Virtual Teams in a Gig Economy*

Teng Ye, Wei Ai, Yan Chen, Qiaozhu Mei, Jieping Ye, Lingyu Zhang

May 13, 2021

While the gig economy provides flexible jobs for millions of workers globally, a lack of organization identity and co-worker bonds contributes to their low engagement and high attrition rates (*1*). To test the impact of virtual teams on worker productivity, retention and well-being, 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 generate significantly higher revenue. Furthermore, drivers in the team ranking treatment continue to be more engaged three months after the end of the experiment. Survey data suggest that peer learning and team identity contribute to the virtual team efficacy.

*Y. Chen: School of Information, University of Michigan, 105 South State Street, Ann Arbor, MI 48109-2112. Email: yanchen@umich.edu.

Introduction

According to a recent Gallup poll, 36% of U.S. workers participate in the gig economy as either their primary or secondary jobs (*2*). The gig economy provides workers with the benefits of autonomy and flexibility (*3*), but it does so at the expense of work-related identity and co-worker 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" (*1, 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 as those in a gig economy. By August 2020, 42 percent of the U.S. labor force was identified as working from home full time (*5*). Given that we expect at least some portion of this remote work to remain post pandemic, 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 lab 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 pro-social behavior in fruit harvesting (*16*) and online peer-to-peer pro-social 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 DiDi Chuxing (DiDi), which is the largest ride-sharing platform in Asia. We then examine the effect of team contests on individual driver behavior. Although virtual teams have been studied in the laboratory (*20*), to our knowledge, this is the first large-scale multi-city randomized field experiment to examine the effect of virtual teams.

To design our contests, we draw on insights obtained from an earlier field experiment conducted in the Chinese city of Dongguan in 2017 (*21*). In this earlier experiment, we randomly assigned 2,100 drivers into seven-person teams to compete for a cash prize across a five-day period. Team compositions are determined either randomly or based on homophily 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.

Encouraged by the results of this first field experiment, in 2018, DiDi conducted 1,548 team contests across 180 cities in China, involving over two million drivers placed into teams based on hometown or age similarity. These contests, typically one week 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 DiDi ran in 2018 is that they were all one-week contests with cash incentives, and the teams existed only for the duration of the contest. As a result of this latter factor, the DiDi 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* a monetary incentive to participate in the contest. Specifically, in fall 2018, we conducted a natural field experiment on the DiDi platform involving 27,790 drivers across three cities: Beijing, Kunming, and Taiyuan. The intervention ran for three weeks, 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 monetary incentive in any of the experimental conditions.

Across the three-week 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 post-experiment survey indicates that drivers in virtual teams made friends, shared information about order-acceptance strategies, and learned collaboration skills from their teammates.

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*), identifying factors contributing to the gender wage gap in ride-sharing (*24*), the value of flexible work (*25*), 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, their bonds with co-workers, and their team identity. As such, our field experiment uses insights from behavioral market design to help structure the future of work (*32, 33*).

Experiment Design

As mentioned, we conduct a natural field experiment on the DiDi platform involving 27,790 drivers across Beijing, Kunming, and Taiyuan, three cities chosen to exemplify diversity in location, size, order fulfillment rate, and the number of team contests hosted on DiDi prior to our experiment (Table S1 in the Supplemental Materials, SM henceforth). Our experiment is approved by the University of Michigan IRB (HUM00153090) and pre-registered at the AEA RCT Registry (*34*). The experiment was conducted from October 22, 2018 to December 3, 2018. To evaluate our treatment effect on driver retention, we continue to collect data for three months after our experiment, until March 15, 2019. In addition to our recruitment and team formation stage, our experiment is organized into pre-intervention, intervention, and post-intervention stages. Figure S1 in SM presents the experimental process.

Driver recruitment and team formation. For the driver recruitment and team formation stage, the platform used its built-in process which was informed by our earlier experiment (*21*). The platform sent an invitation on October 22, 2018 to all 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 recruitment stage, the system then randomly selects 90% of the drivers in under-sized 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 two weeks 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. We now describe the three stages of the team contest.

The pre-intervention contest. Following previous studies that find that inter-group competition is among the most successful methods used for creating a strong sense of group identity (*11*), we conduct a pre-intervention 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 DiDi'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.[1] The cash prize is 1,000 RMB per winning team for Beijing, and 650 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 (*35*), and is credited to their driver accounts immediately after the contest.

