpha g(C_i) - \beta h(Q_i) - \rho \xi(S_{C,i}) - \gamma \varphi(S_{Q,i}) - \kappa_1 \sum_{j \neq i} S_{C,j} - \kappa_2 \sum_{j \neq i} S_{Q,j} \quad (20)$$

The above expressions imply that incorporating workers' non-material motives for winning is mathematically equivalent to increasing the effective pay differential to $(1 + \mu)(M - m)$. As a result, the equilibrium strategy profile remains as defined in Equations (7), (8), (15), (16), (17), and (18), but scaled by a factor of $(1 + \mu)$. Hence, both work effort and sabotage are predicted to rise with increasing pay differentials, reinforcing the predictions in Hypotheses 1(a) and 1(b).

However, workers' scheme preferences may differ between the two assessment methods, because their expected payoff excluding work- and sabotage-related costs is $(1+\mu)(M - m)/3 + m$; and it is not constant across the treatments, unlike the analysis in Section 4.1, where $(M - m)/3 + m = 1{,}400/3$. For instance, if $\mu = 1$, the expected payoffs are calculated as Rs. 1,600/3, 1,900/3, and 2,200/3 under Schemes S, M, and L, respectively. Thus, these gains when Δ grows could exceed the sum of their additional work and sabotage costs. Whether the gains outweigh the costs is ultimately an empirical question. However, simple algebra suggests that in the expert-assessment treatments, the equilibrium utility, i.e., $u_i(e_i^*, e_{-i}^*)$, increases with the pay differential Δ, as long as Δ is not too large.[18] Whether $\Delta = 200$ (Scheme S), 500 (Scheme M), and 800 (Scheme L) are sufficiently large is again an empirical question.

The attractiveness of greater pay differentials is more readily offset by the associated costs in the peer-assessment treatments than in the expert-assessment treatments. In the three peer-assessment treatments, workers bear costs not only for their work effort but also for engaging in and receiving sabotage. Hence, the presence of the sabotage costs (i.e., $\rho\xi(S_{C,i}^{**}) + \gamma\varphi(S_{Q,i}^{**}) +$

(e.g., Kahneman and Tversky, 1979; Tversky and Wakker, 1995). This possibility can also be modelled the same as the ones expressed in Equations (19) and (20).

[18] Substituting the equilibrium strategy C_i^* and Q_i^* into Equation (19), we obtain that $u_i(e_i^*, e_{-i}^*) = \frac{\mu\Delta}{3} - \frac{(1+\mu)^3\Delta^3}{108\alpha\beta\eta^3}$. Therefore, the partial derivative with respect to Δ is: $\frac{\partial u_i}{\partial \Delta} = \frac{\mu}{3} - \frac{(1+\mu)^3\Delta^2}{36\alpha\beta\eta^3}$, which is positive when $\Delta < \sqrt{\frac{12\alpha\beta\mu\eta^3}{(1+\mu)^3}}$.

$\kappa_1 \sum_{j \neq i} S^{**}_{C,j} + \kappa_2 \sum_{j \neq i} S^{**}_{Q,j}$) leads to the directional hypothesis that the worker's preferences for the more equitable of the two pay structures is stronger under peer assessment than under expert assessment, whether the SM, SL, or ML condition is considered. These behavioral considerations are summarized in Hypothesis 2 below:

Hypothesis 2:

(a) Hypothesis 1(a) continues to hold: greater pay differentials increase task performance.

(b) Hypothesis 1(b) continues to hold: greater pay differentials increase sabotage activity.

(c) Workers' preferences for more equal pay schemes are now expected to vary depending on the method of assessment. Specifically, workers exhibit a stronger preference for equitable pay schemes under peer assessment than under expert assessment. That is, the proportion of workers choosing Scheme S in the SM and SL conditions, and Scheme M in the ML condition, is predicted to be higher in the peer-assessment treatments than in their expert-assessment counterparts.

5. Results

This section presents our main findings. First, we document workers' behaviors in the two exogenous phases (Section 5.1), followed by their sorting decisions in Phase 3 (Section 5.2). Results from an additional treatment are presented for discussion purposes in Section 5.3.

5.1. Work and Sabotage Behavior in Exogenous Phases

Theory predicts that both the expert-assessed count and quality of tomatoes packed should increase with steeper incentives (Section 3). Table 2 summarizes the average values of each objective performance measure. These results indicate that workers strongly responded to incentives. First, consistent with theory, workers packed significantly more tomatoes when pay differentials were larger. This tendency was observed across all treatments (see the "Counts" column). Second, however, quality did not improve with steeper pay differentials, contrary to theoretical predictions. Instead, quality declined due to faster packing speed as performance incentives became steeper. These negative effects were statistically significant at 5% and 1% levels in the SM-E, SL-P, and ML-P treatments. Third, nevertheless, the effects of incentives on quantity (counts) were stronger than those on quality. As a result, larger pay differentials lead to higher total performance as measured by $P = C \times Q$. Significantly positive effects on P were observed in the SL-E, ML-E, and SM-P treatments. Hence, despite some deviations from the predictions, the overall behavioral pattern was consistent with Hypothesis 1(a) and Hypothesis 2(a).

Table 2: *Expert-Assessed Work Performance in Exogenous Phases*

Treatment	Counts (C)[#1]				Quality Rating (Q)				Performance P (= C × Q)			
	Scheme			*p* value for	Scheme			*p* value for	Scheme			*p* value for
	S	M	L	H_0: $\Delta_C = 0$	S	M	L	H_0: $\Delta_Q = 0$	S	M	L	H_0: $\Delta_P = 0$
(i) In the treatments with expert assessment												
SM-E	90.74	**96.59**	---	< 0.001***	**86.33**	81.44	---	0.006***	78.40	78.36	---	0.982
SL-E	89.10	---	**102.62**	< 0.001***	82.33	---	84.78	0.167	73.34	---	**87.11**	< 0.001***
ML-E	---	93.07	**101.44**	< 0.001***	---	78.00	81.89	0.062*	---	72.18	**82.74**	< 0.001***
(ii) In the treatments with peer assessment												
SM-P	86.91	**96.03**	---	< 0.001***	87.00	85.67	---	0.388	75.67	**82.26**	---	< 0.001***
SL-P	87.20	---	**103.28**	< 0.001***	**82.89**	---	71.89	< 0.001***	72.28	---	74.18	0.423
ML-P	---	91.83	**106.57**	< 0.001***	---	**74.56**	67.11	0.004***	---	68.06	71.41	0.237

Notes: The units in the "Counts" and "Quality Rating" columns are the average number of tomatoes packed and the average quality ratings (%), respectively. Δ_C, Δ_Q, and Δ_P represent the average performance differences in C, Q, and P, respectively, between two exogenous schemes under the given treatment. The two-sided *p*-values in each row of the table indicate the test results for the null hypothesis that the performance differences (Δ_C, Δ_Q, or Δ_P) are zero. Each test was conducted using linear regression with robust standard errors. To control for possible order effects, a dummy variable for the SM, SL, or ML order is included in each regression. Numbers are shown in bold when they are significantly different from their paired counterparts. [#1] The research team also recorded the number of each worker's attempts to pack tomatoes ("attempts," hereafter). The expert-assessed counts were almost identical to the number of attempts (see Appendix Table C.1 for details).

*, **, and *** indicate significance at the 0.10 level, at the 0.05 level, and at the 0.01 level, respectively.

The peer-assessed performance data reveal intriguing patterns that differ from those based on expert assessments. Panel (i) of Table 3 summarizes the averages of peer-assessed counts (C), quality ratings (Q), and overall performance P (= $C \times Q$). As with the expert-assessed counts, peer-assessed counts increased with larger pay differentials in each treatment. However, peer-assessed quality ratings declined sharply as the pay differentials increased. According to peer evaluations, the rates of quality decline as the incentive became steeper were 14.6% (= −11.33/77.44), 33.4% (= −25.77/77.17), and 21.8% (= −14.44/66.17) in the SM-P, SL-P, and ML-P treatments, respectively. When considering these opposing effects together using the performance measure P, the total effects were significantly negative and economically substantial in all treatments.

The mechanism underlying the difference between expert- and peer-assessed performance P is that higher incentives to win the tournament led to more active sabotage among peers. Panel (ii) of Table 3 presents the differences between expert- and peer-assessed performance values, which are interpreted as measures of sabotage. These sabotage measures are consistently larger in schemes with greater pay differentials compared to those with smaller differentials. Notably, the sabotage appears much stronger for quality assessments *Q* than for quantity assessments *C*.[19] The negative impact of pay differentials on peer assessments, indicative of sabotage, is consistent with Hypothesis 1(b) and Hypothesis 2(b).

Peer-assessed Ps were used to determine rankings in the three treatments involving peer evaluations. Regardless of the intensity of sabotage, peer evaluations substantially distorted the selection of winners (see Appendix Figure C.1). Specifically, approximately 30% to 40% of groups, dependent on the pay scheme, had rankings based on peer evaluations that differed from those based on expert evaluations (which were conducted secretly) during the exogenous phases.[20]

Table 3: *Sabotage and Peer-Assessed Work Performance in Exogenous Phases*

	Counts (C)				Quality Rating (Q)				Performance P (= C × Q)			
	Scheme			*p* value for	Scheme			*p* value for	Scheme			*p* value for
Treatment	S	M	L	H_0: $\Delta_C = 0$	S	M	L	H_0: $\Delta_Q = 0$	S	M	L	H_0: $\Delta_P = 0$
(i) Peer-assessed performance												
SM-P	84.12	**89.20**	---	< 0.001***	**77.44**	66.11	---	< 0.001***	**65.27**	58.94	---	< 0.001***
SL-P	86.07	---	**98.86**	< 0.001***	**77.17**	---	51.39	< 0.001***	**66.40**	---	50.72	< 0.001***
ML-P	---	84.58	**96.34**	< 0.001***	---	**66.17**	51.72	< 0.001***	---	**55.58**	49.90	0.049**
(ii) Sabotage = Expert-assessed performance (Table 2.ii) minus Peer-assessed performance (Table 3.i above)												
SM-P	2.79	**6.83**	---	< 0.001***	9.56	**19.56**	---	< 0.001***	10.40	**23.32**	---	< 0.001***
SL-P	1.13	---	**4.42**	< 0.001***	5.72	---	**20.49**	< 0.001***	5.88	---	**23.45**	< 0.001***
ML-P	---	7.26	10.22	0.109	---	8.39	**15.39**	< 0.001***	---	12.48	**21.51**	< 0.001***

Notes: The units in the "Counts" and "Quality Rating" columns represent the average number of tomatoes packed and the average quality rating (%), respectively. Δ_C, Δ_Q, and Δ_P represent the average performance differences in C, Q, and P, respectively, between two exogenous schemes under the given treatment. The two-sided *p*-values reported in each row indicate the test results for the null hypothesis that the performance differences (in Panel i) or the difference in the sabotage measures (in Panel ii) are zero. Each test was conducted using linear regression with robust standard errors. To control for possible order effects, a dummy variable for SM, SL, or ML order is included as a control variable in each regression. Numbers are shown in bold when they are significantly different from their paired counterparts in the tests.
*, **, and *** indicate significance at the 0.10 level, at the 0.05 level, and at the 0.01 level, respectively.

[19] To compare the two dimensions of sabotage—counts and quality ratings—a regression analysis was conducted using the percentage loss due to peer assessment as the dependent variable under each condition. As shown in Appendix Table C.2, quality (*Q*) is found to be significantly more affected than quantity (*C*) by peers' sabotage activities in all treatment conditions.

[20] An exception is the SL-P-EG treatment (an additional treatment). In Scheme L'—which features a larger total surplus and greater pay differentials—strikingly, 73.3% of groups experience ranking reversals between peer-evaluation-based and expert-based rankings.

Although workers were unaware that an expert assessed performance (and thus did not know the ranking), they might have been skeptical about the accuracy of peer evaluations, as they had the opportunity to observe their peers' work during the assessment process.

