Virtual teams in a gig economy

Teng Ye[a], Wei Ai[b] , Yan Chen[c,d,1] , Qiaozhu Mei[c], Jieping Ye[e], and Lingyu Zhang[f]

While the gig economy provides flexible jobs for millions of workers globally, a lack of organization identity and coworker bonds contributes to their low engagement and high attrition rates. To test the impact of virtual teams on worker productivity and retention, we conduct a field experiment with 27,790 drivers on a ride-sharing platform. We organize drivers into teams that are randomly assigned to receiving their team ranking, or individual ranking within their team, or individual performance information (control). We find that treated drivers work longer hours and generate significantly higher revenue. Furthermore, drivers in the team-ranking treatment continue to be more engaged 3 mo after the end of the experiment. A machine-learning analysis of 149 team contests in 86 cities suggests that social comparison, driver experience, and within-team similarity are the key predictors of the virtual team efficacy.

virtual teams | gig economy | social identity | team leaderboard | field experiment

According to a recent Gallup poll, 36% of US workers participate in the gig economy as either their primary or their secondary job (1). The gig economy provides workers with the benefits of autonomy and flexibility (2), but it does so at the expense of work-related identity and coworker bonds. Indeed, many gig platforms have experienced low engagement and high attrition rates among their workers, who note that they typically work alone with no interaction or relationship with other colleagues, on jobs "that don't lead to anything" (3, 4). The COVID-19 pandemic has created a work structure that has placed exponentially more workers in a work-from-home scenario that is susceptible to the same issues related to the lack of in-person interaction with coworkers as those in a gig economy. In September 2021, 45% of full-time US employees worked from home either all or part of the time (5). This trend continues into 2022. Given that we expect at least some portion of this remote work to remain postpandemic, an important question is how organizations can help their workers create and maintain positive work-related social connections while working remotely.

To answer this question, we conduct a large-scale natural field experiment using a global ride-sharing platform. Specifically, we form drivers into virtual teams and engage the teams in contests to strengthen team identity. We then evaluate the effects of these virtual teams on worker productivity, retention, and well-being.

Our research applies insights from the social identity research in psychology (6, 7) as well as studies in behavioral economics (8–10). In a laboratory setting, this research shows that, when people feel a stronger sense of common identity with a group using either induced (11–13) or natural identities (14, 15), they exert higher effort and make more contributions to improve group outcomes. Field experiments show a similar positive effect of identity-based teams in increasing prosocial behavior in fruit harvesting (16) and online peer-to-peer prosocial lending (17, 18). By contrast, other field experiments have found that when workers are paid by piece rate, providing team ranking information might reduce average worker productivity for teams that are not randomly assigned (19). To estimate the causal effects of team incentives on productivity and retention, we randomly assign teams into different experimental conditions using a large ride-sharing platform in Asia (the platform henceforth). We then examine the effect of team contests on individual driver behavior. Although virtual teams have been studied in the laboratory (20), we examine the effect of virtual teams in the field using a large-scale multicity randomized experiment.

To design our contests, we draw on insights obtained from an earlier field experiment conducted in the Chinese city of Dongguan in 2017. In this earlier experiment, we randomly assigned 2,100 drivers into seven-person teams to compete for a cash prize across a 5-d period. Team compositions are determined either randomly or based on similarity in age, hometown location, or productivity. The results from this earlier experiment show that, compared to those in the control condition, treatment drivers work longer hours and earn 12% higher revenue during the contest period (21).

Encouraged by the results of this first field experiment, in 2018, the ride-sharing platform conducted 1,548 team contests across 180 cities in China, involving over 2 million drivers placed into teams based on hometown or age similarity. These contests,

Significance

More than one-third of US workers participate in the gig economy as either their primary or their secondary job. The gig economy provides workers with the benefits of autonomy and flexibility, but at the expense of work-related identity and coworker bonds. How can organizations help their workers create and maintain positive work-related social connections while working remotely? We show that virtual team contests increase gig worker productivity and retention through a large-scale field experiment. Within-team diversity in productivity activates social comparison that increases below-median workers' effort, whereas within-team similarity in natural identities facilitates team communication and friendship formation. More broadly, this research contributes to our understanding of nonmonetary incentives.

typically 1 wk in duration, helped the platform meet the high tourist demands during national holidays and increased both driver income and retention (22). A common feature among the 1,548 team contests the platform ran in 2018 is that they were approximately 1-wk contests with cash incentives, and the teams existed only for the duration of the contest. As a result of this latter factor, the contest initiative did not provide an opportunity to study the longer-term effects of team membership on organization identity and teammate bonds.

Our study investigates the longer-term effects of team formation on the same platform in the context of contests, but without additional monetary incentives during the intervention. Specifically, in fall 2018, we designed and conducted a natural field experiment on the platform involving 27,790 drivers across three cities: Beijing, Kunming, and Taiyuan. Before the intervention, we engaged drivers in a 1-wk team contest for cash prizes, which was designed to build team identity (11). The intervention ran for the subsequent 3 wk, where we vary whether teams receive social information through the provision of a leaderboard that indicates team ranking or individual ranking within a team (treatments) or whether they receive only individual performance information (control). With the exception of the normal piece rate, there is no additional monetary incentive in any of the experimental conditions during the intervention. We then repeated the 1-wk contest for cash prizes postintervention to measure any persistent effects.

Across the 3-wk contest intervention, we find that drivers in the team and individual leaderboard treatments generated significantly higher revenue than those in the control condition. We also find interesting heterogeneous treatment effects across different cities. Three months after the experiment ended, we find that drivers in the team leaderboard treatment continue to work longer hours on the platform. Within virtual teams, those identified as "laggards" benefit the most from team contests. Our postexperiment survey, albeit with a 15% response rate, provides suggestive evidence that drivers in virtual teams made friends, shared information about order-acceptance strategies, and learned collaboration skills from their teammates.

To corroborate the underlying mechanisms identified using our experimental data, we deploy machine-learning models to uncover the most important features that predict increased revenue at the individual driver level, using data from 149 team-contest experiments in 86 cities, including our own as well as those conducted by the platform.

Our research contributes to the rapidly growing literature on the gig economy and the future of work more broadly. This literature has uncovered important insights related to labor market outcomes (23, 24), identifying factors contributing to the gender wage gap in ride sharing (25), the value of flexible work (2), consumer surplus generated by ride sharing (26), the determinants of tipping (27, 28), the effects of apologies for late trips (29), the value of passenger waiting time (30), and decentralized dynamic matching efficiency (31). Our findings contribute to this stream of research by showing that a team-based approach can significantly increase driver revenue and retention. As such, our field experiment uses insights from behavioral market design to help structure the future of work (32, 33). Our research also contributes to the management and organization literature on the effects of nonmonetary incentives (34). In particular, researchers find that a participatory organization structure that enables workers to voice opinions in their group increases worker productivity and job satisfaction compared to a hierarchical structure (35) and that a purely symbolic award has a sizable and persistent impact on the retention of new editors in Wikipedia (36). Finally, our results

corroborate several findings from the social network literature that investigates peer effects in exercise (37), peer observation on savings (38), and friendship formation (39).

Experiment Design

As mentioned, we design and conduct a natural field experiment on the platform involving 27,790 drivers across Beijing, Kunming, and Taiyuan, three cities chosen to exemplify diversity in location, size, and the number of team contests hosted on the platform prior to our experiment (*SI Appendix*, Table S1). Our experiment is approved by the University of Michigan Institutional Review Board (IRB) (HUM00153090) and preregistered at the American Economic Association's Registry for Randomized Controlled Trials (AEA RCT) Registry (AEARCTR-0003537) (40). The IRB approved our request to waive the informed consent, because awareness of the research study can affect behavior (41). The experiment was conducted from 22 October 2018 to 3 December 2018. To evaluate our treatment effect on driver retention, we continue to collect data for 3 mo after our experiment, until 10 March 2019. In addition to our recruitment and team formation stage, our experiment is organized into preintervention, intervention, and postintervention stages. *SI Appendix*, Fig. S1 presents the experimental process.

Driver Recruitment and Team Formation. For the driver recruitment and team formation stage, the platform used its built-in process that was informed by our earlier experiment (21). The platform sent an invitation on 22 October 2018 to all active drivers in our three cities to participate in a week-long team contest for a cash prize.* Interested drivers are invited to sign up for the contest and start forming teams. Drivers can create a new team as a captain, invite others to join their team, or join an existing team if invited to do so.

While teams are designed to have seven members, 36% of our teams achieved the desired size during the team formation period. Those that reached the desired size during the team formation period are referred to as self-formed teams. At the end of the recruitment stage, the system then randomly selects 90% of the drivers in undersized teams and groups them into full-sized teams, which we refer to as system-formed teams. The system-formed teams are based on either hometown or age similarity, two of the most successful team formation algorithms from our earlier experiment (21). The remaining 10% are not assigned to any team and do not participate in the contest. These drivers are referred to as solo drivers. In our analysis, we control for whether a team is self-formed.