During this stage, all drivers participating in the contest can use the DiDi app to access both

[1]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 plate number.

a team leaderboard and an individual leaderboard for social information, as illustrated in Figure S2 in SM. The team leaderboard shows the cumulative revenue of each of the five teams in the contest group in descending order (top left panel of Figure S2). 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 (top right panel of Figure S2). In addition, we mark the average performance of that team with a line on the individual leaderboard to enhance the effect of ranking (*36, 37*). 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 Section 1.4 of SM.

The intervention: A status contest. Immediately after the pre-intervention contest, we randomly assign each leaderboard to one of three experimental conditions and conduct a three-week status contest between November 5-25 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 pre-intervention 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 their team name on the team leaderboard (the default interface), they can further access the individual leaderboard within their 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 (bottom panel of Figure S2).

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 two weeks prior to the start of the experiment. Kolmogorov-Smirnov tests show that the distribution of pre-experiment revenue, age, gender, length of time with DiDi, team formation, and hometown distance to the contest city is not significantly different in pairwise comparisons across the three conditions within each city ($p > 0.10$, Table S4 in SM). Table S4 also reveals interesting facts about our drivers: more than 95% of them are male, with an average age of 37. Looking at their hometown distance to the contest city, we conclude that Taiyuan drivers are predominantly local, whereas Beijing and Kunming drivers are mostly migrants. In China, DiDi drivers comprise of workers laid off from their traditional jobs, veterans, migrant workers from rural areas, and commuters who offer rides during their daily commute.

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

The post-experiment survey. After the post-intervention contest, all drivers receive a survey which evaluates whether they like the status contest, what they get out of the contest, their sense of belonging related to their team as well as to the organization (DiDi). The survey questions and responses are included in Section 9 of SM.

Results

Our experiment yields findings related to the immediate and longer-term effects of virtual teams on driver revenue and retention, both overall (*34*) and at the city level. On the DiDi 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.

We first examine the average treatment effect on driver revenue during the experiment period. In Figure 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 pre-experiment period. The y-axis presents the revenue difference between a given week and the baseline week in the pre-experiment period. Note that the three lines coincide up to the start of the pre-intervention contest period. However, during the status contest intervention, since drivers in different treatment conditions receive different social information, the lines in Figure 1a start to diverge. Pooling all three cities, we observe that our treatment drivers are more productive on average than those in the control condition both during and after the intervention.

In our first pre-registered hypothesis, based on prior laboratory experiments on social identity and team competition (*11*) as well as those on individual performance ranking (*38*), 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 (*39, 40*).

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 panel regression model:

$$\Delta \mathrm{Y}_{i,t} = \beta_0 + \beta_1 \text{Treated} + \alpha_c + \epsilon_{i,t}, \tag{1}$$

where $\Delta \mathrm{Y}_{i,t}$ represents the outcome change in the t-th week in the current period compared to the corresponding pre-contest week(s),[2] and α_c captures city fixed effects. Hypothesis 1(a) implies that $\beta_1 > 0$ in Equation (1).

We report the main results in Tables 1 to 4 in the main text, and the results of our robustness checks in SM. To correct for multiple hypothesis testing, we report the false discovery rate adjusted q-values in square brackets (*41, 42*). To claim significance, we use a 5% (10%) cutoff for our p-values (q-values) (*43*).

The results in column 1 of Table 1 show that our treatment conditions increase driver revenue by 34.53 RMB, or 1.66% of the average weekly revenue per driver, during the three-week intervention ($p < .05$). Therefore, we reject the null hypothesis in favor of Hypothesis 1(a). We further find a significant treatment effect for drivers in Beijing (41.67 RMB, $p < .05$, 1.69% of average weekly revenue), but not in Taiyuan or Kunming. Our findings are strengthened (39.08 RMB, or 1.88% of average weekly revenue) when we control for demographics and self versus system formation of teams (columns 5-8). Consistent with social identity theories focusing on the effects of social status and social distance on individual identification with social groups (*44, 45*), a driver's hometown distance from the contest city is negatively correlated with their productivity ($p < 0.05$, column 7). Interestingly, self-formed teams generate lower revenue compared to system-formed teams using hometown or age similarity ($p < 0.01$, columns 5-6). We also note that older drivers and those who have joined the platform earlier generate higher

[2]For the pre-experiment baseline week(s), we use the week before the experiment (Oct. 15-21, 2018) for one-week target periods, i.e., the pre-intervention contest, the post-intervention contest, and retention. For the status contest, we use the two weeks before the experiment (Oct. 8-21, 2018) as our baseline, as the week of October 1-7, 2018 was a national holiday with drastically different demand and supply for ride-sharing.

revenue.