Result 1: *(a) The larger the pay differentials among workers, the higher the expert-assessed performance they achieved, driven primarily by increased "counts."*

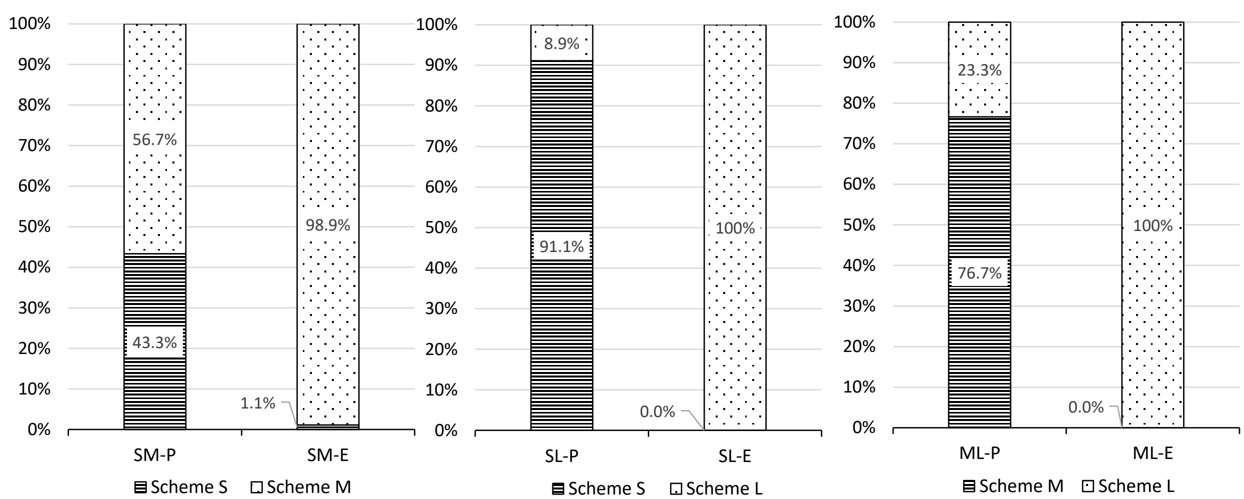
(b) In the SM-P, SL-P, and ML-P treatments, peer-assessed performance P was negatively affected by the size of the pay differentials. The larger the pay differentials, the more intensely workers engaged in sabotage, particularly when rating the quality of their peers' work.

5.2. Labor Market Sorting

The popularity of work environments markedly differs by who assesses performance (Figure 1). On the one hand, when workers are assessed by an uninvolved expert, almost every worker selected the less equitable pay scheme in each condition. The percentages of the workers who chose Scheme M in the SM-E treatment, Scheme L in the SL-E treatment, and Scheme L in the ML-E treatment are 98.9%, 100%, and 100%, respectively.

On the other hand, when peer evaluations were used, workers' preferences for more equitable schemes were strong. Strikingly, 43.3% of the workers in the SM-P treatment, 91.1% of the workers in the SL-P treatment, and 76.7% of the workers in the ML-P treatment selected Scheme S, Scheme S, and Scheme M, respectively. The distributions of workers' scheme preferences are significantly different between the expert and peer assessment conditions at two-

Figure 1: *Scheme Choice by Assessment Method*



(A) SM condition (B) SL condition (C) ML condition sided $p < 0.001$, whether the SM, SL, and ML treatments are considered.[21] This result is consistent with Hypothesis 2(c), not Hypothesis 1(c).

As discussed in Section 3, behavioral predictions based on workers' non-material motives for winning suggest that more intense sabotage activity reduces the attractiveness of the competitive, less equitable pay scheme in each treatment. The reason is that workers incur greater psychological losses when they engage in stronger sabotage and are subjected to more intense sabotage from peers. Recall that, indeed, in the experiment, the more unequal the tournament, the more serious the sabotage activity among peers in each treatment (Table 3. ii, Result 1.b). A regression analysis was conducted to test how experience in the exogenous phases affected workers' scheme choice decisions. Differences in workers' experience of sabotage activity (prior to scheme choices) were constructed as independent variables as follows. First, sabotage rates for counts and quality rating were each calculated using outcome data from Phases 1 and 2:

$$S_{counts}^{Scheme} = (\text{expert-assessed } C - \text{peer-assessed } C)/\text{expert-assessed } C;$$

$$S_{quality}^{Scheme} = (\text{expert-assessed } Q - \text{peer-assessed } Q)/\text{expert-assessed } Q,$$

where the superscript "Scheme" ∈ {unequal, equal} refers to two schemes each worker experienced in Phases 1 and 2. The term "unequal" (or "equal") designates the more unequal (equal) of the two exogenously assigned schemes. For example, in the SM-P treatment, "unequal" refers to Scheme M and "equal" to Scheme S. The differences in the *S* variable between the two schemes are used as independent variables:

$$\text{diff. in sabotage rate for counts } (= S_{counts}^{unequal} - S_{counts}^{equal}).$$

$$\text{diff. in sabotage rate for ratings } (= S_{quality}^{unequal} - S_{quality}^{equal})$$

As no sabotage was possible in the SM-E, SL-E, and ML-E treatments, these two variables are set at 0 for observations in the three treatments (reference group).

Further, two additional factors were included as independent variables. The first is workers' winning experience prior to their scheme choice decisions. For example, if a worker won under an unequal scheme (e.g., Scheme L in the SL-P treatment) in Phase 1 but lost under an equal scheme (e.g., Scheme S in the SL-P treatment) in Phase 2, the worker might be inclined to choose the unequal scheme in Phase 3, influenced by the winning outcomes themselves. Every worker

[21] To control for potential order effects, this is based on a linear regression in which the dependent variable is the worker's scheme choice decision and the independent variables are a peer assessment dummy and an order dummy (SM, SL, or ML order dummy). Two-sided Fisher exact tests were also performed, finding that the differences in the distribution of the workers' scheme preferences in the SM, SL, and ML conditions are all significant at $p < 0.0001$.

experienced two different schemes exogenously. Accordingly, two dummy variables were added to the regressions: "Won in the more unequal scheme" and "Won in the more equal scheme." The second factor is the workers' ability to react to changes in incentives. Hypothesis 1(a) and Hypothesis 2(a) propose that individuals exert greater effort under the more unequal pay schemes than the more equal ones. This prediction was supported by the experimental results (Result 1.a). However, the ability to respond to incentive changes may vary across individuals. Therefore, an ability measure was constructed by calculating the difference in the expert-assess performance P between the two exogenous phases (= $EP^{unequal} - EP^{equal}$) for each worker. This measure was included as an independent variable in the analysis.

The regression results are summarized in Table 4. Since sabotage targeting counts and quality ratings were both affected by the magnitude of pay differentials (as shown in Table 3), only one of the "diff. in sabotage rate" variables was included as an independent variable. The results show that the differences in sabotage rate for counts and for quality ratings are both positive predictors for workers' selection of the more equal pay scheme. In particular, peer behavior aimed at under-evaluating workers' quality under the more unequal scheme led the workers to opt for the more equal pay scheme. This finding is parallel to our earlier observation that sabotage related to quality ratings was more pronounced than sabotage related to counts (Table 3).

The results also show that those with good packing skills (i.e., those who can respond to larger pay differentials by increasing expert-assessed P) were more likely to choose the more unequal scheme in the SL-P and ML-P treatments. Lastly, workers' winning experience before their choices had little effect on their choice of work environment.[22]

Result 2: *(a) When expert evaluations were used (i.e., in the SM-E, SL-E, and ML-E treatments), almost all workers selected the more unequal pay scheme in Phase 3.*
(b) In sharp contrast, in the SM-P, SL-P, and ML-P treatments, where peer evaluations were used, 43.3%, 91.1%, and 76.6% of workers, respectively, selected the more equal pay scheme from the available options in Phase 3.
(c) The tendency to shy away from more unequal pay schemes in (b) was driven by sabotage from peers, the intensity of which increased with the size of the pay differential.

[22] One might think that workers prone to sabotage would prefer less equitable pay schemes and then engage in a high degree of sabotage to improve their rankings in the competition. However, this possibility was not supported by our data. Focusing on the three peer-assessment treatments, we ran the same regression, this time additionally including a variable capturing workers' sabotage behavior—measured as the average under-evaluation of their peers' performance relative to an impartial expert's assessment—during the exogenous phases. In all three treatments, the estimated coefficient on the sabotage variable was not significant (the results are omitted for brevity).

Work and sabotage behaviors in Phase 3 are not the primary focus of this paper. Nevertheless, the behavioral patterns were analyzed and are summarized in Appendix Table C.3. It revealed that, as in the exogenous phases, greater pay differentials among peers led to more intense sabotage by workers. However, one notable difference emerges: In Phase 3 of the three treatments with peer evaluations, workers did not respond to incentive changes. That is, the observed expert-assessed performance (*P*) did not increase with the size of pay differentials. This may suggest that workers anticipated being under-evaluated by their peers regardless of their effect, given that peer evaluations and mutual sabotage determine tournament outcomes.

Table 4: *Determinants of Workers' Sorting*

Dependent variable: A dummy which equals 1 if worker *i* sorted into the more equal scheme in Phase 3 (i.e., a worker selected Scheme S in the SM-E, SM-P, SL-E, and SL-P treatments, and Scheme M in the ML-E and ML-P treatments); 0 otherwise

Treatment:	SM-P and SM-E		SL-P and SL-E		ML-P and ML-E	
Independent variable:	(1)	(2)	(3)	(4)	(5)	(6)
diff. in sabotage rate for counts (= $S_{counts}^{unequal} - S_{counts}^{equal}$)[#1]	2.399*** (0.838)	---	3.212*** (0.654)	---	0.610 (0.380)	---
diff. in sabotage rate for quality rating (= $S_{quality}^{unequal} - S_{quality}^{equal}$)[#1]	---	1.007*** (0.368)	---	1.471*** (0.210)	---	0.666*** (0.235)
diff. in expert-assess P (= $EP^{unequal} - EP^{equal}$)	-0.002 (0.003)	-0.001 (0.003)	-0.008*** (0.002)	-0.008*** (0.001)	-0.006*** (0.002)	-0.005*** (0.002)
Won in the more unequal scheme	0.030 (0.073)	0.015 (0.072)	0.144* (0.083)	0.217*** (0.065)	0.104 (0.080)	0.101 (0.076)
Won in the more equal scheme	0.026 (0.072)	0.038 (0.071)	-0.063 (0.078)	-0.159** (0.067)	-0.075 (0.081)	-0.055 (0.072)
Constant	0.164*** (0.051)	0.157*** (0.051)	0.400*** (0.063)	0.328*** (0.057)	0.319*** (0.058)	0.268*** (0.064)
Observations	180	180	180	180	180	180
F	1.73	1.74	10.46	28.03	4.43	6.88
Prob > F	0.1295	0.1272	0.0000	0.0000	0.0008	0.0000
R-squared	0.0647	0.0693	0.1721	0.3955	0.1171	0.1556

Notes: Linear regressions with robust standard errors. To control for possible order effects, a dummy variable for SM order, SL order, or ML order was included as a control in each regression (the estimates are omitted to conserve space).
[#1] The two diff. measures are zero for the SM-E, SL-E, and ML-E treatments, as sabotage was not possible in these cases.
*, **, and *** indicate significance at the 0.10 level, at the 0.05 level, and at the 0.01 level, respectively.

5.3. Concerns about Sabotage Drive Workers' Selection of Inefficient Environments

The original six treatments were designed such that the sum of three workers' earnings in each group was the same. That is, the two available pay schemes in each treatment were equivalent in terms of total material surplus. Under this condition, peer evaluations strongly discouraged workers from choosing the less equitable scheme due to sabotage activities by their peers (Figure 1). This raises the question: might workers even forgo efficiency gains because they wish to avoid sabotage-prone environments? If the answer is yes, it suggests that firms may benefit from implementing mechanisms to curb sabotage (e.g., by adopting industrial policies that compress pay disparities among workers [Lazear, 1989], or screening out aggressive individuals prone to sabotage during the hiring process [Lazear, 1995]). Such a measure could make firms more attractive to high-performing talent.