Finally, we sort teams into contest groups. To assign the teams into contest groups, we first sort teams within each city decreasingly based on their prior revenue (the sum of individual team members' revenue in the 2 wk prior to the beginning of the experiment). We then partition every five adjacent teams into a contest group, also referred to as a leaderboard. Teams compete only with other teams in the same leaderboard. Our grouping method ensures that teams in the same leaderboard have similar prior productivity, to approximate the assumption of ex ante symmetry found in most of the theoretical contest

---

*While the preintervention contest was designed to build group identity, the recruitment of drivers based on the attraction for cash incentives might have attracted more competitive drivers. This self-selection of drivers might affect the effect of the intervention. While we do not have the data on those who did not sign up for our experiment, our companion paper reporting the Dongguan field experiment shows that those who signed up for the team contest with a cash prize were more productive prior to the experiment than a random sample of drivers who were not contacted by the experimenter (21).

literature (42, 43). In practice, grouping similar teams or players in a contest is often observed in sports. A machine-learning analysis of incentivized short-term team contests conducted by the platform also indicates that drivers exert higher effort when their team's precontest revenue is closer to that of the best team in the leaderboard (22). We now describe the three stages of the team contest.

The Preintervention Contest. Following Eckel and Grossman (11) who find that intergroup competition is among the most successful methods used for creating a strong sense of group identity, we conduct a preintervention best-of-five team contest. In this contest, within each leaderboard, the team with the highest cumulative team revenue during the contest week wins a cash prize, whereas the other four teams receive no prize. Following the platform's current contest practice, we exclude the lowest driver revenue in a given team on each day when calculating the team's daily cumulative team revenue. This allows one driver on a team to take a day off without affecting team performance.[†] The cash prize is 1,000 Chinese yuan renminbi (CNY) per winning team for Beijing and 650 CNY for Taiyuan and Kunming, respectively, adjusted by the drivers' average hourly revenue in each city. The prize is allocated to members of the winning team proportional to their contributions to the cumulative team revenue, an allocation shown to incentivize group members in laboratory contests (44), and is credited to their driver accounts immediately after the contest.

During this stage, all drivers participating in the contest can use the platform app to access both a team leaderboard and an individual leaderboard for social information, as illustrated in *SI Appendix*, Fig. S2. The team leaderboard shows the cumulative revenue of each of the five teams in the contest group in descending order (*SI Appendix*, Fig. S2, *Top Left*). The top three teams are highlighted with badges. The individual leaderboard shows individual members' daily revenue in descending order for those within a given team (*SI Appendix*, Fig. S2, *Top Right*). In addition, we mark the average performance of that team with a line on the individual leaderboard to enhance the effect of ranking (45, 46). The team ranking is updated every hour while individual revenue is updated in real time. We send each driver a daily reminder of the contest and the leaderboards at the end of each day. The communication messages for each stage of the experiment can be found in *SI Appendix*, section 1D.

The Intervention: A Status Contest. Immediately after the preintervention contest, we randomly assign each leaderboard to one of three experimental conditions and conduct a 3-wk status contest between 5 and 25 November to examine the effect of team identity on driver revenue and retention.

- Team Leaderboard. In this treatment, drivers continue to have access to both the team and individual leaderboards as in the preintervention contest. We send out a daily reminder to these drivers to check the rankings of the same five teams within their leaderboard. When a driver taps the driver's team name on the team leaderboard (the default interface), the driver can further access the individual leaderboard within the team.
- Individual Leaderboard. In this treatment, drivers have access to only the individual leaderboard within their team. Again, we send out a daily reminder to drivers to check their individual rankings.

- Control. In the control condition, drivers cannot access either leaderboard. However, to keep the same communication frequency, drivers continue to receive a daily reminder that they can access their own revenue statistics in the app (*SI Appendix*, Fig. S2, *Bottom*).

While drivers continue to earn piece rate, we do not provide additional monetary incentives for the status contest.

The randomization is stratified based on the average revenue of a given leaderboard in the 2 wk prior to the start of the experiment. Kolmogorov–Smirnov tests show that the distribution of preexperiment revenue, age, gender, work experience (platform age), team formation, and hometown distance to the contest city is not significantly different in pairwise comparisons across the three conditions within each city ($P > 0.10$; *SI Appendix*, Table S4). *SI Appendix*, Table S4 also reveals interesting facts about our drivers: More than 95% of them are male, with an average age of 37 y. Looking at their hometown distance to the contest city, we conclude that Taiyuan drivers are predominantly local, whereas Beijing and Kunming drivers are mostly domestic migrants. In China, the platform drivers are composed of workers laid off from their traditional jobs, veterans, migrant workers from rural areas, and commuters who offer rides during their daily commute.

The Postintervention Contest. On 26 November, we send each driver a message announcing a 1-wk contest for a cash prize from 27 November to 3 December under the same leaderboard groups and prize parameters as in the preintervention contest. This postintervention contest is designed to evaluate whether treatment effects on individual driver productivity persist immediately after the intervention.

The Postexperiment Survey. After the postintervention contest, all drivers receive a survey that evaluates whether they like the status contest, what they get out of the contest, and their sense of belonging related to their team as well as to the organization. The survey questions and responses are included in *SI Appendix*, section 4.

Results

Our experiment yields findings related to the immediate and longer-term effects of virtual teams on driver revenue and retention, both overall (40) and at the city level. On the platform, drivers receive 81% of the revenue they generate and give the remaining 19% to the platform. Therefore, using revenue as one of our outcome variables is equivalent to using driver earning or platform profit. In what follows, we first report our preregistered hypotheses and the corresponding results and then explore the underlying mechanisms. Finally, we use a machine-learning model to validate our findings in 149 team contests in 86 cities.

Preregistered Hypotheses and Results. We first examine the average treatment effect on driver revenue during the experiment period. In Fig. 1, we plot the weekly average driver revenue for each experimental group. To better compare the treatments, we realign the lines based on revenue earned during the preexperiment period. The y axis presents the revenue difference between a given week and the baseline week(s) in the preexperiment period. Note that the three lines coincide up to the start of the preintervention contest period. However, during the status contest intervention, since drivers in different treatment conditions receive different social information, the lines in Fig. 1A start to diverge. Pooling all three cities, we observe that our treatment drivers are more

---

[†]In some cities, such as Beijing, to reduce air pollution, each license plate must be off the street on a designated day of the week, typically determined by the last digit of the license plate number.
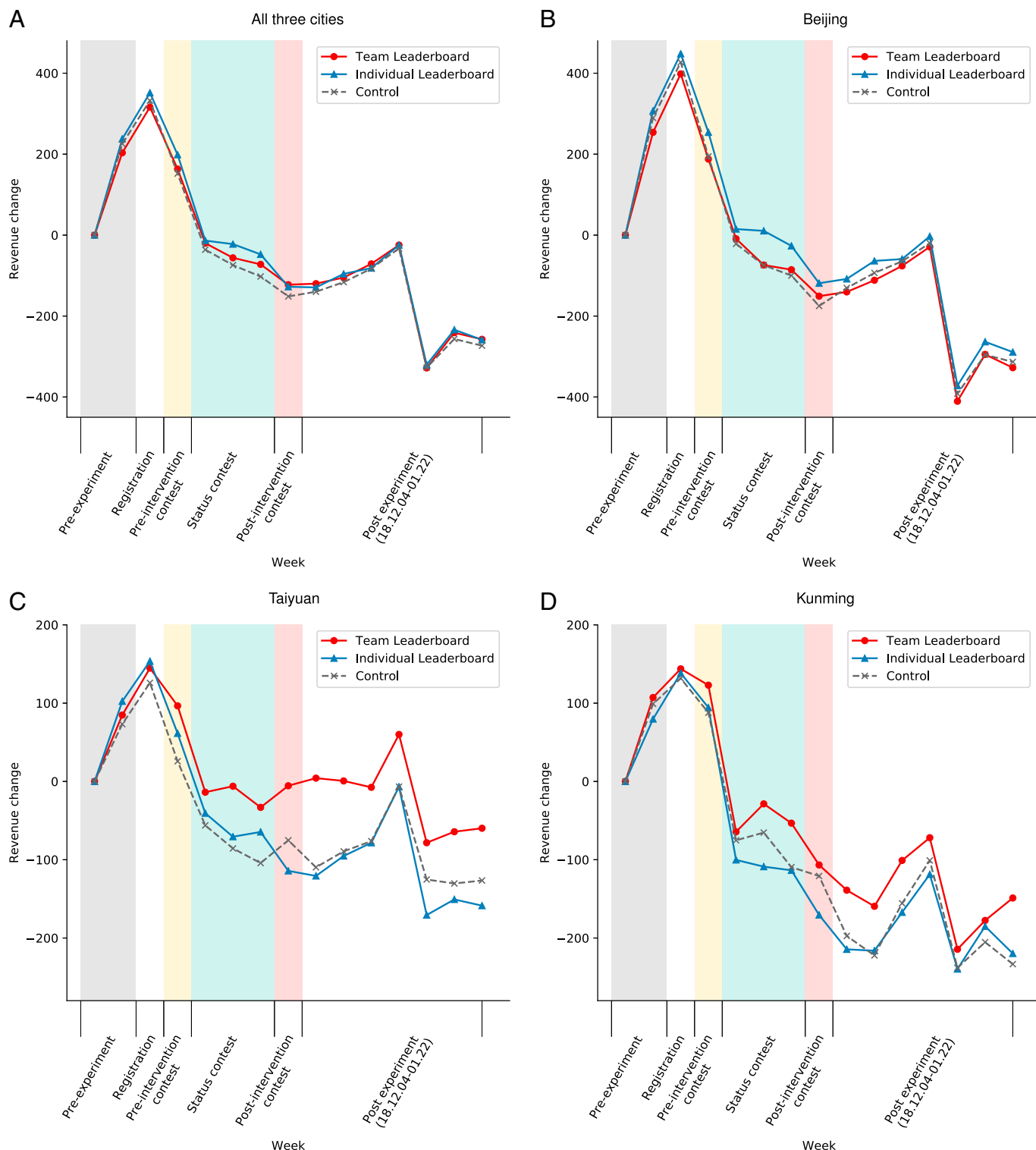
Fig. 1. (*A–D*) Average weekly driver revenue under each experimental condition. To better visualize the changes over time, we rescale the revenue within each experimental condition with reference to its preexperiment average weekly revenue from the week of 8 to 14 October, i.e., 2 wk before the start of the experiment. For example, each point represents the weekly average revenue per driver under that experimental condition minus the preexperiment weekly average revenue per driver under the same experimental condition.