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 that $\beta_1 > \beta_2$ in Equation (2) below.

$$\Delta \mathrm{Y}_{i,t} = \beta_0 + \beta_1 \text{Team Leaderboard} + \beta_2 \text{Individual Leaderboard} + \alpha_c + \epsilon_{i,t}, \tag{2}$$

The results in column 1 of Table 2, show that drivers in the team leaderboard treatment generate 32.12 RMB marginally higher weekly revenue compared with the control group ($p < .10$), while those in the individual leaderboard condition generate 36.96 RMB higher revenue compared with the control group ($p < .05$). After controlling for demographics and self versus system formation of teams (column 5), we see that the team (individual) leaderboard generates 36.70 RMB (41.47 RMB) more weekly revenue, equivalent to a 1.76% (1.99%) increase ($p < .05$ in each case), although the difference between the two treatments is not significant ($p > .10$).

We next examine our city-level results. From Table 2, we see that only the individual leaderboard treatment has a significant effect on revenue (56.32 RMB per week, or 2.29% of the weekly revenue of the control group, $p < .05$) for drivers in Beijing (columns 2 and 6), whereas in Taiyuan (columns 3 and 7), only the team leaderboard treatment has a significant effect on revenue compared to the control condition (58.49 RMB per week, $p < .05$). By contrast, neither treatment has a significant effect on weekly revenue for drivers in Kunming. As shown in Table S1 in SM, passenger order fulfillment rate was already quite high (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. After controlling for demographics and team formation methods (columns 6-8 of Table 2), the city-level treatment effects remain statistically and economically significant, with the size of the individual and team leaderboard effect equal to 59.24 RMB in Beijing and 62.31 RMB in Taiyuan ($p < .05$ in each case). Furthermore, the difference between the two treatments in Taiyuan is in the direction we hypothesize (Hypothesis 1b), albeit marginally significant ($p < .10$, columns 3 and 7). This leads to our first main result.

Result 1 (Virtual teams and productivity). During the three-week status contest intervention, (1) treated drivers generate 1.9% higher revenue than those in the control condition; (2) drivers in the team (individual) leaderboard treatment generate 1.8% (2%) higher revenue than those in the control condition; and (3) at the city level, the team (individual) leaderboard treatment leads to a 5.3% (2.3%) increase in driver 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 (*46*).

We are also interested in the question of whether our team effect persists over time. To evaluate the short-term effect, we implement a one-week best-of-five contest with a monetary reward immediately after the intervention. The post-intervention contest rules are identical to those of the pre-intervention contest. We did not announce this contest until the three-week status contest was over. We expect that the treatment effects will persist during this post-intervention contest (pre-registered Hypothesis 2).

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 post-intervention contest.*

The results in column 1 of Table 3 show that drivers in the team leaderboard treatment generate 49.91 RMB higher weekly revenue during our post-intervention contest, or a 2.49% increase, compared to those in the control group ($p < .05$). By contrast, drivers in the individual leaderboard treatment do not differ significantly from those in the control group ($p > .10$). Although the coefficient for the team leaderboard dummy is greater than that for the individual leaderboard, this difference is not significant at the aggregate level ($p = .11$, column 1). Our results are strengthened when we control for demographics as well as self versus system formation of teams (column 5).

At the city level, Beijing drivers in the team leaderboard treatment generate 59.89 RMB marginally higher revenue during our post-intervention contest than those in the control group ($p < .10$, column 2), and significantly higher when we control for demographics and team formation methods (67.20 RMB, $p < .05$, 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 do not differ significantly from those in the control group ($\beta_1 = 58.03$ RMB, $p = .12$), but do generate significantly higher revenue during our post-intervention contest than those in the individual leaderboard treatment ($\beta_1 \neq \beta_2$, $p < .01$ in columns 3 and 7). It is worth noting that those in the individual leaderboard treatment exhibit a marginally significant reduction in average weekly revenue during the post-intervention contest compared to the control group (-68.26 RMB, $p < .10$, columns 3 and 7). Again, we observe no treatment effect for drivers in Kunming (columns 4 and 8). Based on a theoretical model of individual status contests (*47*), 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 one-week post-intervention contest, drivers in the team leaderboard treatment continue to generate 2.49% more weekly revenue compared to those in the control group, whereas the individual leaderboard treatment no longer has an effect.

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 (our pre-registered Hypothesis 3).