To investigate this further, an additional treatment, referred to as the SL-P-EG treatment, was conducted to examine how workers sort themselves when the total group earnings differ across schemes. The SL-P-EG treatment is identical to the SL-P treatment, except for the pay schemes used. The two schemes in the additional treatment are referred to as S' and L'. Each payoff under Scheme S' is 25% lower than its counterpart under Scheme S, while each payoff under Scheme L' is 25% higher than its counterpart under Scheme L (see Panel A of Table 5). The total earnings for the three group members are Rs. 1,050 under Scheme S' and Rs. 1,750 under Scheme L'. Thus, the efficiency is around 67% (= (1,750 – 1,050)/1,050) higher under Scheme L'. Note that the loser's payoffs are nearly identical between the two schemes (Rs. 250 or Rs. 300), while the winner's payoffs differ substantially. Thus, in the absence of sabotage, Scheme L' would be more attractive than Scheme S'.

The experiment results showed that, strikingly, 60% of the workers selected Scheme S', even though it was materially much less attractive than Scheme L'. The underlying reason was the presence of intense sabotage activities under Scheme L'. Panel B of Table 5 summarizes workers' performance and sabotage behavior prior to their scheme choice decisions. The observed patterns are similar to those found in the SL-P treatment (Table 3). Specifically, workers packed significantly more tomatoes under Scheme L' than under Scheme S'. However, their peers gave lower quality ratings to work done under Scheme L'.

The degree of sabotage in quality rating was remarkably high: the under-evaluation rate was calculated as (44.89 – 73.72)/73.72 × 100% = 39.1% under Scheme L'. As a result, the average peer-assessed performance P was significantly lower in Scheme L' than in Scheme S'. Thus, the psychological costs of being severely sabotaged led many workers to shy away from Scheme L', even though it was superior in terms of material rewards.

Result 3: *Workers shied away from the more unequal scheme (Scheme L') in the SL-P-EG treatment, despite its clear material advantage over the more equal scheme (Scheme S').*

Table 5: *Incentive Structures and Results for Phases 1 and 2 in the SL-P-EG treatment*

(A) Incentive Structure

Remuneration Scheme	Payoff[#2]	
	Winner	Each of the two losers
S' (S with efficiency loss)	450 (= 600×0.75) INR	300 (= 400×0.75) INR
L' (L with efficiency gain)	1,250 (= 1,000×1.25) INR	250 (= 200×1.25) INR

(B) Work Performance and Sabotage Behavior in Exogenous Phases

Counts (C)			Quality Rating (Q)			Performance P (= C × Q)		
Scheme		*p* value for	Scheme		*p* value for	Scheme		*p* value for
S'	L'	H_0: $\Delta_C = 0$	S'	L'	H_0: $\Delta_Q = 0$	S'	L'	H_0: $\Delta_P = 0$
(i) Expert-assessed performance								
85.06	**105.16**	< 0.001***	**82.56**	71.33	< 0.001***	70.23	**74.54**	0.021**
(ii) Peer-assessed performance								
82.76	**98.36**	< 0.001***	**73.72**	44.89	< 0.001***	**61.07**	43.61	< 0.001***
(iii) Sabotage = Expert-assessed performance (Panel i) minus Peer-assessed performance (Panel ii)								
2.29	**6.79**	< 0.001***	8.83	**26.44**	< 0.001***	9.16	**30.93**	< 0.001***

Notes: The units in the "Counts" and "Quality Rating" columns are the average number of tomatoes packed and the average quality ratings (%), respectively. Δ_C, Δ_Q, and Δ_P represent the average performance (or sabotage) differences in C, Q, and P, respectively, between two exogenous schemes under the given treatment. The two-sided *p*-values in each row of the table are the test results for the null hypothesis that the performance (or sabotage) differences are zero. Each test was performed based on linear regressions with robust standard errors. To control for potential order effects, a dummy variable for SM order, SL order, or ML order is included as a control in each regression. The numbers are shown in bold when they are significantly different from their paired counterparts in the tests.
*, **, and *** indicate significance at the 0.10 level, at the 0.05 level, and at the 0.01 level, respectively.

6. Discussion and Conclusion

This study offers the first experimental evidence on how sabotage risk shapes workers' preferences over tournament pay schemes and, in turn, drives endogenous sorting in labor markets. By embedding sabotage opportunities within a realistic production task performed by professional vegetable packers, we isolate the causal role of adverse peer evaluations in influencing sorting behavior under steeper incentives. In a high-validity field setting, we show that, under impartial assessment, workers demand steeper pay schemes and competition as predicted by classic tournament theory. By contrast, when peer evaluation enables sabotage, workers retreat from unequal schemes and select pay structures with greater equity even when this entails foregoing significant material benefits. Thus, we shed light on a previously overlooked factor of selection into performance-based pay schemes: workers' willingness to tolerate destructive peer dynamics.

Our results illuminate the pay compression literature, offering field corroboration of Lazear's (1989) theoretical insight that equitable structures may optimize profits when sabotage is salient, by curbing peer interferences. While Breza et al. (2018) demonstrate compression's morale benefits in Indian manufacturing via firm imposition, we show workers actively demand it under peer reliance, even at efficiency costs. This endogenous preference provides a microfoundation for observed compression in sabotage-vulnerable settings (e.g., gig platforms; The Economist, 2018), suggesting firms' adoption may reflect not only supply-side responses but also worker-driven demand-side preferences. These patterns reveal sabotage not merely as a frictional response to incentives but as a pivotal mechanism influencing compensation choice and workplace morale.

This study bridges two strands of the literature on performance pay, advancing our understanding of both sabotage and sorting. The empirical sabotage literature, especially impressive studies by Harbring and Irlenbusch (2008, 2011) and Carpenter et al. (2010), establishes that wider tournament prize spreads intensify sabotage, undermining both output and fairness in workplace competitions. These studies show that wider prize spreads exacerbate destructive peer interactions, whether through stylized point deductions (Harbring and Irlenbusch, 2008, 2011) or manipulations in real-effort tasks like envelope stuffing (Carpenter et al., 2010). Our findings extend this body of work to a field context with professional subjects, where sabotage manifests in both objective (quantity underreporting) and subjective (quality downgrading) forms within the same production process. Our design also manipulates the feasibility of interference directly via peer versus expert assessment. This yields cleaner identification: sabotage offsets incentive effects on productivity under peer evaluation, mirroring Carpenter et al.'s (2010) observation of effort withdrawal in anticipation of interference, but in a setting with verifiable objective benchmarks. Moreover, by quantifying quality sabotage's outsized role, reaching 39% underreporting in our efficiency-gain treatment, we highlight how subjective evaluations, common in sales or service roles, may amplify tournaments' dark side beyond quantity-based interference. Our results thus go further by documenting that workers actively forgo more lucrative but sabotage-prone environments, reflecting direct psychological and strategic costs of peer interference that prior work on sabotage often inferred but rarely measured in sorting decisions.

Equally, we contribute to the sorting literature by incorporating sabotage risk as a heretofore unexamined driver of pay scheme selection. Existing work emphasizes, among other factors, risk tolerance (Dohmen and Falk, 2011; Eriksson et al., 2009) and preferences for competition (Niederle and Vesterlund, 2007; Buser et al., 2014) as key sorting determinants often in stylized lab environments where sabotage was infeasible by design. In contrast, our experiment endogenizes sorting after exogenous exposure to tournament schemes, revealing how lived experience of sabotage – absent in these studies – reorients preferences toward equity. Under no-sabotage conditions, workers overwhelmingly embrace wage dispersion, aligning with risk-neutral predictions of equivalent expected earnings across schemes (Orrison et al., 2004). Yet, sabotage's salience inverts this: workers forgo up to around 67% higher group surplus to avoid interference-prone tournaments, a magnitude exceeding typical risk-aversion effects documented in lab sorting (Eriksson et al., 2009). This underscores a pragmatic calculus – balancing material gains against the costs of perpetrating and enduring sabotage – distinct from the trait-based mechanisms in prior sorting research. By holding total prizes constant while varying dispersion, we also rule out confounds from aggregate efficiency, sharpening the link between interference risk and pay scheme choice. Our finding concerning pay scheme choices also runs counter to standard risk-aversion prediction. If widening prize spreads increase income variance, risk-averse workers should gravitate toward more equitable schemes. Instead, we observe the opposite: workers selected steeper tournaments. Hence, competitiveness, but not risk preferences or inequality aversion, can explain the behavior observed under impartial evaluation. Our findings, thus, extend the sorting literature by establishing sabotage risk as a causal driver of pay scheme selection.

Our findings also qualify standard tournament theory and extend the literature by demonstrating that sabotage risk is a decisive factor in sorting. Whereas much empirical and experimental research has focused on how competitive pay schemes drive output or attract "the best" workers, our results highlight that the institutional environment of evaluation – peer vs. impartial third party – radically shapes both individual preferences and competitive dynamics. Workers' avoidance of inequitable schemes under peer evaluation suggests that the optimal design of incentives depends critically on who controls performance measurement, and the extent of mutual interference permitted. In contexts where sabotage is hard to monitor, pay compression may promote workplace cohesion and sustain productivity even at the cost of reduced incentives. Our results thus bridge several strands of economic theory – sorting, sabotage, and pay compression – and provide actionable guidance for firms seeking to design labor contracts that balance incentives and cooperation. The findings reveal that compensation scheme preference cannot be decoupled from the social environment of evaluation: ensuring fair outcomes requires careful attention not only to incentive structures but also to institutional mechanisms that limit sabotage. These insights advance understanding of the interplay between material rewards and social costs in organizational design, offering new perspectives for researchers and practitioners.

Limitations warrant note. Our design, while externally valid for low-skill packing, may not generalize to knowledge-intensive roles where sabotage is subtler (e.g., idea theft). Repeated interaction or reputation effects, absent here due to stranger matching, could further modulate behavior. Future work might explore heterogeneous responses by risk or prosocial traits, or test interventions like monitoring to mitigate sorting distortions.

In sum, sabotage risk emerges as a core friction shaping labor market outcomes, compelling workers to trade competitive vigor for cooperative equity. By unveiling this mechanism, our study enriches contract theory's emphasis on incentive alignment (Bolton and Dewatripont, 2004), underscoring the need for designs that safeguard against peers' hidden costs.

References:

Auriol, E., Friebel, G., & Pechlivanos, L. (2002). Career concerns in teams. *Journal of Labor Economics*, *20*(2), 289-307.

Bandiera, O., Barankay, I., & Rasul, I. (2005). Social preferences and the response to incentives: Evidence from personnel data. *Quarterly Journal of Economics*, 120(3), 917–962.

Bandiera, O., Barankay, I., & Rasul, I. (2007). Incentives and the allocation of authority in organizations. *Organization Science*, 18(3), 349–363.

Bartling, B., Fehr, E., Marechal, M., Schunk, D. (2009). Egalitarianism and Competitiveness. *American Economic Review*, 99(2): 93–98.

Bolton, P., & Dewatripont, M. (2004). Contract theory. MIT Press.

Breza, E., Kaur, S., & Shamdasani, Y. (2018). The morale effects of pay inequality. *Quarterly Journal of Economics*, 133(2), 611–663.

Bull, C., Schotter, A., & Weigelt, K. (1987). Tournaments and piece rates: An experimental study. *Journal of Political Economy*, 95(1), 1–33.

Buser, T., Niederle, M., & Oosterbeek, H. (2014). Gender, competitiveness, and career choices. *Quarterly Journal of Economics*, 129(3), 1409–1447.

Cadsby, C. B., Song, F., & Tapon, F. (2007). Sorting and tournament compensation: Theory and experimental evidence. *American Economic Journal: Microeconomics*, 2(4), 66–85.

Camerer, C. (2011). The promise and success of lab-field generalizability in experimental economics: A critical reply to Levitt and List. *Available at SSRN 1977749*.

Card, D., Mas, A., Moretti, E., & Saez, E. (2012). Inequality at work: The effect of peer salaries on job satisfaction. *American Economic Review*, 102(6), 2981–3003.

Carpenter, J., Geber, S., & Matthew, B. (2010). Tournaments and sabotage: Evidence from the lab. *Journal of Economic Behavior & Organization*, 76(3), 359–372.