productive on average than those in the control condition both during and after the intervention.

We report the main results in Tables 1–4 in the main text and the results of our robustness checks in *SI Appendix*. To correct for multiple-hypothesis testing, we report the false discovery rate adjusted *q* values in square brackets (47, 48). To claim significance, we use a 5% (10%) cutoff for our *P* values (*q* values) (49).

In our first hypothesis, based on prior laboratory experiments on social identity and team competition (11) as well as those on individual performance ranking (50), we predict that drivers in our treatment conditions will generate higher revenue than those in the control condition as their exposure to a leaderboard should facilitate a team identity. The comparison between team leaderboard and individual leaderboard is motivated by laboratory experiments in group contests (51, 52).

Table 1. Average and heterogeneous treatment effects on weekly revenue during the intervention (status contest): Difference-in-differences regressions

	Dependent variable: Δ of weekly revenue (CNY)							
	1) All	2) Beijing	3) Taiyuan	4) Kunming	5) All	6) Beijing	7) Taiyuan	8) Kunming
Treated, in a virtual team	34.53** (15.37) —	41.67** (21.01) [0.17]	33.99 (23.86) [0.18]	8.25 (24.97) [0.33]	39.62*** (14.47) —	45.50** (19.67) [0.07]	43.64* (22.67) [0.07]	13.36 (23.48) [0.24]
Age, y					4.56*** (0.81)	4.84*** (1.13)	0.65 (1.24)	6.40*** (1.24)
Platform age, y					35.36*** (7.19)	48.86*** (9.21)	0.43 (11.33)	−1.65 (12.79)
Hometown distance to contest city, km					−0.07*** (0.02)	−0.08*** (0.02)	−0.11** (0.05)	−0.04 (0.02)
Self-formed team					−110.96*** (14.99)	−128.66*** (19.91)	−74.70*** (25.69)	−53.57** (25.04)
Individual ranking within team during the preintervention contest					−77.32*** (3.56)	−95.45*** (4.92)	−30.43*** (5.35)	−49.74*** (5.51)
Team ranking during the preintervention contest					−150.46*** (5.08)	−176.56*** (6.89)	−86.77*** (7.74)	−101.72*** (8.06)
City fixed effect	Yes	—	—	—	Yes	—	—	—
No. of clusters	11,890	8,100	1,625	2,165	11,890	8,100	1,625	2,165
No. of drivers	27,790	18,900	3,815	5,075	27,790	18,900	3,815	5,075

SEs in parentheses are clustered at the team (individual) level for treated (control) drivers. False discovery rate adjusted q values calculated separately for individual cities (columns 2–4 and columns 6–8) are reported in square brackets. Individual ranking is coded from 1 (top) to 7 (bottom), whereas team ranking is coded from 1 (top) to 5 (bottom). *$P < 0.1$, **$P < 0.05$, ***$P < 0.01$.

Hypothesis 1 (Status Contest). *a) Treated drivers are more productive than those in the control condition, and b) drivers in the team leaderboard condition are more productive than those in the individual leaderboard condition during the status contest phase.*

To quantify the average treatment effects on outcome, Y, we construct the following difference-in-differences regression model for each target period:

$$\Delta Y_i = \beta_0 + \beta_1 \text{Treated}_i + \alpha_c + \epsilon_i, \qquad [1]$$

Table 2. Average and heterogeneous treatment effects on weekly revenue during the intervention (status contest): Difference-in-differences regressions investigating the two treatments separately

	Dependent variable: Δ of weekly revenue (CNY)							
	1) All	2) Beijing	3) Taiyuan	4) Kunming	5) All	6) Beijing	7) Taiyuan	8) Kunming
Team leaderboard, β_1	32.12* (17.97) [0.08]	27.03 (24.61) [0.44]	58.49** (26.60) [0.09]	30.54 (29.91) [0.44]	37.37** (16.72) [0.02]	32.17 (22.77) [0.27]	67.62*** (25.02) [0.03]	33.05 (27.59) [0.30]
Individual leaderboard, β_2	36.96** (17.90) [0.08]	56.32** (24.49) [0.09]	8.81 (28.76) [0.86]	−14.50 (28.03) [0.86]	41.89** (16.55) [0.02]	58.83*** (22.45) [0.03]	18.85 (26.68) [0.46]	−6.77 (26.06) [0.53]
Age, y					4.56*** (0.81)	4.84*** (1.13)	0.67 (1.32)	6.35*** (1.24)
Platform age, y					35.35*** (7.18)	48.77*** (9.20)	0.37 (11.36)	−1.79 (12.78)
Hometown distance to contest city, km					−0.07*** (0.02)	−0.08*** (0.02)	−0.11** (0.05)	−0.04 (0.02)
Self-formed team					−110.92*** (14.99)	−128.32*** (19.91)	−74.12*** (25.58)	−53.43** (25.05)
Individual ranking within team during the preintervention contest					−77.32*** (3.56)	−95.45*** (4.92)	−30.48*** (5.35)	−49.76*** (5.52)
Team ranking during the preintervention contest					−150.46*** (5.08)	−176.55*** (6.89)	−86.73*** (7.72)	−101.73*** (8.04)
City fixed effect	Yes	—	—	—	Yes	—	—	—
H_0: $\beta_1 = \beta_2$ (P value)	0.79	0.25	0.08	0.13	0.78	0.23	0.05	0.13
No. of clusters	11,890	8,100	1,625	2,165	11,890	8,100	1,625	2,165
No. of drivers	27,790	18,900	3,815	5,075	27,790	18,900	3,815	5,075

SEs in parentheses are clustered at the team (individual) level for treatment (control) conditions. False discovery rate adjusted q values are calculated separately for all cities (columns 1 and 5) and for individual cities (columns 2–4 and 6–8) and are reported in square brackets. Individual ranking is coded from 1 (top) to 7 (bottom), whereas team ranking is coded from 1 (top) to 5 (bottom). *$P < 0.1$, **$P < 0.05$, ***$P < 0.01$.