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

To examine the effect of team membership on driver retention, we measure driver retention one week, one month, and three months 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. Furthermore, it is possible those who have quit driving may still log into the app. Therefore, we use whether they drive for the platform in a given day rather than app login as our retention measure. Specifically, we measure retention as the number of days that a driver provides at least one ride and separately analyze retention during the week immediately (Table S7), one month (Table S9), and three months (Table 4) after the post-intervention contest.

As shown in Figure 2, drivers in the team leaderboard treatment consistently exhibit higher retention than those from either of the other experimental conditions. From Table 4, we see that drivers in the team leaderboard treatment on average work 0.10 days (or an hour) more than those in the control group in the week three months after the experiment ended ($p < 0.01$, columns 1 and 5). Furthermore, we find that drivers in the team leaderboard treatment also outperform those in the individual leaderboard treatment ($p = 0.02$, columns 1 and 5). The effect is robust and the effect size is stable across different time windows (Tables S7 and S9 in SM). 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. These results are robust after controlling for demographic covariates and team characteristics, such as whether a team has won the post-intervention contest.

Examining our city-level results, columns 2 to 4 in Table 4 show 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 days, $p < 0.01$), with a similar significant effect between the team and individual leaderboard treatments. In Kunming, we find a positive albeit insignificant effect of the team leaderboard treatment on retention only during the one-week window compared to the control group (0.22 days, $p < 0.05$, Table S7). We observe no significant difference between treatments for drivers in Beijing. We summarize the results of our driver retention analysis below.

Result 3 (Virtual teams and retention). For up to three months after the end of the experiment, drivers in the team leaderboard treatment work an average of 0.1 days longer per week than those in the control group. At the city level, Taiyuan drivers in the team leaderboard treatment work 0.3 days 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 last pre-registered hypothesis.

Hypothesis 4 (Leadership) *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.*

Using logit specifications, we report our findings in Table S12 in SM and summarize the result below, which rejects the null in favor of Hypothesis 4.

Result 4 (Leadership). Drivers with a higher revenue, a longer tenure on the platform, and previous contest captain positions prior to our experiment are more likely to volunteer to be team captains.

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 (Tables S5 for Hypothesis 1, S6 for Hypothesis 2, and S8, S10, and S11 for Hypothesis 3). This indicates that captains are not the only people benefiting from the team contests.

To understand who benefits more from virtual team contests, we partition the drivers into two subgroups by whether their pre-experiment revenue was above or below the median in their respective city. As shown in Figure S3, below-median drivers consistently generate a larger revenue increase than their above-median counterparts in the pre-intervention, status, and post-intervention contests. More specifically, in the pre-intervention contest (Table S13 in SM), below-median drivers generate a 628.50 RMB revenue increase compared to their above-median counterparts ($p < .01$). This asymmetric effect has been observed in other information intervention field experiments (*36, 37*), and could be attributed to any combinations of social information, team identity, and monetary rewards. When the latter is removed in the three-week status contest, social information and team identity remain present among treated drivers, whereas none of the three channels is available to drivers in the control condition, although we cannot rule out the possibility that drivers in the control condition continue to use the social information from the pre-intervention contest as a reference point. The fact that below-median drivers in the control condition continue to outperform their more productive counterparts during the three week intervention indicates that social information alone could sustain better performance for workers who used to be lagging behind.

At the end of our experiment, we sent out a survey to all drivers (SM Section 9). While the survey response rate is only 15%, feedback from the 4,295 drivers who completed the survey yields insights on how drivers benefit from virtual teams. More than 82% of the drivers 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 (*48*).

Discussion and Conclusion

Our study examines the effect of virtual teams on gig worker productivity, retention, and well-being on a ride-sharing platform. Using a large-scale natural field experiment with 27,790 drivers, we organize drivers into virtual teams and randomly assign teams to one of three conditions: team leaderboard, individual leaderboard, and no leaderboard (control) conditions. We find that treated drivers generate significantly more revenue than those in the control condition during the three-week 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. Using survey data, we find evidence of peer information sharing and learning, as well as increased team belonging, co-worker bonds, and organization identity. This research points to the promise of virtual teams for the gig economy and for the future of work more generally.

References

1. A. J. Ravenelle, *Hustle and Gig: Struggling and Surviving in the Sharing Economy* (University of California Press, Oakland, California, 2019), first edn.

2. S. Mcfeely, R. Pendell, What workplace leaders can learn from the real gig economy (August 2018 (retrieved March 19, 2021)).