Charness, G., Masclet, D., & Villeval, M. C. (2014). Competitive preferences and status: Experimental evidence. *Journal of Economic Behavior & Organization*, 99, 180–190.

Chowdhury, S. M., & Gürtler, O. (2015). Sabotage in contests: A survey. Public Choice, 164, 135–155.

Dohmen, T., & Falk, A. (2011). Performance pay and multidimensional sorting: Productivity, preferences, and gender. *American Economic Review*, 101(2), 556–590.

Dube, A., Freeman, R. B., & Reich, M. (2022). Do corporations pay workers less than competitive market rates? *American Economic Review*, 112(8), 2510–2558.

Dechenaux, E., Kovenock, D., & Sheremeta, R. (2015). A survey of experimental research on contests, all-pay auctions, and tournaments. *Experimental Economics*, 18(4), 609–669.

Ehrenberg, R. G., & Bognanno, M. L. (1990). The incentive effects of tournaments revisited: Evidence from the European PGA Tour. *Industrial & Labor Relations Review*, 43(3), 74–88.

Eriksson, T. (1999). Executive compensation and tournament theory: Empirical tests on Danish data. *Journal of Labor Economics*, 17(2), 262–284.

Eriksson, T., Teyssier, S., & Villeval, M. C. (2009). Self-selection and the efficiency of tournaments. *Economic Inquiry*, 47(3), 530–548.

Eriksson, T., & Villeval, M. C. (2008). Other-regarding preferences and performance pay: An experiment on incentives and sorting. *Journal of Economic Behavior & Organization*, 68(2), 412–421.

Faravelli, M., Friesen, L., & Gangadharan, L. (2015). Selection and moral hazard in competitions: Evidence from a laboratory experiment. *Journal of Economic Behavior & Organization*, 118, 50–66.

Fehr, E., & Schmidt, K. M. (1999). A theory of fairness, competition, and cooperation. *The Quarterly Journal of Economics*, *114*(3), 817-868.

Fehr, E., & Schmidt, K. M. (2006). The economics of fairness, reciprocity, and altruism–experimental evidence and new theories. *Handbook of the economics of giving, altruism and reciprocity*, *1*, 615-691.

Flory, J. A., Leibbrandt, A., & List, J. A. (2015). Do competitive workplaces deter female workers? A large-scale natural field experiment on job entry decisions. *Review of Economic Studies*, 82(1), 122–155.

Harbring, C., & Irlenbusch, B. (2008). How many contestants are optimal? Group size and rent-seeking in tournaments. *Journal of Economic Behavior & Organization*, 65(3–4), 682–702.

Harbring, C., & Irlenbusch, B. (2011). Sabotage in tournaments: Evidence from a laboratory experiment. *Management Science*, 57(4), 611–627.

Harrison, G. W., & List, J. A. (2004). Field experiments. *Journal of Economic Literature*, *42*(4), 1009-1055.

Holmström, B. (1979). Moral hazard and observability. *The Bell journal of economics*, 74-91.

Holmstrom, B., & Milgrom, P. (1991). Multitask principal–agent analyses: Incentive contracts, asset ownership, and job design. *The Journal of Law, Economics, and Organization*, *7*(special issue), 24-52.

Ichniowski, C., Shaw, K., & Prennushi, G. (1997). The effects of human resource management practices on productivity: A study of steel finishing lines. *American Economic Review*, 87(3), 291–313.

Kahneman, D. A. N. I. E. L., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, *47*(2), 363-391.

Kandel, E., & Lazear, E. P. (1992). Peer pressure and partnerships. *Journal of Political Economy*, 100(4), 801–817.

Konrad, K. A. (2009). Strategy and dynamics in contests. Oxford University Press.

Kuhn, P. J., & Villeval, M. C. (2015). Should we distrust the review process in science? Laboratory experiments on sabotage in science. *Journal of Economic Behavior & Organization*, 116, 469–482.

Lavy, V. (2004). Performance pay and teachers' effort, productivity, and grading ethics. *American Economic Review*, 94(5), 1979–2011.

Lazear, E. P. (1989). Pay equality and industrial politics. *Journal of Political Economy*, 97(3), 561–580.

Lazear, E. P. (1995). *Personnel Economics*. MIT Press.

Lazear, E. P. (2000). *Personnel economics for managers*. Wiley.

Lazear, E. P., & Rosen, S. (1981). Rank-order tournaments as optimum labor contracts. *Journal of Political Economy*, 89(5), 841–864.

Leuven, E., Oosterbeek, H., & Sloof, R. (2011). Incentives for teachers: A meta-analysis. *Economics of Education Review*, 30(3), 449–462.

Levitt, S. D., & List, J. A. (2007). What do laboratory experiments measuring social preferences reveal about the real world?. *Journal of Economic Perspectives*, 21(2), 153-174.

Mas, A., & Moretti, E. (2009). Peers at work. *American Economic Review*, 99(1), 112–145.

Milgrom, P., Roberts, J. (1990). Bargaining Costs, Influence Costs, and the Organization of Economic Activity. pp. 57-89 in *Perspectives on Positive Political Economy* (eds. J. Alt and K. Shepsle), Cambridge University Press.

Niederle, M., & Vesterlund, L. (2007). Do women shy away from competition? Do men compete too much? *Quarterly Journal of Economics*, 122(3), 1067–1101.

Orrison, A., Schotter, A., & Weigelt, K. (2004). Multiperson tournaments: An experimental examination. *Management Science*, *50*(2), 268-279.

Sobel, J. (2005). Interdependent preferences and reciprocity. *Journal of Economic Literature*, *43*(2), 392-436.

The Economist. (2018, December 22). Warehouse wars: The secret sabotage of gig workers. Retrieved from https://www.economist.com/business/2018/12/22/warehouse-wars-the-secret-sabotage-of-gig-workers

Tversky, A., & Wakker, P. (1995). Risk attitudes and decision weights. *Econometrica: Journal of the Econometric Society*, 1255-1280.

Wall Street Journal. (2020, March 15). Sabotage on the shop floor: How workers disrupt competitors. https://www.wsj.com/

Zhou, W., & Hey, J. (2018). Context matters. *Experimental economics*, *21*(4), 723-756.

Online Supplementary Material 1 for Dugar and Kamei (2025)

Appendix A: Instructions and Logistics

As the literacy rate was low among the vegetable packing workers in the villages where the experimental sessions were conducted, all the instructions were verbally communicated to the workers (subjects) in Bengali. All our workers were native speakers of the Bengali language. One of the authors (Dugar) is native in Bengali, and he directed the field experiment sessions carefully in Bengali. At least six research assistants who were also native speakers of Bengali supported the workers as necessary, and helped any aspect of logistics (e.g., bringing a worker's packed baskets to his/her group members' workstation for counts and quality rating). The following is the English translation of those verbal instructions.

A.1. Instructions for the Treatments with Peer Assessments

[The following is the instructions for the SM-P treatment as an example:]

Welcome to today's work.

General Rules:

Please pick a piece of paper from the basket. The piece of paper you just picked has an alpha-numeric code written on it. This code is your registration ID. Each of you has an ID randomly assigned to you. Each ID code is different from the rest of the codes. This ID is your private information. Please do not share (or show) your registration ID with (to) anybody other than the research team members. We will use this registration ID to identify you during today's work. Upon request, show your registration ID to a research team member who will escort you to a workstation. Today's work will approximately take about 90 minutes of your time.

For your participation you will receive 200 Rupees. In addition, you can earn a considerable amount of money depending on your own performance on a task, the performances of two other participants on the same task, other decisions you and two other participants will make, and partly on chance. Below, we explain the task and other decisions you and two other participants will have to make in today's work. Your earnings, including the participation fee of 200 Rupees, will be privately paid to you in cash at the end of today's work. If you leave before the whole work is over, you will not receive any money. Today's work is designed so that the other participants cannot trace your performance on the task, the other decisions you will make, or the additional money you will earn from today's work back to your real identity. Your performances and decisions in today's work will remain anonymous. In other words, all the tasks you will complete and the decisions you will make in today's work will remain confidential information between you and the researchers. To maintain anonymity and privacy of all your choices, we will only use your ID when we must disclose your performances and other decisions in today's work to other participants. No one other than the researchers will ever be able to make a connection between your private registration ID and your real identity in today's work.

During today's work, you are not allowed to talk or communicate with anybody, verbally or via physical gestures. Please switch off your mobile phone. Any violation of these rules will disqualify you from participating in today's work and receiving any money. If you have any questions during today's work, please raise your hand and a research team member will privately assist you. We will not start today's work until all the participants have understood all the instructions.

{Half of the 90 workers work under Scheme S (M), and the other half work under Scheme M (S) in Stage One (Two). The following are the instructions for those confronted with Scheme S (M) in Stage One (Two).}

Three Stages & Three-Person Groups:

Today's work will proceed in three stages: Stage One, Stage Two, and Stage Three. As explained below, you will work on the same task and make similar decisions in each stage. Forty-four other participants, excluding you, are participating in today's work. Thus, there are 45 participants. In each stage, we will randomly form 15 three-person groups. In each of the first two stages, you will be randomly matched with two other anonymous participants to form a three-person group. In other words, the other two members of your group will change between Stage One and Stage Two. In the interest of time, we have already created 15 random three-person groups using 45 participants for Stage One and Stage Two. In Stage Three, who will be the other two members of your group will be determined by a specific choice you will make at the beginning of Stage Three, which we will explain below. Throughout all stages, you will not know the true identities of your group members, and they will not know yours either. In each stage, we will present you with the performances and decisions of two other group members, but only their registration IDs will be visible, not their real identities. Similarly, your performance and decisions in each stage will be shared with the other two group members using your registration ID instead of your actual identity.

The details of the task and the decisions you will make in a specific stage will be explained to you before that stage begins. For example, the Stage Two task and the decisions will be explained to you once the Stage One task and the decisions end.

15 People on A Specific Floor: {See Appendix A.4 for the floor image}

The building has four floors. We will use the top three floors for today's work. In each stage, 15 participants or five three-person groups will occupy a specific floor. Which 15 participants will occupy a specific floor in a stage will depend upon the registration IDs. Please keep in mind that the other two members of your group may be on a different floor during any of the three stages.

The Makeshift Workstation:

Although all three group members will perform the same task in each stage, all of them will work on the task in the same makeshift workstation throughout today's work. A research team member will escort you to a workstation assigned to you. In the workstation, you will find a chair, a pen, a wooden basket, a pile of tomatoes on the floor, and two record sheets. Unless directed, please sit in the chair facing the wall and wait for instructions from one of our research team members on when to start working on the task, which is explained below.

The Payment:

As previously mentioned, you will receive 200 Rupees for participating today. Additionally, you can earn additional money by performing a task and making other decisions in each of the three stages. However, your payment will only be based on one of these stages. At the end of today's work, a research team member will randomly draw one of three pieces of paper labeled 1, 2, or 3 from a pot without looking. The number selected – 1, 2, or 3 – will correspond to Stage One, Stage Two, or Stage Three, respectively, and will determine which stage your payment will be based on. For example, if the number drawn is 2, you will receive the amount earned in Stage Two, reflecting your performance on the task and other decisions as well as those of the other two group members during that stage. In this case, you would earn your 200 Rupee show-up fee plus any additional earnings from Stage Two, which will depend on the performances and other decisions you and your group members made at that stage. All the earned money will be paid to you privately in cash at the end of today's work (you will be asked to write your name and signature on a payment form). Below, we describe the task and other decisions you will make in each stage, and then we explain the payment schemes.

The Task:

In each of the three stages, all of you will perform the same task described below:

• Packing tomatoes in a wooden basket both of which have been provided to you.

Note that the packed tomatoes will be shipped to the city for sale.

Stage One:

After a research team member demonstrates how this task should be completed, you will perform the task **for 10 minutes**. You will find a pile of tomatoes on the floor in your workstation. Within 10 minutes, you can pack as many tomatoes as possible in the wooden basket. The research team member will wait outside the room while you work on the task. As soon as the 10-minute time window is over, the research team member will ask all of you to stop working on the task.