Table 3. Average and heterogeneous treatment effects on weekly revenue in the postintervention contest: Difference-in-differences regressions

	Dependent variable: Δ of weekly revenue (CNY)							
	1) All	2) Beijing	3) Taiyuan	4) Kunming	5) All	6) Beijing	7) Taiyuan	8) Kunming
Team leaderboard, β_1	49.91**	59.89*	58.03	6.05	56.52**	67.42**	64.55*	10.64
	(23.80)	(32.49)	(37.50)	(39.57)	(22.41)	(30.47)	(35.01)	(36.89)
	[0.08]	[0.32]	[0.32]	[0.56]	[0.02]	[0.19]	[0.19]	[0.35]
Individual leaderboard, β_2	11.75	38.98	−68.26*	−30.36	18.12	42.63	−60.63*	−19.23
	(24.30)	(33.12)	(39.25)	(39.52)	(22.81)	(31.00)	(36.39)	(36.84)
	[0.46]	[0.32]	[0.32]	[0.36]	[0.27]	[0.24]	[0.19]	[0.34]
Age, y					9.34***	10.10***	3.85**	10.38***
					(1.06)	(1.48)	(1.63)	(1.68)
Platform age, y					83.80***	99.65***	35.47**	33.63**
					(9.46)	(12.12)	(15.32)	(16.72)
Hometown distance to contest city, km					−0.06**	−0.08***	−0.15**	0.01
					(0.02)	(0.03)	(0.06)	(0.03)
Self-formed team					−76.15***	−96.46***	−24.77	−15.80
					(20.74)	(27.52)	(37.26)	(35.56)
Individual ranking within team during the preintervention contest					−23.73***	−28.86***	−11.75	−17.28**
					(4.45)	(6.11)	(7.41)	(7.28)
Team ranking during the preintervention contest					−127.65***	−146.00***	−84.77***	−92.18***
					(6.82)	(9.25)	(11.32)	(10.73)
City fixed effect	Yes	—	—	—	Yes	—	—	—
H_0: $\beta_1 = \beta_2$ (P value)	0.11	0.53	0.00	0.33	0.09	0.42	0.00	0.40
No. of clusters	3,970	2,700	545	725	3,970	2,700	545	725
No. of drivers	27,790	18,900	3,815	5,075	27,790	18,900	3,815	5,075

SEs in parentheses are clustered at the team level. False discovery rate adjusted q values are calculated separately for all cities (columns 1 and 5) and for individual cities (columns 2–4 and 6–8) and are reported in square brackets. Individual ranking is coded from 1 (top) to 7 (bottom), whereas team ranking is coded from 1 (top) to 5 (bottom). *$P < 0.1$, **$P < 0.05$, ***$P < 0.01$.

where ΔY_i represents the outcome change in the current period compared to the corresponding precontest week(s), and α_c captures city fixed effects.[‡] *Hypothesis 1a* implies that $\beta_1 > 0$ in Eq. **1**.

The results in column 1 of Table 1 show that our treatment conditions increase driver revenue by 34.53 CNY, or 1.66% of the

[‡]For the preexperiment baseline week(s), we use the week before the experiment (15 to 21 October 2018) for 1-wk target periods, i.e., the preintervention contest, the

postintervention contest, and retention. For the status contest, we use the 2 wk before the experiment (8 to 21 October 2018) as our baseline, as the week of 1 to 7 October 2018 was a national holiday with drastically different demand and supply for ride sharing.

Table 4. Average and heterogeneous treatment effects on weekly number of working days during the second week of March (4 to 10 March 2019), about 3 mo after the experiment ended: Difference-in-differences regressions

	Dependent variable: Δ of weekly no. of work days							
	1) All	2) Beijing	3) Taiyuan	4) Kunming	5) All	6) Beijing	7) Taiyuan	8) Kunming
Team leaderboard, β_1	0.10**	0.06	0.33***	0.05	0.11**	0.08	0.33***	0.06
	(0.05)	(0.06)	(0.12)	(0.10)	(0.04)	(0.05)	(0.11)	(0.10)
	[0.06]	[1.00]	[0.03]	[1.00]	[0.02]	[0.50]	[0.02]	[1.00]
Individual leaderboard, β_2	−0.01	−0.01	−0.02	0.01	0.01	−0.00	−0.01	0.04
	(0.05)	(0.06)	(0.12)	(0.10)	(0.04)	(0.05)	(0.12)	(0.10)
	[0.70]	[1.00]	[1.00]	[1.00]	[0.77]	[1.00]	[1.00]	[1.00]
Age, y					0.03***	0.03***	0.02***	0.03***
					(0.00)	(0.00)	(0.01)	(0.00)
Platform age, y					0.22***	0.24***	0.08	0.18***
					(0.02)	(0.02)	(0.05)	(0.05)
Hometown distance to contest city, km					−0.00***	−0.00***	−0.00**	−0.00
					(0.00)	(0.00)	(0.00)	(0.00)
Self-formed team					−0.07*	−0.16***	0.10	0.16*
					(0.04)	(0.05)	(0.10)	(0.09)
Team won in postintervention contest					0.66***	0.68***	0.63***	0.61***
					(0.05)	(0.06)	(0.12)	(0.11)
City fixed effect	Yes	—	—	—	Yes	—	—	—
H_0: $\beta_1 = \beta_2$ (P value)	0.02	0.17	0.00	0.66	0.02	0.12	0.00	0.85
No. of drivers	27,790	18,900	3,815	5,075	27,790	18,900	3,815	5,075

False discovery rate adjusted q values are calculated separately for all cities (columns 1 and 5) and for individual cities (columns 2–4 and 6–8) and are reported in square brackets. The results hold if we alternatively control for the number of wins in the two short contests instead of the team that wins the postintervention contest. *$P < 0.1$, **$P < 0.05$, ***$P < 0.01$.

average weekly revenue per driver, during the 3-wk intervention ($P < 0.05$). Therefore, we reject the null hypothesis in favor of *Hypothesis 1a*. We further find a significant treatment effect for drivers in Beijing (41.67 CNY, $P < 0.05$, 1.69% of average weekly revenue), but not in Taiyuan or Kunming. Our findings are strengthened (39.62 CNY for all cities, or 1.90%, $P < 0.01$) when we control for demographics and self-formed versus system-formed teams, as well as individual and team ranking in the preintervention contest (Table 1, columns 5–8). Consistent with social identity theories focusing on the effects of social status and social distance on individual identification with social groups (53, 54), a driver's hometown distance from the contest city is negatively correlated with the driver's productivity ($P < 0.01$; Table 1, columns 5 and 6; $P < 0.05$, column 7). Interestingly, self-formed teams generate lower revenue compared to system-formed teams using hometown or age similarity ($P < 0.01$; Table 1, columns 5–7; $P < 0.05$, column 8). We conjecture that drivers in self-formed teams knew each other prior to our experiment and therefore did not feel the need to impress or to signal that they were responsible. By contrast, system-formed teams group strangers together, who might have felt a stronger need to signal to their teammates (38). We also note that older drivers and those who have joined the platform earlier generate higher revenue. Finally, a poor preintervention individual or team ranking is associated with a reduction in revenue increase. For example, a one-place drop in team ranking is associated with a 150.46 CNY reduction in revenue increase ($P < 0.01$; Table 1, column 5).

SI Appendix, Table S7 repeats the same analysis with working hours as the outcome variable. Results indicate that the treatment effects are driven by longer working hours.[§] We further investigate whether there is any adverse treatment effect on safety. *SI Appendix*, section 5 reports our analysis of driver safety scores, which shows that our intervention has no adverse effect on safety (*SI Appendix*, Table S6).

Investigating the two types of interventions separately (*Hypothesis 1b*), we further expect that drivers in the team leaderboard treatment will generate higher revenue than those in the individual leaderboard treatment, who in turn will generate higher revenue than those in the control group during our intervention period. This hypothesis implies that $\beta_1 > 0$, $\beta_2 > 0$, and $\beta_1 > \beta_2$ in Eq. **2** below:

$$\Delta Y_i = \beta_0 + \beta_1 \text{TeamLeaderboard}_i + \beta_2 \text{IndividualLeaderboard}_i + \alpha_c + \epsilon_i. \quad [2]$$

The results in column 5 of Table 2 show that the team (individual) leaderboard generates 37.37 (41.89) CNY higher weekly revenue compared with the control group, equivalent to a 1.79% (2.01%) increase ($P < 0.05$ in each case), after controlling for covariates. Furthermore, the difference between the two treatments is not significant ($P > 0.10$). Again, *SI Appendix*, Table S10 reports the same analysis for working time.

We next examine our city-level results. From Table 2, we see that, in Beijing (columns 2 and 6), only the individual leaderboard treatment has a significant effect on revenue (58.83 CNY, or 2.39% of the weekly revenue change of the control group, $P < 0.01$), whereas in Taiyuan (columns 3 and 7), only the team leaderboard treatment has a significant effect on revenue compared to the control condition (67.62 CNY, or 6.07%, per week, $P < 0.01$). By contrast, neither treatment has a significant

effect on weekly revenue for drivers in Kunming. As shown in *SI Appendix*, Table S1, passenger order fulfillment rate was already 98% in Kunming before our experiment; thus, there was little room for a substantial improvement in revenue. In comparison, 90% of the orders were fulfilled in Beijing and Taiyuan during the same time period. Furthermore, we reject the null in favor of *Hypothesis 1b* that drivers under the team leaderboard treatment generate higher revenue than those under the individual leaderboard in Taiyuan ($P = 0.05$; Table 2, column 7).

Finally, unbeknownst to us during our experiment, the platform's Operations Department implemented an individual threshold cash bonus in Beijing for each weekday during the 3 wk of our intervention, whereas Taiyuan and Kunming did not receive this interference. It is therefore plausible that the heterogeneous effects in Beijing and Taiyuan might be partially due to the different incentives in each city, with Taiyuan implementing a pure status contest during our intervention.[¶] This leads to our first main result.

Result 1 (virtual teams and productivity): During the 3-wk status contest intervention, 1) treated drivers work 42 min longer per week and generate 1.9% higher revenue than those in the control condition; 2) drivers in the team (individual) leaderboard treatment work 45 (39) min longer and generate 1.79% (2.01%) higher revenue than those in the control condition; and 3) at the city level, drivers in the team (individual) leaderboard treatment work 125 (48) min longer, leading to a 6.07% (2.39%) increase in revenue in Taiyuan (Beijing) compared to the control group, whereas neither treatment has a significant effect in Kunming.

Note that our status contest belongs to the class of information provision experiments. The effect sizes reported in *Result 1* are largely consistent with the meta-analysis results using 126 randomized control trials covering 23 million individuals (55). *Result 1* indicates that virtual team contests increase driver working hours, which leads to increased revenue among treated drivers. However, we do not find any significant treatment effect on revenue per hour, indicating that drivers might have been working during the busiest time blocks prior to our intervention.[#]

We are also interested in the question of whether our team effect persists over time. To evaluate the short-term effect, we implement a 1-wk best-of-five contest with a monetary reward immediately after the intervention. The postintervention contest rules are identical to those of the preintervention contest. We did not announce this contest until the 3-wk status contest was over. We expect that the treatment effects will persist during this postintervention contest.

Hypothesis 2 (Treatment Persistence). *Drivers in the team leaderboard condition are more productive than those in the individual leaderboard condition, who in turn are more productive than those in the control condition during the postintervention contest.*

The results in column 5 of Table 3 show that drivers in the team leaderboard treatment generate 56.52 CNY higher revenue (or 2.82%, $P < 0.05$) during our postintervention contest, compared to those in the control group. By contrast, drivers in the individual leaderboard treatment do not differ significantly

---

[§]Note that to protect drivers from fatigue driving, the ride-sharing app imposes an upper limit of 10 working hours per day and logs a driver out once the driver reaches the upper limit.

[¶]A threshold cash bonus consists of a preannounced cash bonus for any individual driver whose number of trips on a given day exceeds a preannounced threshold. Unfortunately, we do not have data on the types of threshold cash bonuses in Beijing during our intervention.

[#]*SI Appendix*, Table S12 presents treatment effects on revenue per hour during the status contest and shows that there is no significant treatment effect on revenue per hour overall ($P > 0.10$, columns 1 and 5) and a negative effect in Taiyuan (-0.72, $P < 0.05$ for individual leaderboard, and -0.71, $P < 0.01$ for team leaderboard, column 7).

from those in the control group ($P > 0.10$). The coefficient for the team leaderboard dummy is marginally greater than that for the individual leaderboard ($P = 0.09$; Table 3, column 5). *SI Appendix*, Table S11 reporting the same analysis for working time shows that drivers in the team leaderboard treatment work 64 min longer per week than the control group during our postintervention contest ($P < 0.01$; Table 3, column 5). Again, we do not find any significant treatment effect on revenue per hour.‖

At the city level, Beijing drivers in the team leaderboard treatment generate significantly higher revenue during our postintervention contest than those in the control group (67.42 CNY, $P < 0.05$; Table 3, column 6). By contrast, we find no persistent effect of the individual leaderboard treatment for Beijing drivers.

For drivers in Taiyuan, those in the team leaderboard treatment generate significantly higher revenue than those in the individual leaderboard treatment ($\beta_1 \neq \beta_2$, $P < 0.01$ in Table 3, columns 3 and 7) and marginally higher revenue than those in the control condition (64.55 CNY, $P < 0.10$; Table 3, column 7). It is worth noting that those in the individual leaderboard treatment exhibit a marginally significant reduction in average weekly revenue during the postintervention contest compared to the control group (-60.63 CNY, $P < 0.10$; Table 3, column 7). Again, we observe no treatment effect for Kunming drivers (Table 3, columns 4 and 8). Based on a theoretical model of individual status contests (56), depending on the properties of the ability distribution function, the aggregate revenue under an individual leaderboard can be lower than that under the control condition, as we observe in Taiyuan. We state our results related to the persistence of our treatment effect below.

Result 2 (treatment persistence): During the 1-wk postintervention contest, drivers in the team leaderboard treatment work 64 min longer and generate 2.82% higher weekly revenue compared to those in the control group, whereas the individual leaderboard treatment no longer has an effect. The team treatment effects are driven by drivers in Beijing and Taiyuan.

In addition to testing whether teams incentivize individual drivers to generate more revenue, we are interested in whether these individuals are more likely to continue working as drivers. Driver retention is a key challenge for ride-sharing platforms across the globe. As such, an important goal for our intervention is to evaluate the effects of virtual teams on driver retention. Specifically, we hypothesize that drivers who are part of a virtual team are more likely to continue as drivers than those in the control group.

Hypothesis 3 (Retention). *Drivers in the team and individual leaderboard conditions are more likely to stay in the platform than those in the control condition both during and after our experiment.*

To examine the effect of team membership on driver retention, we measure driver retention 1 wk, 1 mo, and 3 mo after the end of our experiment. Unlike workers in traditional sectors whose departure is unambiguous, gig workers who quit typically do not delete their app, so it is possible that those who have quit driving may still log into the app. Therefore, we use three different retention measures: 1) whether they drive for the platform on a given day, 2) the volume of working hours, and 3) whether they quit before the end of an observation period.

We present measures 1 and 2 in the main text and relegate measure 3 to *SI Appendix*, section 7. For the first retention measure, we count the number of days that a driver provides at least one ride and separately analyze retention during the status contest (*SI Appendix*, Table S15), the week immediately after (*SI Appendix*, Table S16), 1 mo after (*SI Appendix*, Table S17), and 3 mo after (Table 4) the postintervention contest.

As shown in *SI Appendix*, Fig. S4, drivers in the team leaderboard treatment consistently exhibit higher retention than those from either of the other experimental conditions. From Table 4 and *SI Appendix*, Table S20, we see that drivers in the team leaderboard treatment on average work 0.11 d, or 1.01 h, more than those in the control group in the week 3 mo after the experiment ended ($P < 0.05$; Table 4 and *SI Appendix*, Table S20, column 5). Furthermore, we find that drivers in the team leaderboard treatment also outperform those in the individual leaderboard treatment ($P = 0.02$; Table 4, column 5). The effect size is stable across different time windows, using the number of either working days or hours (*SI Appendix*, Tables S16–S19). Finally, we observe no significant difference in retention across any of the periods between those in the individual leaderboard treatment and those in the control group except in Kunming (2.02, $P < 0.05$; column 8 in *SI Appendix*, Table S20).

Examining our city-level results, Table 4, columns 2–4 shows significant differences in driver retention across cities. Indeed, only in Taiyuan do we see a consistent positive effect of the team leaderboard treatment on retention (0.33 d, $P < 0.01$ in Table 4, or 3.05 h, $P < 0.01$ in *SI Appendix*, Table S20), with a similar significant effect between the team and individual leaderboard treatments ($P < 0.01$). In Kunming and Beijing, we find a positive albeit insignificant effect of the team leaderboard treatment on retention. As a robustness check, we report treatment effects on the volume of working hours (*SI Appendix*, Tables S18–S20) and on whether a driver quits in our observation window (*SI Appendix*, Tables S21 and S22) in *SI Appendix*, section 7.

Result 3 (virtual teams and retention): For up to 3 mo after the end of the experiment, drivers in the team leaderboard treatment work an average of 1 h (3.16%) longer per week than those in the control group. At the city level, Taiyuan drivers in the team leaderboard treatment work 3 h (11.83%) longer per week, whereas treated drivers in Beijing and Kunming do not behave differently from those in their respective control groups.

To better understand driver incentives within each team, we conduct analyses on driver preferences to be a team captain based on our preregistered *Hypothesis 4*, which postulates that drivers with higher productivity prior to our experiment, a longer tenure on the platform, and previous contest captain positions will be more likely to volunteer to be a team captain. We indeed reject the null in favor of *Hypothesis 4* (*SI Appendix*, Table S23).

To rule out the possibility that captains are the main drivers of our treatment effects, we rerun all analyses excluding team captains and find that our results are robust to this specification (*SI Appendix*, section 8). This indicates that captains are not the only people benefiting from the team contests.