At the end of 10 minutes, a research team member will come to your workstation and temporarily take away your wooden basket packed with tomatoes while you sit in the workstation facing the wall. We will show your packed basket to your two group members from that stage. Each of your group members will make two decisions based on your wooden basket packed with tomatoes:

- Each of them will rate the quality of your output (i.e., the total number of tomatoes you packed) and record them on the record sheet provided to them. The quality rating we will use for each group member will be a percentage between 0% and 100% (with increments of 10%), where 0% indicates that the quality of the output is definitely unacceptable, and 100% indicates that the quality of the output is definitely acceptable. There are 11 possibilities for rating: {0%, 10%, 20%, …, 100%}.
- Each of them will also count and record on the record sheet how many tomatoes you packed.

A research team member will collect the record sheets immediately once each of them records them. Similarly, we will bring each of your group members' packed baskets to your cubicle after Stage One. You will rate the quality of your group members' outputs and count how many tomatoes they packed. You will record them on the record sheet provided to you. A research team member will collect the record sheet immediately once you record them.

We have provided you with two record sheets, one for each of the other two members of your group. Do not write your name on any of the record sheets you are given. Only write down your registration ID in the space provided at the top of the record sheet. We do this to maintain anonymity. By only using your registration ID, the other group members **will not be able to** link your name to your assessment of their output, and the research team member will also never be able to link your decisions to you by name.

We will calculate each of your performance in Stage One according to the following formula:

$$\text{Your performance, } P\text{, in Stage One} = N \times Q,$$

where,

N = the average count by the other two group members of the number of tomatoes you packed.

Q = the average quality rating you received from the other two group members.

More specifically, N and Q will be computed as follows.

$$N = \frac{\text{the first group member's count + the second group member's count}}{2}$$

$$Q = \frac{\text{the first group member's quality rating+ the second group member's quality rating}}{2}$$

For example, suppose that the first and second group members recorded that you packed 100 and 120 tomatoes within 10 minutes. Then, $N = (100+120)/2 = 110$. Also, suppose that the first and second group members rated the quality of your work as 60% and 80%, respectively. Then, $Q = (60\%+80\%)/2 = 70\%$. Your performance in this example will be calculated as $N \times Q = 110 \times 0.7 = 77$. Thus, your count is reduced by 30% as the quality of your work is not perfect.

If your P number is higher than the other two group members' P numbers, your earnings will be 600 Rupees. Together with the show-up fee of 200 Rupees, if this stage is selected for payout, **you will receive a total of 800 Rupees**.

If your P number is not the highest in your group, your earnings will be 400 Rupees. Together with the show-up fee of 200 Rupees, if this stage is selected for payout, **you will receive a total of 600 Rupees**.

If your P number is the highest, but it equals another group member's P number, we will flip a coin, and if heads come up, you will receive a total of 800 Rupees. If Tails come up, you will receive a total of 600 Rupees. If the P numbers are the same for all three members, then the member with the highest number in their alpha-numeric code or registration ID in their group will receive a total of 800 Rupees, whereas each of the other two members will receive a total of 600 Rupees.

For example, if SM41, SM12, and SM23 form a group and their *P* numbers are the same, then the group member with SM41 will receive a total of 800 Rupees, whereas each of the other two members, SM12 and SM32, will receive a total of 600 Rupees.

After Stage One is over, we will inform you of the total number of tomatoes you packed in Stage One, the other two group members' counts of your output, the other two group members' quality ratings of your output, your value of *N*, your value of *Q*, your value of *P*, and whether you won 800 or 600 Rupees in Stage One.

Do you have any questions? Once all questions are answered, we will move to Stage Two.

Stage Two:

You will be randomly matched with two different group members at this stage.

The task in Stage Two is the same as that in Stage One. You will again be given 10 minutes to perform your task. We will count your *P* number in the second stage according to the same formula as in Stage One.

However, your earnings in this stage will be determined differently from how they were determined in Stage One, which we explain below.

If your *P* number is higher than the other two group members' *P* numbers, your earnings will be 800 Rupees. Together with the show-up fee of 200 Rupees, if this stage is selected for payout, **you will receive a total of 1,000 Rupees**.

If your *P* number is not the highest in your group, your earnings will be 300 Rupees. Together with the show-up fee of 200 Rupees, if this stage is selected for payout, **you will receive a total of 500 Rupees**.

If your *P* number is the highest, but it equals another group member's *P* number, we will flip a coin, and if heads come up, you will receive a total of 1,000 Rupees. If Tails come up, you will receive a total of 500 Rupees. If the *P* numbers are the same for all three members, then the member with the highest number in their alpha-numeric code or registration ID in their group will receive a total of 1,000 Rupees, whereas each of the other two members will receive a total of 500 Rupees. For example, if SM41, SM12, and SM23 form a group and their *P* numbers are the same, then the group member with SM41 will receive a total of 1,000 Rupees, whereas each of the other two members, SM12 and SM32, will receive a total of 500 Rupees.

After Stage Two is over, we will inform you of the total number of tomatoes you packed in Stage Two, the other two group members' counts of your output, the other two group members' quality ratings of your output, your value of *N*, your value of *Q*, your value of *P*, and whether you won 1000 or 500 Rupees in Stage Two.

Do you have any questions? Once all questions are answered, we will move to Stage Three.

Summary of Stages One and Two:

In each of the first two stages, you will be randomly assigned to a group of three and pack as many tomatoes as possible within 10 minutes.

Your two group members in Stage Two are different from your two group members in Stage One.

Your performance in each stage is determined by your two group members' assessments on (a) count of the number of tomatoes you packed and (b) the quality of your work.

Your earnings (show-up fee plus the earnings from the task) depend on whether your performance is the best in your group or not, as follows:

Table 1: Your earnings

	When your performance is the best in your group	When your performance is not the best in your group
Stage One	800 Rupees	600 Rupees
Stage Two	1,000 Rupees	500 Rupees

Do you have any questions? Once the first two stages end, we will distribute instructions for Stage Three.

{Once Stage Two is over, the following instructions are explained:}

Stage Three:

Stage Two has now ended.

In Stage Three, you can choose whether to perform the same task under the Stage One or Stage Two payment scheme. You will be provided with a choice sheet. The choice sheet will list the first two stages of payment schemes. Next to each payment scheme, you will find a box. If you would like to perform the same task under the Stage One payment scheme in this stage, then tick the box next to the Stage One payment scheme. If you would like to perform the same task under the Stage Two payment scheme in this stage, then tick the box next to the Stage Two payment scheme.

If you choose to perform under the Stage One payment scheme in this stage, you will be randomly matched with two other participants who also chose to perform the task under the Stage One payment scheme in this stage, just like you. On the other hand, if you choose to perform under the Stage Two payment scheme in this stage, you will be randomly matched with two other participants who also chose to perform the task under the Stage Two payment scheme in this stage, just like you.

Once everyone completes their choice, one of the research team members will collect all the choice sheets. After that, a staff member will indicate that you can begin performing the same task within the 10-minute time interval. We will count your P number in Stage Three according to the same formula as in Stage One or Stage Two. However, your earnings in this stage will be determined

based on your choice of payment scheme. If you choose to perform under the Stage One payment scheme in this stage, your payment will be determined just like under the Stage One scheme.

On the other hand, if you choose to perform under the Stage Two payment scheme in this stage, your payment will be determined just like under the Stage Two scheme. The summary of the payment scheme can be found in Table 1 in the instructions for Stage Two. You can also go back and re-read the payment schemes of these two stages in detail for the instructions.

After Stage Three is over, we will inform you of the total number of tomatoes you packed in Stage Three, the other two group members' counts of your output, the other two group members' quality ratings of your output, your value of N, your value of Q, your value of P, whether you won in Stage Three or not.

Today's work will be over once everyone completes the Stage Three task. A research team member will escort all of you to a different room where you will wait for the payments. Please allow us some time to calculate your earnings from all three stages. After calculating the payments, a staff member will come to the waiting room and show all of you a pot containing three numbers: 1, 2, or 3. The staff member will randomly pick one of the three numbers without looking at the pot. The randomly selected number, which can be 1, 2, or 3, will determine which of the three stages we will use to pay you. We will put your earnings from that stage inside an envelope. Each envelope will have a registration ID written on it. Therefore, all envelopes will have registration IDs randomly assigned to you at the beginning of the work. We will put these envelopes on a table showing the registration IDs. On your way out of the building, please pick up your envelope showing your registration ID. After you have picked up your envelope, a staff member will ask you to complete a payment form asking for your name and signature. This will conclude today's work for you.

Thank you for your participation!

Do you have any questions?

<u>Remark</u>: An expert—a 43-year-old male with 14 years of experience in vegetable packing and recruitment for similar roles—was hired for 10 Rupees to rate and count a worker's output secretly. The same expert was used in all sessions without revealing his identity to the workers. Each floor of the building had a room in one of which the expert was stationed. The expert arrived before the workers and left the building after all the workers were sent home. Each worker's output was initially presented to the expert for a quality rating. Subsequently, the packed tomatoes from the same worker were taken to two group members for evaluation. The same output was then brought back to the expert's room for counting before being sequentially returned to the group members for their count. The workers were not made aware of this process. As discussed in this paper, the expert's ratings and counts are used to estimate the extent of sabotage.

A.2. Assessment Criterion

As explained in Section A.1, each worker's output quality is rated on a scale of 0 to 100 (%).

Classification	**Points**
Outstanding	100
Exceptional	90
Excellent	80
Very Good	70
Good	60
Competent	50
Weak	40
Poor	30
Very Poor	20
Incompetent	10
Unacceptable	0

A.3. Instructions for the Treatments with Expert Assessments

[The following is the instructions for the SM-E treatment as an example:]

Welcome to today's work.

General Rules:

Please pick a piece of paper from the basket. The piece of paper you just picked has an alpha-numeric code written on it. This code is your registration ID. Each of you has an ID randomly assigned to you. Each ID code is different from the rest of the codes. This ID is your private information. Please do not share (or show) your registration ID with (to) anybody other than the research team members. We will use this registration ID to identify you during today's work. Upon request, show your registration ID to a research team member who will escort you to a workstation. Today's work will approximately take about 90 minutes of your time.

For your participation you will receive 200 Rupees. In addition, you can earn a considerable amount of money depending on your own performance on a task, the performances of two other participants on the same task, other decisions you and two other participants will make, and partly on chance. Below, we explain the task and other decisions you and two other participants will have to make in today's work. Your earnings, including the participation fee of 200 Rupees, will be privately paid to you in cash at the end of today's work. If you leave before the whole work is over, you will not receive any money. Today's work is designed so that the other participants cannot trace your performance on the task, the other decisions you will make, or the additional money you will earn from today's work back to your real identity. Your performances and decisions in today's work will remain anonymous. In other words, all the tasks you will complete and the decisions you will make in today's work will remain confidential information between you and the researchers. To maintain anonymity and privacy of all your choices, we will only use your ID when we must disclose your performances and other decisions in today's work to other participants. No one other than the researchers will ever be able to make a connection between your private registration ID and your real identity in today's work.

During today's work, you are not allowed to talk or communicate with anybody, verbally or via physical gestures. Please switch off your mobile phone. Any violation of these rules will disqualify you from participating in today's work and receiving any money. If you have any questions during today's work, please raise your hand and a research team member will privately assist you. We will not start today's work until all the participants have understood all the instructions.

{Half of the 90 workers work under Scheme S (M), and the other half work under Scheme M (S) in Stage One (Two). The following are the instructions for those confronted with Scheme S (M) in Stage One (Two).}

Three Stages & Three-Person Groups:

Today's work will proceed in three stages: Stage One, Stage Two, and Stage Three. As explained below, you will work on the same task and make similar decisions in each stage. Forty-four other participants, excluding you, are participating in today's work. Thus, there are 45 participants. In each stage, we will randomly form 15 three-person groups. In each of the first two stages, you will be randomly matched with two other anonymous participants to form a three-person group. In other words, the other two members of your group will change between Stage One and Stage Two. In the interest of time, we have already created 15 random three-person groups using 45 participants for Stage One and Stage Two. In Stage Three, who will be the other two members of your group will be determined by a specific choice you will make at the beginning of Stage Three, which we explain below. Throughout all stages, you will not know the true identities of your group members, and they will not know yours either. In each stage, we will present you with the performances and decisions of two other group members, but only their registration IDs will be visible, not their real identities. Similarly, your performance and decisions in each stage will be shared with the other two group members using your registration ID instead of your actual identity.