Exploring Underlying Mechanisms. In this section, we explore the underlying mechanisms that drive the treatment effects. More specifically, we investigate three channels—social comparison, momentum effects, and team communication. To explore the effects of social comparison, we partition the drivers into two subgroups by whether their preexperiment revenue was above or below the median in their respective city in the 2 wk prior

---

‖ *SI Appendix*, Table S13 presents treatment effects on revenue per hour during the postintervention contest and shows that there is no significant treatment effect on revenue per hour overall ($P > 0.10$, columns 1 and 5) or in any of the three cities ($P > 0.10$, columns 2–4 and 6–8).

to the start of our experiment and report the subgroup analysis in *SI Appendix*, section 9. As shown in *SI Appendix*, Fig. S5, below-median drivers consistently generate a larger revenue increase than their above-median counterparts in the preintervention, status, and postintervention contests. More specifically, in the preintervention contest (*SI Appendix*, Table S32), below-median drivers generate a 628.50 CNY revenue increase compared to their above-median counterparts ($P < 0.01$). This asymmetric effect has been observed in other social comparison field experiments in the context of online movie ratings (45) and traffic violations (46). While the effects in the preintervention contest could be attributed to any combinations of social comparison, team identity, and monetary rewards, the latter is removed in the 3-wk status contest. Specifically, social comparison and team identity remain present among treated drivers, whereas only social comparison is available to drivers in the control condition if they continue to use the social information from the preintervention contest as a reference point. We summarize our subgroup analyses below.

Result 4 (social comparison: below- vs. above-median drivers): Below-median drivers demonstrate a significant increase in revenue compared to their above-median counterparts in each experimental condition. The fact that below-median drivers in the control condition outperform their above-median counterparts by 782.07 CNY/wk ($P < 0.01$; *SI Appendix*, Table S33) during the intervention indicates that information provision alone could sustain better performance for the below-median drivers.

Result 4 is consistent with Festinger's social comparison theory (57), which posits that we compare ourselves to others who are better off for guidance (upward comparison) and to others who are worse off to increase our self-esteem (downward comparison). A large body of literature in economics and social psychology shows that social comparison affects behavior; however, there is no consensus on when the effects for upward or downward comparison dominate. In the context of exercise, researchers show that the effects for downward comparison are larger than those for upward comparison (37). However, when participants' efforts contribute to public goods, upward comparison dominates downward comparison (45, 46). Since our drivers' revenue contributes to their individual as well as their team's ranking, our context has a public goods component. Consistent with prior literature, we

also find that the effects for the below-median drivers (upward comparison) are larger.

Related to social comparison, we investigate contest dynamics and find that drivers whose team won the preintervention contests show a significant and sizeable increase in revenue in every experimental condition (*SI Appendix*, Table S35), i.e., a momentum effect, which is also identified in other empirical contest literature (58).

Figs. 2 and 3 present the ranking dynamics throughout our experiment at the individual and team levels, respectively. We note that both individuals and teams are likely to move up or down one to two ranks, whereas radical rank changes are less often observed. Despite considerable movements, both individual ranking within a team and team ranking within a leaderboard are moderately and significantly correlated with those in the preintervention contest (Spearman's rank correlation coefficient between the last day of the preintervention contest and that of the status contest is $\rho = 0.65$ for individual ranking and $\rho = 0.52$ for team ranking, $P < 0.01$ in each case). We summarize our analyses below.

Result 5 (momentum effect): Drivers whose team won the preintervention contest show a sizeable increase in revenue in every experimental condition ($P < 0.01$; *SI Appendix*, Table S35). The rank correlation coefficient is 0.65 for individual ranking and 0.52 for team ranking ($P < 0.01$).

We next explore the channel of within-team communication on driver revenue, as prior laboratory experiments demonstrate that within-group communication is crucial in strengthening group identity (13, 18). To facilitate between-driver communications, prior to our experiment, the platform implemented a direct message forum within its app, so that drivers can send direct messages to each other. However, the forum allows only pairwise communications, which are clumsy to use to communicate with a team. Indeed, we saw forum messages where a driver sent the same message six times, once to each of the driver's teammates. Based on our survey data, 79% of the respondents said that their team moved to WeChat, which facilitates easier group communication with functionalities similar to those of WhatsApp. The forum data corroborates this view, as 77% of the forum messages were communicated during the team building week and 93% before the start of the intervention. However, we do not have access to their WeChat communication data, as it belongs to a different
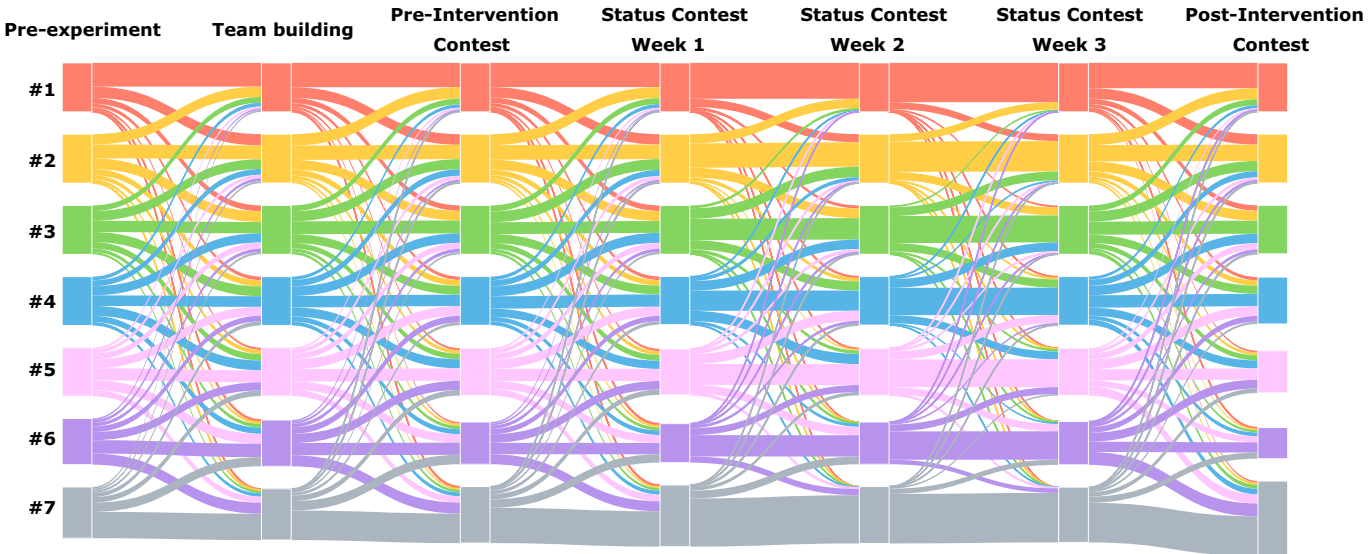

Fig. 2. Sankey diagram of individual ranking dynamics within the individual's team. The *x* axis denotes the week and the *y* axis denotes ranking, with #1 being the top rank. The width of the branches is proportional to the flow rate.
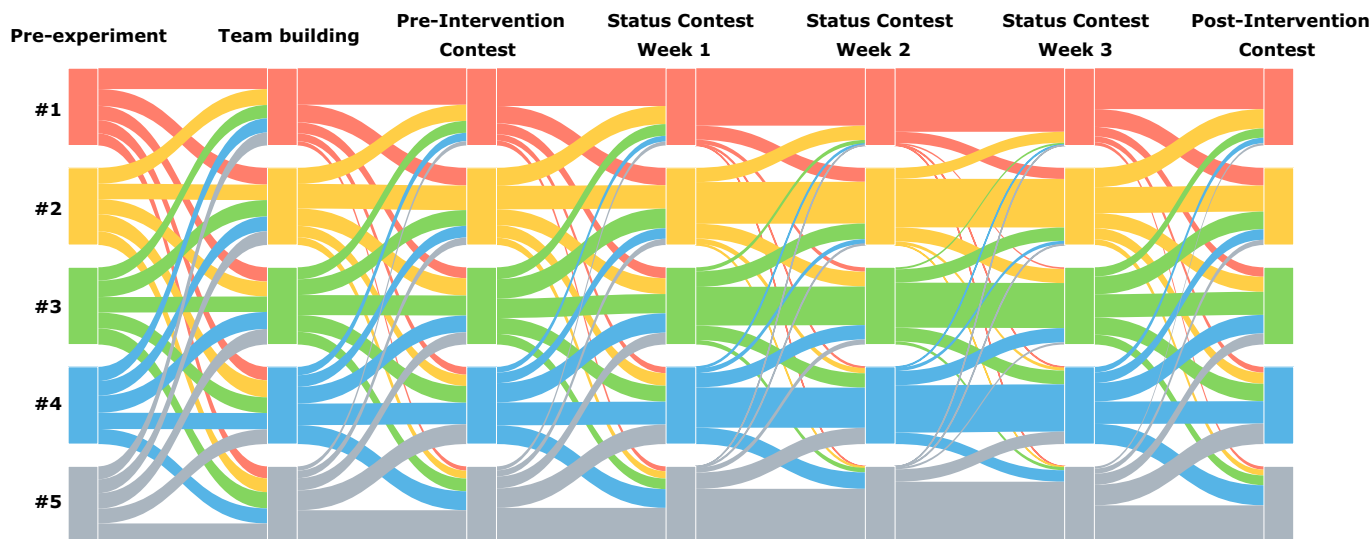
Fig. 3. Sankey diagram of team ranking dynamics within its leaderboard. The *x* axis denotes the week and the *y* axis denotes ranking, with #1 being the top rank. The width of a branch is proportional to the flow rate.

IT company. Nonetheless, we use "team forum usage," a binary variable to indicate whether at least one team member used the forum to contact teammates between the team formation and the status contest phases. Our analysis of team forum usage is reported in *SI Appendix*, section 11 and summarized below.

Result 6 (team communication): Drivers in Beijing and Taiyuan are significantly more likely to use the team forum compared to their counterparts in Kunming ($P < 0.01$; *SI Appendix*, Table S38). Furthermore, team forum usage is associated with a significant increase in weekly revenue overall (32.31 CNY, $P < 0.05$; column 1 in *SI Appendix*, Table S39) and in Taiyuan (60.93 CNY, $P < 0.01$; column 3 in *SI Appendix*, Table S39).

Result 6 corroborates findings in the management literature that a participatory organization structure that enables workers to voice opinions in their group increases worker productivity (35). In our setting, since team forum usage is not randomly assigned, we cannot conclude that team communication is the cause of improved performance. Nonetheless, *Result 6* provides correlational evidence that team communication is associated with increased revenue.

While we do not have the content of drivers' WeChat communication, our postexperiment survey (*SI Appendix*, section 4) yields some insights into the nature of their information exchange, even though the survey response rate is only 15%. Of the 4,295 drivers who responded, more than 82% like the contests (Q1), citing team belonging (Q17), making friends (Q2, Q6), and identification with the organization (Q18) as benefits. We also find evidence of peer information exchange, learning, and skill improvement among team members (Q4e), providing empirical evidence for information sharing in teams (59).

A Machine-Learning Analysis of 149 Team Contests. To further explore the robustness of findings in our experiment, we perform a machine-learning analysis using the data of our six short-term, pre- and postintervention contests (i.e., two per city) together with an additional dataset comprised of 143 short-term team contests in 86 cities conducted by the platform between January and August 2018 (22). The common features among these contests are that they are short term (3 to 10 d), with a best-of-five team contest format, and with a cash prize for the best-performing team. Our status contest is unique among all the contests run on the platform in that it does not provide additional monetary prizes. While it is not possible to use other status contests to check

the robustness of our results, since the short-term contests contain identical leaderboard information, they are useful to validate our main findings. *SI Appendix*, section 12 contains our sample selection criteria, the implementation details of the random forest model (60), and its performance metrics.[**]

Our set of features includes driver age, work experience (platform age), the driver's precontest revenue difference from the team average, team similarity (based on age, hometown, precontest work area, etc.), and precontest ranking on the team leaderboard, as well as city-level precontest order fulfillment rate, prize amount, the presence of individual threshold bonus, the number of drivers on the platform in a city, and the weather condition (rain or snow).

SI Appendix, Table S43 presents the features in decreasing importance in predicting individual treatment effects. The feature importance ranking provides a robustness check for our main results. The most important feature, the difference between a driver's precontest revenue and the team average, accounts for 32% of the node impurity score. This is consistent with *Result 4* that below-median drivers exhibit a stronger revenue increase than their above-median counterparts. It is also consistent with *Result 5* (momentum effect) that those who won the preintervention contest continue to do better in the intervention. The second most important feature, a driver's work experience, accounts for 13% of the node impurity score, corroborating similar findings in our regression analysis (Tables 1–3) and those using Uber data (25). Next, the three team similarity measures (precontest work area, age, hometown) together account for 29% of the score, followed by city-level precontest order fulfillment rate, prize amount, the presence of individual threshold bonus, and inclement weather.

Our machine-learning analysis provides a robustness check for the main mechanism identified by our heterogeneity analyses (social comparison). It also helps explain the different treatment effects across our three contest cities. The effects observed in Beijing are affected by the presence of individual threshold bonus

---

[**]A random forest model integrates a number of decision trees, each of which uses a random sample of the training data. In each tree, it iteratively examines the set of possible splitting points among a random subset of features. Given a splitting point, it calculates the reduction of node impurity, which is measured by the mean-squared error when the outcome variable is continuous, before and after the split. A larger reduction of node impurity represents better prediction performance. The feature importance (in making accurate predictions) is then measured by the extent to which a feature helps reduce the node impurity across the "forest."

implemented by another department, whereas the lack of any treatment effect in Kunming might be explained by the near-perfect precontest order fulfillment rate (98%). Absent this type of interference and with a high level of within-team similarity as most of its drivers are local, Taiyuan provides the ideal environment for testing the efficacy of team contests. The importance of within-team similarity in natural identities, such as hometown or age, is consistent with social network models of friendship formation (39).

Conclusion

Our study examines the effect of virtual teams on gig worker productivity and retention on a ride-sharing platform. We find that treated drivers generate significantly more revenue than those in the control condition during the 3-wk intervention. Three months after the experiment ended, we find that drivers in the team leaderboard treatment continue to work longer hours on the platform, indicating that virtual teams have the potential to increase worker productivity and retention. We identify social comparison, momentum effect, and team communication as the underlying mechanisms that drive the treatment effects. Our machine-learning analysis with 149 short-term team contests provides robustness checks for our main findings and explains the heterogeneous treatment effects across the three cities. This research points to the promise of virtual teams for the gig economy and for the future of work.

More broadly, our research contributes to the understanding of nonmonetary incentives. For the team status contest to work, three conditions should be satisfied. First, within-team diversity in productivity activates social comparison. The public good nature of team ranking promotes upward comparison that increases below-median workers' effort. Second, within-team similarity in natural identities, such as hometown or age, facilitates team communication, identity building, and friendship formation. Finally,

a gap in the labor supply and work demand ensures that there is room for improvement and that an increase in labor supply is efficient.

Author affiliations: [a]Department of Information and Decision Sciences, Carlson School of Management, University of Minnesota, Minneapolis, MN 55455-9940; [b]College of Information Studies, University of Maryland, College Park, MD 20742; [c]School of Information, University of Michigan, Ann Arbor, MI 48109; [d]Department of Economics, School of Economics and Management, Tsinghua University, Beijing 100084, China; [e]Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI 48109; and [f]School of Computer Science and Technology, Shandong University, Qingdao 266237, China