The details of the task and the decisions you will make in a specific stage will be explained to you before that stage begins. For example, the Stage Two task and the decisions will be explained to you once the Stage One task and the decisions end.

15 People on A Specific Floor:

The building has four floors. We will use the top three floors for today's work. In each stage, 15 participants or five three-person groups will occupy a specific floor. Which 15 participants will occupy a specific floor in a stage will depend upon the registration IDs. Please keep in mind that the other two members of your group may be on a different floor during any of the three stages.

The Makeshift Workstation:

Although all three group members will perform the same task in each stage, all of them will work on the task in the same makeshift workstation throughout today's work. A research team member will escort you to a workstation assigned to you. In the workstation, you will find a chair, a pen, a wooden basket, a pile of tomatoes on the floor, and two record sheets. Unless directed, please sit in the chair facing the wall and wait for instructions from one of our research team members on when to start working on the task, which is explained below.

The Payment:

As previously mentioned, you will receive 200 Rupees for participating today. Additionally, you can earn additional money by performing a task and making other decisions in each of the three stages. However, your payment will only be based on one of these stages. At the end of today's work, a research team member will randomly draw one of three pieces of paper labeled 1, 2, or 3 from a pot without looking. The number selected – 1, 2, or 3 – will correspond to Stage One, Stage Two, or Stage Three, respectively, and will determine which stage your payment will be based on. For example, if the number drawn is 2, you will receive the amount earned in Stage Two, reflecting your performance on the task and other decisions as well as those of the other two group members during that stage. In this case, you would earn your 200 Rupee show-up fee plus any additional earnings from Stage Two, which will depend on the performances and other decisions you and your group members made at that stage. All the earned money will be paid to you privately in cash at the end of today's work (you will be asked to write your name and signature on a payment form). Below, we describe the task and other decisions you will make in each stage, and then we explain the payment schemes.

The Task:

In each of the three stages, all of you will perform the same task described below:

• Packing tomatoes in a wooden basket both of which have been provided to you.

Note that the packed tomatoes will be shipped to the city for sale.

Stage One:

After a research team member demonstrates how this task should be completed, you will perform the task **for 10 minutes**. You will find a pile of tomatoes on the floor in your workstation. Within 10 minutes, you can pack as many tomatoes as possible in the wooden basket. The research team member will wait outside the room while you work on the task. As soon as the 10-minute time window is over, the research team member will ask all of you to stop working on the task.

At the end of 10 minutes, a research team member will come to your workstation and temporarily take away your wooden basket packed with tomatoes while you sit in the workstation facing the wall. We will show your packed basket to an expert who has 14 years of experience in the vegetable packing industry. We will never reveal the expert's identity to you, nor will the expert know your identity. We will show your packed basket to the expert without showing your registration ID. We do this to maintain anonymity. The expert will make two decisions based on your wooden basket packed with tomatoes:

- The expert will rate the quality of your output (i.e., the total number of tomatoes you packed) and record it on the record sheet provided to the expert. The quality rating we will use for the expert will be a percentage between 0% and 100% (with increments of 10%), where 0% indicates that the quality of the output is definitely unacceptable, and 100% indicates that the quality of the output is definitely acceptable. There are 11 possibilities for rating: {0%, 10%, 20%, …, 100%}.
- The expert will also count and record on the record sheet how many tomatoes you packed.

A research team member will collect the record sheet immediately once the expert records it and show it to you.

We will calculate each of your performance in Stage One according to the following formula:

Your performance, P, in Stage One = $N \times Q$,

where,

N = the expert count of the number of tomatoes you packed.

Q = the expert quality rating you received from the expert.

More specifically, N and Q will be computed as follows.

$N = \text{ the expert count of the number of tomatoes you packed}$

$Q = \text{ the expert quality rating you received from the expert}$

For example, suppose the expert recorded that you packed 110 tomatoes within 10 minutes. Then, $N = 110$. Also, suppose that the expert rated the quality of your work as 70%. Then, $Q = 70\%$. Your performance in this example will be calculated as $N \times Q = 110 \times 0.7 = 77$. Thus, your count is reduced by 30% as the quality of your work is not perfect.

If your P number is higher than the other two group members' P numbers, your earnings will be 600 Rupees. Together with the show-up fee of 200 Rupees, if this stage is selected for payout, **you will receive a total of 800 Rupees**.

If your P number is not the highest in your group, your earnings will be 400 Rupees. Together with the show-up fee of 200 Rupees, if this stage is selected for payout, **you will receive a total of 600 Rupees**.

If your P number is the highest, but it equals another group member's P number, we will flip a coin, and if heads come up, you will receive a total of 800 Rupees. If Tails come up, you will receive a total of 600 Rupees. If the P numbers are the same for all three members, then the member with the highest number in their alpha-numeric code or registration ID in their group will receive a total of 800 Rupees, whereas each of the other two members will receive a total of 600 Rupees. For example, if SM41, SM12, and SM23 form a group and their P numbers are the same, then the group member with SM41 will receive a total of 800 Rupees, whereas each of the other two members, SM12 and SM32, will receive a total of 600 Rupees.

After Stage One is over, we will inform you of the total number of tomatoes you packed in Stage One, the expert's count of your output, the expert's quality rating of your output, your value of N, your value of Q, your value of P, and whether you won 800 or 600 Rupees in Stage One.

Do you have any questions? Once all questions are answered, we will move to Stage Two.

Stage Two:

You will be randomly matched with two different group members at this stage.

The task in Stage Two is the same as that in Stage One. You will again be given 10 minutes to perform your task. We will count your P number in the second stage according to the same formula as in Stage One.

However, your earnings in this stage will be determined differently from how they were determined in Stage One, which we explain below.

If your P number is higher than the other two group members' P numbers, your earnings will be 800 Rupees. Together with the show-up fee of 200 Rupees, if this stage is selected for payout, **you will receive a total of 1,000 Rupees**.

If your *P* number <u>is not the highest</u> in your group, your earnings will be <u>300 Rupees</u>. Together with the show-up fee of 200 Rupees, if this stage is selected for payout, **you will receive a total of 500 Rupees**.

If your *P* number is the highest, but it equals another group member's *P* number, we will flip a coin, and if heads come up, you will receive a total of 1,000 Rupees. If Tails come up, you will receive a total of 500 Rupees. If the *P* numbers are the same for all three members, then the member with the highest number in their alpha-numeric code or registration ID in their group will receive a total of 1,000 Rupees, whereas each of the other two members will receive a total of 500 Rupees. For example, if SM41, SM12, and SM23 form a group and their *P* numbers are the same, then the group member with SM41 will receive a total of 1,000 Rupees, whereas each of the other two members, SM12 and SM32, will receive a total of 500 Rupees.

After Stage One is over, we will inform you of the total number of tomatoes you packed in Stage One, the expert's count of your output, the expert's quality rating of your output, your value of *N*, your value of *Q*, your value of *P*, and whether you won 1000 or 500 Rupees in Stage One.

Do you have any questions? Once all questions are answered, we will move to Stage Three.

Summary of Stages One and Two:

In each of the first two stages, you will be randomly assigned to a group of three and pack as many tomatoes as possible within 10 minutes.

Your two group members in Stage Two are different from your two group members in Stage One.

Your performance in each stage is determined by the expert's assessments of (a) count of the number of tomatoes you packed and (b) the quality of your work.

Your earnings (show-up fee plus the earnings from the task) depend on whether your performance is the best in your group or not, as follows:

Table 1: Your earnings

	When your performance is the best in your group	When your performance is not the best in your group
Stage One	800 Rupees	600 Rupees
Stage Two	1,000 Rupees	500 Rupees

Do you have any questions? Once the first two stages end, we will distribute instructions for Stage Three.

{Once Stage Two is over, the following instructions are explained:}

Stage Three:

Stage Two has now ended.

In Stage Three, you can choose whether to perform the same task under the Stage One or Stage Two payment scheme. You will be provided with a choice sheet. The choice sheet will list the first two stages of payment schemes. Next to each payment scheme, you will find a box. If you would like to perform the same task under the Stage One payment scheme in this stage, then tick the box next to the Stage One payment scheme. If you would like to perform the same task under the Stage Two payment scheme in this stage, then tick the box next to the Stage Two payment scheme.

If you choose to perform under the Stage One payment scheme in this stage, you will be randomly matched with two other participants who also chose to perform the task under the Stage One payment scheme in this stage, just like you. On the other hand, if you choose to perform under the Stage Two payment scheme in this stage, you will be randomly matched with two other participants who also chose to perform the task under the Stage Two payment scheme in this stage, just like you.

Once everyone completes their choice, one of the research team members will collect all the choice sheets. After that, a staff member will indicate that you can begin performing the same task within the 10-minute time interval. We will count your P number in Stage Three according to the same formula as in Stage One or Stage Two. However, your earnings in this stage will be determined **based on your choice of payment scheme**. If you choose to perform under the Stage One payment scheme in this stage, your payment will be determined just like under the Stage One scheme.

On the other hand, if you choose to perform under the Stage Two payment scheme in this stage, your payment will be determined just like under the Stage Two scheme. The summary of the payment scheme can be found in Table 1 in the instructions for Stage Two. You can also go back and re-read the payment schemes of these two stages in detail for the instructions.

After Stage Three is over, we will inform you of the total number of tomatoes you packed in Stage Three, the expert's count of your output, the expert's quality rating of your output, your value of N, your value of Q, your value of P, whether you won in Stage Three or not.

Today's work will be over once everyone completes the Stage Three task. A research team member will escort all of you to a different room where you will wait for the payments. Please allow us some time to calculate your earnings from all three stages. After calculating the payments, a staff member will come to the waiting room and show all of you a pot containing three numbers: 1, 2, or 3. The staff member will randomly pick one of the three numbers without looking at the pot. The randomly selected number, which can be 1, 2, or 3, will determine which of the three stages we will use to pay you. We will put your earnings from that stage inside an envelope. Each envelope will have a registration ID written on it. Therefore, all envelopes will have registration IDs randomly assigned to you at the beginning of the work. We will put these envelopes on a table showing the registration IDs. On your way out of the building, please pick up your envelope showing your registration ID. After you have picked up your envelope, a staff member will ask you to complete a payment form asking for your name and signature. This will conclude today's work for you.