1. S. Mcfeely, R. Pendell, What workplace leaders can learn from the real gig economy (2018). https://www.gallup.com/workplace/240929/workplace-leaders-learn-real-gig-economy.aspx. Accessed 19 March 2021.
2. M. K. Chen, J. A. Chevalier, P. E. Rossi, E. Oehlsen, The value of flexible work: Evidence from uber drivers. *J. Polit. Econ.* **127**, 2735–2794 (2019).
3. N. Heller, Is the gig economy working? *The New Yorker* (May 8, 2017). https://www.newyorker.com/magazine/2017/05/15/is-the-gig-economy-working. Accessed 1 November 2022.
4. A. J. Ravenelle, *Hustle and Gig: Struggling and Surviving in the Sharing Economy* (University of California Press, Oakland, CA, ed. 1, 2019).
5. L. Saad, B. Wigert, Remote work persisting and trending permanent (2021). https://news.gallup.com/poll/355907/remote-work-persisting-trending-permanent.aspx. Accessed 26 January 2022.
6. H. Tajfel, J. Turner, "The social identity theory of intergroup behavior" in *The Social Psychology of Intergroup Relations*, S. Worchel, W. Austin, Eds. (Nelson-Hall, Chicago, IL, 1986), pp. 7–24.
7. M. B. Brewer, The psychology of prejudice: Ingroup love and outgroup hate? *J. Soc. Issues* **55**, 429–444 (1999).
8. G. A. Akerlof, R. E. Kranton, Economics and identity. *Q. J. Econ.* **115**, 715–753 (2000).
9. G. A. Akerlof, R. E. Kranton, *Identity Economics: How Our Identities Shape Our Work, Wages, and Well-Being* (Princeton University Press, Princeton, NJ, 2010).
10. R. Fryer, M. O. Jackson, A categorical model of cognition and biased decision making. *B.E. J. Theor. Econ.* **8**, 1–42 (2008).
11. C. C. Eckel, P. J. Grossman, Managing diversity by creating team identity. *J. Econ. Behav. Organ.* **58**, 371–392 (2005).
12. G. Charness, L. Rigotti, A. Rustichini, Individual behavior and group membership. *Am. Econ. Rev.* **97**, 1340–1352 (2007).
13. R. Chen, Y. Chen, The potential of social identity for equilibrium selection. *Am. Econ. Rev.* **101**, 2562–2589 (2011).
14. L. Goette, D. Huffman, S. Meier, M. Sutter, Competition between organizational groups: Its impact on altruistic and antisocial motivations. *Manage. Sci.* **58**, 948–960 (2012).
15. Y. Chen, S. X. Li, T. X. Liu, M. Shih, Which hat to wear? Impact of natural identities on coordination and cooperation. *Games Econ. Behav.* **84**, 58–86 (2014).
16. I. Erev, G. Bornstein, R. Galili, Constructive intergroup competition as a solution to the free rider problem: A field experiment. *J. Exp. Soc. Psychol.* **29**, 463–478 (1993).
17. W. Ai, R. Chen, Y. Chen, Q. Mei, W. Phillips, Recommending teams promotes prosocial lending in online microfinance. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 14944–14948 (2016).
18. G. Charness, Y. Chen, Social identity, group behavior, and teams. *Annu. Rev. Econ.* **12**, 691–713 (2020).
19. O. Bandiera, I. Barankay, I. Rasul, Team incentives: Evidence from a firm level experiment. *J. Eur. Econ. Assoc.* **11**, 1079–1114 (2013).
20. G. M. Olson, J. S. Olson, Distance matters. *Hum. Comput. Interact.* **15**, 139–178 (2000).
21. W. Ai, Y. Chen, Q. Mei, J. Ye, L. Zhang, Putting teams into the gig economy: A field experiment at a ride-sharing platform. *Manage. Sci.* (2022), in press.
22. T. Ye et al., "Predicting individual treatment effects of large-scale team competitions in a ride-sharing economy" in *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, R. Gupta, Y. Liu, Eds. (Virtual, 2020), pp. 2368–2377.
23. J. V. Hall, A. B. Krueger, An analysis of the labor market for Uber's driver-partners in the United States. *Ind. Labor Relat. Rev.* **71**, 705–732 (2018).
24. E. Jackson, A. Looney, S. Ramnath, The rise of alternative work arrangements: Evidence and implications for tax filing and benefit coverage. Office of Tax Analysis Working Paper 114, January 2017. https://www.census.gov/content/dam/Census/about/about-the-bureau/adrm/FESAC/meetings/Ramnath%20Background%20Document.pdf. Accessed 1 November 2022.
25. C. Cook, R. Diamond, J. V. Hall, J. A. List, P. Oyer, The gender earnings gap in the gig economy: Evidence from over a million rideshare drivers. *Rev. Econ. Stud.* **88**, 2210–2238 (2021).
26. P. Cohen, R. Hahn, J. Hall, S. Levitt, R. Metcalfe, Using big data to estimate consumer surplus: The case of uber (Working Paper 22627, National Bureau of Economic Research, 2016). https://www.nber.org/papers/w22627. Accessed 1 November 2022.
27. B. Chandar, U. Gneezy, J. A. List, I. Muir, "The drivers of social preferences: Evidence from a nationwide tipping field experiment" (Tech. Rep. 26380, National Bureau of Economic Research, Cambridge, MA, 2019).
28. B. K. Chandar, A. Hortaçsu, J. A. List, I. Muir, J. M. Wooldridge, "Design and analysis of cluster-randomized field experiments in panel data settings" (Tech. Rep. 26389, National Bureau of Economic Research, Cambridge, MA, 2019).
29. B. Halperin, B. Ho, J. A. List, I. Muir, Toward an understanding of the economics of apologies: Evidence from a large-scale natural field experiment. *Econ. J. (Lond.)* **132**, 273–298 (2022).
30. A. Goldszmidt et al., "The value of time in the United States: Estimates from nationwide natural field experiments" (Tech. Rep. 28208, National Bureau of Economic Research, Cambridge, MA, 2020).
31. T. X. Liu, Z. Wan, C. Yang, The efficiency of a dynamic decentralized two-sided matching market. (SSRN, 2019). https://ssrn.com/abstract=3339394. Accessed 1 November 2022.
32. A. E. Roth, The economist as engineer: Game theory, experimentation, and computation as tools for design economics. *Econometrica* **70**, 1341–1378 (2002).
33. A. E. Roth, R. B. Wilson, How market design emerged from game theory: A mutual interview. *J. Econ. Perspect.* **33**, 118–143 (2019).
34. L. Cassar, S. Meier, Nonmonetary incentives and the implications of work as a source of meaning. *J. Econ. Perspect.* **32**, 215–238 (2018).
35. S. J. Wu, E. L. Paluck, Participatory practices at work change attitudes and behavior toward societal authority and justice. *Nat. Commun.* **11**, 2633 (2020).
36. J. Gallus, Fostering public good contributions with symbolic awards: A large-scale natural field experiment at Wikipedia. *Manage. Sci.* **63**, 3999–4015 (2016).

37. S. Aral, C. Nicolaides, Exercise contagion in a global social network. *Nat. Commun.* **8**, 14753 (2017).

38. E. Breza, A. G. Chandrasekhar, Social networks, reputation, and commitment: Evidence from a savings monitors experiment. *Econometrica* **87**, 175–216 (2019).

39. S. Currarini, M. O. Jackson, P. Pin, An economic model of friendship: Homophily, minorities, and segregation. *Econometrica* **77**, 1003–1045 (2009).

40. T. Ye, W. Ai, Y. Chen, M. Qiaozhu, J. Ye, *Inter-Team Status Competition for Ride Sharing: A Large-Scale Field Experiment at a Ride-Sharing Platform* (AEA RCT Registry, 2018).

41. P. Couture, Informed consent in social science. *Science* **322**, 672–672, author reply 672 (2008).

42. K. A. Konrad, *Strategy and Dynamics in Contests* (Oxford University Press, New York, NY, 2009).

43. M. Vojnović, *Contest Theory: Incentive Mechanisms and Ranking Methods* (Cambridge University Press, Cambridge, UK, 2016).

44. R. M. Sheremeta, Behavior in group contests: A review of experimental research. *J. Econ. Surv.* **32**, 683–704 (2018).

45. Y. Chen, F. M. Harper, J. Konstan, S. X. Li, Social comparisons and contributions to online communities: A field experiment on movielens. *Am. Econ. Rev.* **100**, 1358–1398 (2010).

46. Y. Chen, F. Lu, J. Zhang, Social comparisons, status and driving behavior. *J. Public Econ.* **155** (suppl. C), 11–20 (2017).

47. Y. Benjamini, Y. Hochberg, Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. R. Stat. Soc. B* **57**, 289–300 (1995).

48. M. L. Anderson, Multiple inference and gender differences in the effects of early intervention: A reevaluation of the abecedarian, perry preschool, and early training projects. *J. Am. Stat. Assoc.* **103**, 1481–1495 (2008).

49. B. Efron, *Large-Scale Inference: Empirical Bayes Methods for Estimation, Testing, and Prediction* (Cambridge University Press, 2012), vol. 1.

50. G. Charness, D. Masclet, M. C. Villeval, The dark side of competition for status. *Manage. Sci.* **60**, 38–55 (2014).

51. S. Chowdhury, A. Mukherjee, R. Sheremeta, In-group versus out-group preferences in intergroup conflict: An experiment (MPRA Working Paper, 2021). https://mpra.ub.uni-muenchen.de/105690/1/MPRA_paper_105690.pdf. Accessed 1 November 2022.

52. S. Chowdhury, *The Economics of Identity and Conflict* (Oxford University Press (OUP), 2021).

53. M. Shayo, A model of social identity with an application to political economy: Nation, class, and redistribution. *Am. Polit. Sci. Rev.* **103**, 147–174 (2009).

54. M. Bernard, F. Hett, M. Mechtel, Social identity and social free-riding. *Eur. Econ. Rev.* **90**, 4–17 (2016).

55. S. DellaVigna, E. Linos, RCTs to scale: Comprehensive evidence from two nudge units. *Econometrica* **90**, 81–116 (2022).

56. B. Moldovanu, A. Sela, X. Shi, Contests for status. *J. Polit. Econ.* **115**, 338–363 (2007).

57. L. Festinger, A theory of social comparison processes. *Hum. Relat.* **7**, 117–140 (1954).

58. A. Descamps, C. Ke, L. Page, How success breeds success. *Quant. Econom.* **13**, 355–385 (2022).

59. D. Bergemann, S. Morris, Robust predictions in games with incomplete information. *Econometrica* **81**, 1251–1308 (2013).

60. L. Breiman, Random forests. *Mach. Learn.* **45**, 5–32 (2001).

61. T. Ye *et al.*, Data and code for "Virtual teams in a gig economy." (Available via the platform's contact at gaia@didichuxing.com. 28 January 2022).