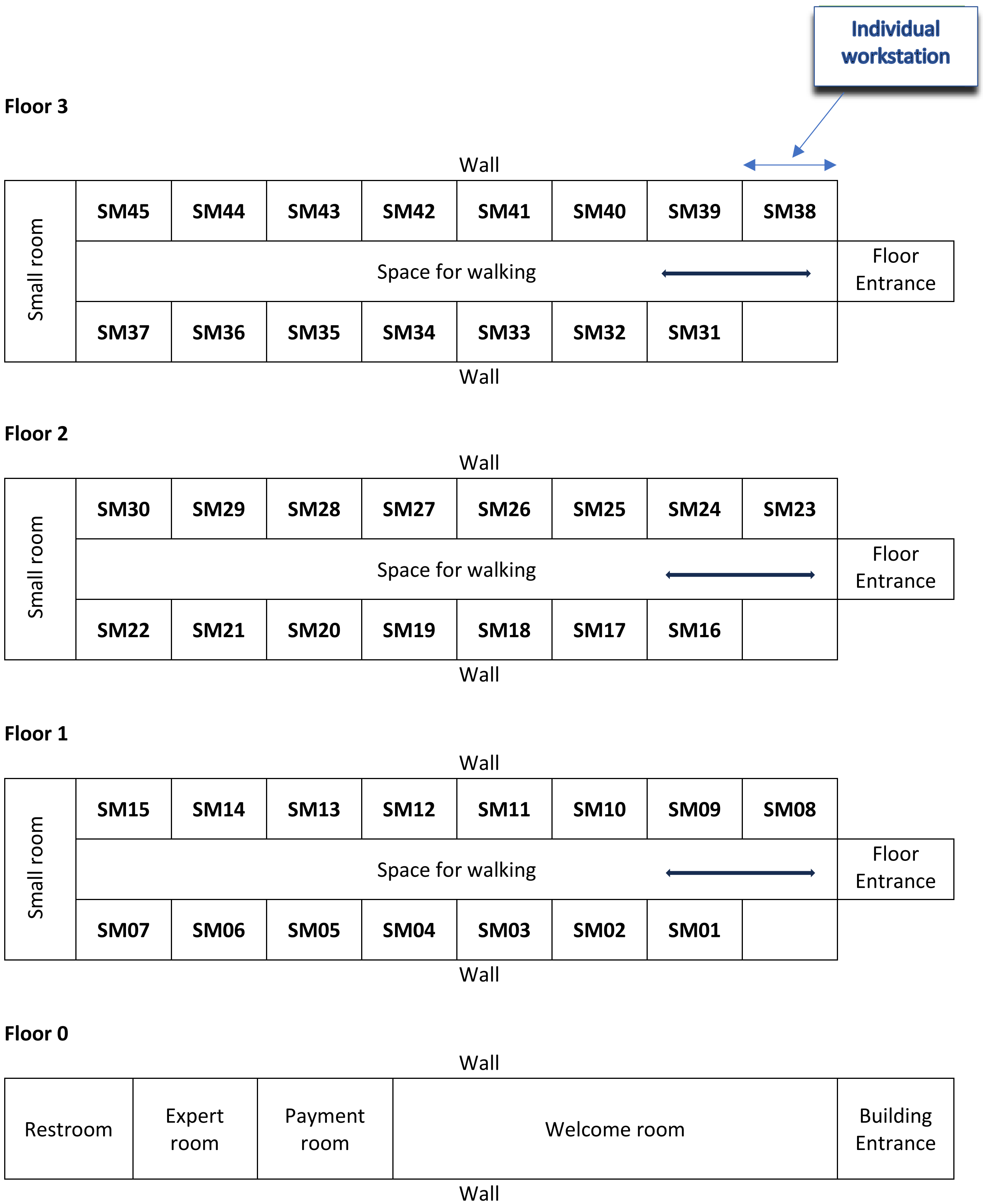
Thank you for your participation!

Do you have any questions?

Remark: An expert—a 43-year-old male with 14 years of experience in vegetable packing and recruitment for similar roles—was hired for 10 Rupees to rate and count a worker's output secretly. The same expert was used in all sessions without revealing his identity to the workers. Each floor of the building had a room in one of which the expert was stationed. The expert arrived before the workers and left the building after all the workers were sent home. Each worker's output was initially presented to the expert for a quality rating. Subsequently, the packed tomatoes from the same worker were taken to two group members for evaluation. The same output was then brought back to the expert's room for counting before being sequentially returned to the group members for their count. The workers were not made aware of this process. As discussed in this paper, the expert's ratings and counts are used to estimate the extent of sabotage.

A.4. Floor Map for the SM-P treatment

(All treatments use the same floor structure other than workers having different IDs)



A5. Photos of the Workspace (Laboratory) and Tasks in the Field

(a) Workspace (cubicle) created in the field

(b) Wooden basket and tomatoes placed on the floor for a worker

(c1) Work given the full mark

(c2) Work given a score of 50 marks

(c3) Work givens a score of 0 marks

(c) Examples of workers’ task performance by quality rating

A6. Invitation Letter for Participants

THE UNIVERSITY OF UTAH
DEPARTMENT OF ECONOMICS

To Whom It May Concern

Dear participant: November 19, 2023

আপনি একটি অর্থনীতি সংক্রান্ত গবেষণাতে অংশগ্রহণ করার জন্য আমন্ত্রিত। আপনি একটি অর্থনীতি সংক্রান্ত গবেষণাতে অংশগ্রহণ করার জন্য আমন্ত্রিত। এই গবেষণাটির উদ্দেশ্য হলো মানুষেরা কিভাবে বিভিন্ন অর্থনৈতিক বিকল্পর মধ্যে থেকে কোনো একটি বিকল্প বেছে নেন ।

এই গবেষণাটি যুগ্ম ভাবে পরিচালনা করছেন আমেরিকার University of Utah র Associate Professor of Economics, Dr. Subhasish Dugar এবং জাপানের Keio University র Professor of Economics, Dr. Kenju Kamei। এই গবেষণাটির সম্পর্কে যদি আপনার কোনো প্রশ্ন থাকে, তাহলে আপনি নীচে দেওয়া ই-মেল বা mobile ফোনের মাধ্যমে যে কোনো একজনের সাথে যোগাযোগ করতে পারেন।

এই অর্থনীতি সংক্রান্ত গবেষণাটিতে অংশগ্রহণ করার জন্য যোগ্যতা:

আপনি এই অর্থনীতি সংক্রান্ত গবেষণাতে স্বেচ্ছায় অংশগ্রহণ করতে পারবেন যদি আপনার বয়েস 18 বছর বা তার বেশি হয়।

এই অর্থনীতি সংক্রান্ত গবেষণাটিতে আপনাকে কি করতে হবে:

আপনি যদি এই অর্থনীতি সংক্রান্ত গবেষণাটিতে স্বেচ্ছায় অংশগ্রহণ করেন, তাহলে আপনাকে একটি সিদ্ধান্ত নিতে হবে। আপনাকে কি সিদ্ধান্ত নিতে হবে সেটি পরে আরও বিস্তারিতভাবে বর্ণনা করা হবে। গবেষণাটিতে অংশগ্রহণ করলে আপনাকে show-up fee হিসাবে Rs. 200 দেওয়া হবে। এ ছাড়াও আপনি আরো কিছু অতিরিক্ত টাকা আয় করতে পারবেন যেটির পরিমান নির্ভর করবে আপনার নেওয়া সিদ্ধান্ত এবং আরেকজন অজ্ঞাতপরিচয় ব্যক্তির সিদ্ধান্তের ওপর। গবেষণাটির থেকে আপনার আয় করা মোট টাকা গবেষণাটির শেষে আপনাকে দিয়ে দেওয়া হবে। গবেষণাটিতে অংশগ্রহণ করতে আপনার এক ঘন্টা বা তার কম সময় লাগবে। আপনি যদি স্বেচ্ছায় অংশগ্রহণ করতে চান, তাহলে গবেষণাটি কোন দিন, কোন সময়ে, এবং কোথায় অনুষ্ঠিত হবে সেই তথ্যগুলি আপনাকে পরে জানিয়ে দেওয়া হবে।

এই অর্থনীতি সংক্রান্ত গবেষণাটিতে অংশগ্রহণ করে আপনার কি লাভ হতে পারে:

এই গবেষণাটিতে অংশগ্রহণ করে আপনি show-up fee হিসাবে Rs. 200 পাবেন এবং আরো কিছু অতিরিক্ত টাকা আয় করতে পারেন যেটির পরিমান নির্ভর করবে আপনার নেওয়া সিদ্ধান্ত এবং আরেকটি মানুষের সিদ্ধান্তের ওপর। এই ব্যতীত আপনার আর কোনো প্রকার অর্থনৈতিক বা অনর্থনৈতিক লাভ নেই।

Department of Economics
260 Central Campus Drive
Gardner Commons, Room 4039
Salt Lake City, Utah 84112
801-581-7481
FAX 801-585-5649

এই অর্থনীতি সংক্রান্ত গবেষণাটিতে অংশগ্রহণ করে আপনাকে কি ব্যয় করতে হবে:
এই গবেষণাটিতে অংশগ্রহণ করলে আপনাকে এক ঘন্টা বা তার কম সময় ছাড়া আর কোনো ব্যয় বাহন করতে হবে না।

আপনার নেওয়া সিদ্ধান্তের গোপনীয়তা:
এই গবেষণাটিতে নেওয়া আপনার সিদ্ধান্তটি আপনি ছাড়া আর কেউই জানতে পারবেন না। আমরাও আপনার সিদ্ধান্তটি জানতে পারবো না। আপনার নেওয়া সিদ্ধান্তটি সম্পূর্ণ গোপন থাকবে। আপনার নেওয়া সিদ্ধান্তটি আমরা শুধুমাত্র গবেষণামূলক কাজের জন্য ব্যবহার করবো। এই গবেষণাটির লক্ষ্য রাজনৈতিক, ধর্মীয় বা অন্যথায় সংবেদনশীল তথ্য সংগ্রহ করা নয়। গবেষণাটি সম্পর্কে আপনার যদি আরও কোনো প্রশ্ন বা উদ্বেগ থাকে, তাহলে অনুগ্রহ করে আমাদের মধ্যে যে কোনো একজনের সাথে যোগাযোগ করুন।

Best regards,

Subhasish Dugar
Associate Professor
Department of Economics
University of Utah
260 Central Campus Drive
Gardner Commons, Room 4039
Salt Lake City, Utah 84112
Email: subhasish.dugar@economics.utah.edu
Phone: +1 801 587 1294 (U.S.)
Phone: +1 91 70444 30055 (India)
Website: https://sites.google.com/view/subhasishdugar/main

Kenju Kamei
Professor
Department of Economics
Keio University
2-15-45 Mita, Minato-ku, Tokyo, 108-8345, Japan
Email: kenjukamei@keio.jp
Phone: + 8103-5427-1493 (Japan)
Website: https://sites.google.com/site/kenjukamei2011/

Online Supplementary Material 2 for Dugar and Kamei (2025)

Appendix B: Second-Order Conditions for the Tournament Model

This appendix derives the second-order conditions (SOCs) for the expert- and peer-assessment models discussed in Section 3. While checking the SOCs, we will continue with the assumed cost functions for the two choice variables, C and Q, in the expert-assessment model and the four choice variables, C, Q, S_C, and S_Q, in the peer-assessment model. We remove the subscript i for all the choice variables as all optimal values of these variables have been derived under the symmetry conditions.

We begin with the expert-assessment model by reproducing below the two first-order conditions (FOCs) pertaining to the two choice variables, C and Q. The two FOCs are as follows:

$$(M-m)\frac{Q}{2\eta} = 3\alpha C^2 \tag{1}$$

$$(M-m)\frac{C}{2\eta} = 3\beta Q^2 \tag{2}$$

We solved the above two equations simultaneously in subsection 3.1 to obtain the optimal values of C and Q, which we reproduce below:

$$C^* = \frac{M-m}{6\eta(\alpha^2\beta)^{\frac{1}{3}}} \tag{3}$$

$$Q^* = \frac{M-m}{6\eta(\alpha\beta^2)^{\frac{1}{3}}} \tag{4}$$

Let us check for the SOC by computing the corresponding Hessian matrix at (C^*, Q^*) which we call H_1, and checking whether the matrix is negative definite. The conditions for the negative definiteness of H_1 are as follows:

The first principal minor, $\frac{\partial^2 u}{\partial C^2}$, of the matrix should be strictly negative and the second principal minor, $det(H_1)$, must be strictly positive, where $det(H_1)$ is the corresponding determinant value of the Hessian matrix. We write the H_1 below, which can be visually verified to be symmetric:

$$H_1 = \begin{pmatrix} -\frac{(M-m)\alpha^{\frac{1}{3}}}{\eta\beta^{\frac{1}{3}}} & \frac{(M-m)}{2\eta} \\ \frac{(M-m)}{2\eta} & -\frac{(M-m)\beta^{\frac{1}{3}}}{\eta\alpha^{\frac{1}{3}}} \end{pmatrix}$$

Note that the first principal minor, $-\frac{(M-m)\alpha^{\frac{1}{3}}}{\eta\beta^{\frac{1}{3}}}$, is unambiguously negative since, by assumption, $\alpha, \beta, \eta, (M-m) > 0$. The second principal minor is the determinant value of H_1. The necessary operation for finding the determinant value of the above matrix, H_1, leads to the expression $\frac{3(M-m)^2}{4\eta^2}$, which is strictly positive since η^2 and $(M-m)^2$ are both > 0. Thus, C^* and Q^* are indeed the maximum values of the corresponding optimization problem as laid out in subsection 3a.

Next, we move onto checking the SOC for the peer-assessment model as discussed in Section 3. We rewrite the four FOCs pertaining to the four choice variables, C, Q, S_C, and S_Q. We continue with the four assumed cost functions for them. The four FOCs are as follows:

$$(M-m)\frac{Q-S_Q}{2\eta} = 3\alpha C^2 \tag{5}$$

$$(M-m)\frac{C-S_C}{2\eta} = 3\beta Q^2 \tag{6}$$

$$(M-m)\frac{Q-S_Q}{4\eta} = 3\rho S_C^2 \tag{7}$$

$$(M-m)\frac{C-S_C}{4\eta} = 3\gamma S_Q^2 \tag{8}$$

We solved the above four equations simultaneously in subsection 3.2 to obtain the optimal values of C, Q, S_C, and S_Q, which we reproduce below.

$$C^* = (M-m)\left(\frac{\left(\sqrt{\frac{2\gamma}{\beta}}-1\right)^2\left(\sqrt{\frac{2\rho}{\gamma}}-1\right)}{1728\eta^3\gamma\rho^2}\right)^{\frac{1}{3}}\sqrt{\frac{2\rho}{\alpha}} \tag{9}$$

$$S_C^* = (M-m)\left(\frac{\left(\sqrt{\frac{2\gamma}{\beta}}-1\right)^2\left(\sqrt{\frac{2\rho}{\gamma}}-1\right)}{1728\eta^3\gamma\rho^2}\right)^{\frac{1}{3}} \tag{10}$$

$$Q^* = (M-m)\left(\frac{\left(\sqrt{\frac{2\rho}{\gamma}}-1\right)^2\left(\sqrt{\frac{2\gamma}{\beta}}-1\right)}{1728\eta^3\rho\gamma^2}\right)^{\frac{1}{3}}\sqrt{\frac{2\gamma}{\beta}} \tag{11}$$

$$S_Q^* = (M-m)\left(\frac{\left(\sqrt{\frac{2\rho}{\gamma}}-1\right)^2\left(\sqrt{\frac{2\gamma}{\beta}}-1\right)}{1728\eta^3\rho\gamma^2}\right)^{\frac{1}{3}} \tag{12}$$

Let us now check for the SOC by computing the Hessian matrix, which we call H_2, at (C^*, Q^*, S_C^*, S_Q^*) and checking whether the matrix is negative definite. The conditions for the negative definiteness of H_2 are as follows: all leading principal minors must alternate in sign, starting with a negative sign. We write the H_2 below, which can be visually verified to be symmetric:

$$H_2 = \begin{pmatrix} -6\alpha C & \frac{M-m}{2\eta} & 0 & 0 \\ \frac{M-m}{2\eta} & -6\beta Q & 0 & 0 \\ 0 & 0 & -6\rho S_C & -\frac{M-m}{2\eta} \\ 0 & 0 & -\frac{M-m}{2\eta} & -6\gamma S_Q \end{pmatrix}$$

The first leading principal minor, $-6\alpha C$, is negative since C^* is > 0. The second leading principal minor reduces to the expression $36\alpha\beta C^*Q^* - \frac{(M-m)^2}{4\eta^2}$, which is required to be strictly positive. This leads to the following restriction: $C^*Q^* > \frac{(M-m)^2}{144\alpha\beta\eta^2}$. The third leading principal minor, $-6\rho S_C^*$, which is required to be strictly negative. This condition holds since $S_C^* > 0$. Finally, the fourth leading principal minor, which is the determinant value of H_2. The determinant value of H_2 is required to be strictly positive. The expression for H_2 reduces to $36\rho\gamma S_C^* S_Q^* - \frac{(M-m)^2}{4\eta^2}$. This leads to the following restriction: $S_C^* S_Q^* > \frac{(M-m)^2}{144\rho\gamma\eta^2}$.

Recall that $C^* = \sqrt{\frac{2\rho}{\alpha}} S_C^*$ and $Q^* = \sqrt{\frac{2\gamma}{\beta}} S_Q^*$. Plugging their product in the second restriction above (i.e., $C^*Q^* > \frac{(M-m)^2}{144\alpha\beta\eta^2}$) yields the following inequality: $S_C^* S_Q^* >$

$\frac{(M-m)^2}{288\sqrt{\alpha\beta\rho\gamma}\eta^2}$. Combining this restriction and the fourth restriction above (i.e., $S_C^* S_Q^* > \frac{(M-m)^2}{144\rho\gamma\eta^2}$) yields, we obtain:

$$S_C^* S_Q^* > Max\left\{\frac{(M-m)^2}{144\rho\gamma\eta^2}, \frac{(M-m)^2}{288\sqrt{\alpha\beta\rho\gamma}\eta^2}\right\} \tag{13}$$

A comparison of the above two terms inside the bracket leads to the following conditions: $S_C^* S_Q^* > \frac{(M-m)^2}{144\rho\gamma\eta^2}$ if $\rho\gamma > 4\alpha\beta$; and $S_C^* S_Q^* > \frac{(M-m)^2}{288\sqrt{\alpha\beta\rho\gamma}\eta^2}$ if $\rho\gamma \leq 4\alpha\beta$.

Hence, the final conditions for H_2 to be negative definite are summarized below:

i. $C^* > 0$, which is satisfied due to the interior solution.

ii. $S_C^* > 0$, which is satisfied due to the interior solution.

iii. $S_Q^* > 0$, which is satisfied due to the interior solution.

iv. $S_C^* S_Q^* > \frac{(M-m)^2}{288\sqrt{\alpha\beta\rho\gamma}\eta^2}$ if $\rho\gamma \leq 4\alpha\beta$.

v. $S_C^* S_Q^* > \frac{(M-m)^2}{144\rho\gamma\eta^2}$ if $\rho\gamma > 4\alpha\beta$.

Online Supplementary Material 3 for Dugar and Kamei (2025)

Appendix C: Additional Figures and Tables

Table C.1: *Average Number of Attempts in Exogenous Phases*

Treatment	S	Scheme M	L	p value for H_0: $\Delta_A = 0$
Under expert assessment				
SM-E	90.76	**96.59**	---	< 0.001***
SL-E	89.11	---	**102.80**	< 0.001***
ML-E	---	93.09	**101.49**	< 0.001***
Under peer assessment				
SM-P	86.99	**96.00**	---	< 0.001***
SL-P	87.18	---	**103.40**	< 0.001***
ML-P	---	92.04	**106.57**	< 0.001***

Notes: The average numbers of tomatoes attempted to pack. Δ_A represents the average differences in attempts between two exogenous schemes under the given treatment. The two-sided p-values in each row of the table are based on linear regressions with robust standard errors. To control order effects (if any), a SM order, SL order or ML order dummy is included as a control variable in each regression. The numbers are written in bold when they are significantly different from the paired counterparts in the tests.

*, **, and *** indicate significance at the 0.10 level, at the 0.05 level, and at the 0.01 level, respectively.

Table C.2: *Sabotaging Counts or Quality Ratings in Exogenous Phases*

Dependent variable: (Expert-assessed *C* – Peer assessed *C*)/Expert-assessed *C*, or (Expert-assessed *R* – Peer assessed *R*)/Expert-assessed *R, in Phases 1 and 2*

Put differently, the dependent variable is the subject's experience of being under-evaluated by peers. The following regression examines how under-evaluation differs between "Counts" (*C*) and "Quality Rating" (*Q*). The results show that sabotage was more severe for quality rating rather than for counts.

Treatment:	SM-P		SL-P		ML-P	
	Scheme S	Scheme M	Scheme S	Scheme L	Scheme M	Scheme L
Independent variable:	(1)	(2)	(3)	(4)	(5)	(6)
"Assessing quality *Q*" dummy	0.077***	0.156***	0.052***	0.240***	0.041**	0.147***
	(0.011)	(0.011)	(0.009)	(0.021)	(0.020)	(0.025)
Constant	0.034***	0.068***	0.006	0.039***	0.077***	0.077***
	(0.007)	(0.008)	(0.005)	(0.013)	(0.014)	(0.018)
Observations	180	180	180	180	180	180
F	24.80	89.98	19.40	65.41	2.02	17.35
Loglikelihood	< 0.0001	< 0.0001	< 0.0001	< 0.0001	0.1352	< 0.0001
Pseudo R2	0.2197	0.4996	0.1558	0.4198	0.0230	0.1619

Notes: Linear regressions with robust standard errors. The reference groups are sabotage acts related to counting the numbers of tomatoes packed. To control potential order effects, a dummy variable for SM order, SL order, or ML order is included as a control in each regression.

*, **, and *** indicate significance at the 0.10 level, at the 0.05 level, and at the 0.01 level, respectively.

Table C.3: *Expert-Assessed, and Peer-Assessed Performance in Part 3*

A. Expert-Assessed Performance (*C*, *Q* and *P*)

Treatment	Counts (C)[#1]				Quality Rating (Q)				Performance P (= C × Q)			
	Scheme			*p* value for	Scheme			*p* value for	Scheme			*p* value for
	S	M	L	H_0: $\Delta_C = 0$	S	M	L	H_0: $\Delta_Q = 0$	S	M	L	H_0: $\Delta_P = 0$
(i) In the treatments with expert assessment												
SM-E	n.a.	93.99	---	n.a.	n.a.	86.21	---	n.a.	n.a.	80.50	---	n.a.
SL-E	n.a.	---	104.76	n.a.	n.a.	---	88.00	n.a.	n.a.	---	92.18	n.a.
ML-E	---	n.a.	99.17	n.a.	---	n.a.	82.00	n.a.	---	n.a.	80.82	n.a.
(ii) In the treatments with peer assessment												
SM-P	95.75	95.52	---	0.896	84.17	84.17	---	1.000	80.66	80.32	---	0.896
SL-P	88.40	---	85.00	0.913	78.15	---	68.33	0.103	68.87	---	57.70	0.144
ML-P	---	89.67	**98.44**	< 0.001***	---	75.45	71.67	0.823	---	67.88	69.89	0.091*

B. Peer-Assessed Performance (*C*, *Q* and *P*), and Sabotage activities for the SM-P, SL-P, and ML-P treatment

Treatment	Counts (C)				Quality Rating (Q)				Performance P (= C × Q)			
	Scheme			*p* value for	Scheme			*p* value for	Scheme			*p* value for
	S	M	L	H_0: $\Delta_C = 0$	S	M	L	H_0: $\Delta_Q = 0$	S	M	L	H_0: $\Delta_P = 0$
(i) Peer-assessed performance												
SM-P	93.17	93.07	---	0.952	**73.19**	66.98	---	0.002***	**68.28**	62.13	---	0.004***
SL-P	85.71	---	75.58	0.104	**70.99**	---	59.17	0.019**	**60.54**	---	44.84	0.013**
ML-P	---	82.47	80.39	0.394	---	64.32	59.44	0.907	---	53.83	48.71	0.697
(ii) Sabotage = Expert-assessed performance (Panel A.ii) minus Peer-assessed performance (Panel B.i above)												
SM-P	2.58	2.45	---	0.823	10.97	**17.19**	---	< 0.001***	12.38	**18.19**	---	< 0.001***
SL-P	2.69	---	9.42	0.074*	7.16	---	9.17	0.730	8.33	---	12.86	0.301
ML-P	---	7.20	**18.06**	0.046**	---	11.14	12.22	0.917	---	14.04	21.18	0.246

Notes: The units in the "Counts" and "Quality Rating" columns represent the average number of tomatoes packed and the average quality ratings (in %), respectively. Δ_C, Δ_Q, and Δ_P represent the average performance differences in C, Q, and P, respectively, between two exogenous schemes under the given treatment. The two-sided *p*-values in each row of the table indicate the test results for the null hypothesis that the performance differences (Δ_C, Δ_Q, or Δ_P) are zero. Each testing was performed based on a linear regression with robust standard errors. To control for potential order effects, a dummy variable for SM order, SL order, or ML order is included as a control in each regression. The numbers are shown in bold when they are significantly different from their paired counterparts in the tests. *, **, and *** indicate significance at the 0.10 level, at the 0.05 level, and at the 0.01 level, respectively.

Figure C.1: *Percentage of Groups with Peer-Evaluation-Based Rankings Different from Expert-Based Rankings in the Exogenous Phases*