3. M. K. Chen, P. E. Rossi, J. A. Chevalier, E. Oehlsen, *Journal of Political Economy* **127**, 2735 (2019).

4. N. Heller, *The New Yorker* (2017).

5. J. M. Barrero, N. Bloom, S. J. Davis, Covid-19 is also a reallocation shock (2020). Working paper available at SSRN.

6. H. Tajfel, J. Turner, *The Social Psychology of Intergroup Relations*, S. Worchel, W. Austin, eds. (Nelson-Hall, Chicago, 1986).

7. M. B. Brewer, *Journal of Social Issues* **55**, 429 (1999).

8. G. A. Akerlof, R. E. Kranton, *Quarterly Journal of Economics* **115**, 715 (2000).

9. G. A. Akerlof, R. E. Kranton, *Identity Economics: How Our Identities Shape Our Work, Wages, and Well-Being* (Princeton University Press, Princeton, New Jersey, 2010).

10. R. Fryer, M. O. Jackson, *The B.E. Journal of Theoretical Economics* **8** (2008).

11. C. C. Eckel, P. J. Grossman, *Journal of Economic Behavior & Organization* **58**, 371 (2005).

12. G. Charness, L. Rigotti, A. Rustichini, *The American Economic Review* **97**, 1340 (2007).

13. R. Chen, Y. Chen, *The American Economic Review* **101**, 2562 (2011).

14. L. Goette, D. Huffman, S. Meier, M. Sutter, *Management Science* **58**, 948 (2012).

15. Y. Chen, S. X. Li, T. X. Liu, M. Shih, *Games and Economic Behavior* **84**, 58 (2014).

16. I. Erev, G. Bornstein, R. Galili, *Journal of Experimental Social Psychology* **29**, 463 (1993).

17. W. Ai, R. Chen, Y. Chen, Q. Mei, W. Phillips, *Proceedings of the National Academy of Sciences* **113**, 14944 (2016).

18. G. Charness, Y. Chen, *Annual Review of Economics* (2020).

19. O. Bandiera, I. Barankay, I. Rasul, *Journal of the European Economic Association* **11**, 1079 (2013).

20. G. M. Olson, J. S. Olson, *Human–Computer Interaction* **15**, 139 (2000).

21. W. Ai, Y. Chen, Q. Mei, J. Ye, L. Zhang, *Working paper* (2019).

22. T. Ye, *et al.*, *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining* (2020), pp. 2368–2377.

23. J. V. Hall, A. B. Krueger, *ILR Review* **71**, 705 (2018).

24. C. Cook, R. Diamond, J. Hall, J. A. List, P. Oyer, The gender earnings gap in the gig economy: Evidence from over a million rideshare drivers. Review of Economic Studies (forthcoming).

25. M. K. Chen, J. A. Chevalier, P. E. Rossi, E. Oehlsen, *Journal of Political Economy* **127**, 2735 (2019).

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).

27. B. Chandar, U. Gneezy, J. A. List, I. Muir, The drivers of social preferences: Evidence from a nationwide tipping field experiment, *Tech. rep.*, National Bureau of Economic Research (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.*, National Bureau of Economic Research (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, *Tech. rep.*, National Bureau of Economic Research (2019).

30. A. Goldszmidt, *et al.*, The value of time in the united states: Estimates from nationwide natural field experiments, *Tech. rep.*, National Bureau of Economic Research (2020).

31. T. X. Liu, Z. Wan, C. Yang, The efficiency of a dynamic decentralized two-sided matching market (2018). University of Rochester Working Paper.

32. A. E. Roth, *Econometrica* **70**, 1341 (2002).

33. A. E. Roth, R. B. Wilson, *Journal of Economic Perspectives* **33**, 118 (2019).

34. T. Ye, W. Ai, Y. Chen, Q. Mei, J. Ye, *AEA RCT Registry* (2018).

35. R. M. Sheremeta, *Journal of Economic Surveys* **32**, 683 (2018).

36. Y. Chen, F. M. Harper, J. Konstan, S. X. Li, *American Economic Review* **100**, 1358 (2010).

37. Y. Chen, F. Lu, J. Zhang, *Journal of Public Economics* **155**, 11 (2017).

38. G. Charness, D. Masclet, M. C. Villeval, *Management Science* **60**, 38 (2014).

39. S. Chowdhury, A. Mukherjee, R. Sheremeta (2021).

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

41. Y. Benjamini, Y. Hochberg, *Journal of the Royal Statistical Society, Series B (Methodological)* **57**, 289 (1995).

42. M. L. Anderson, *Journal of the American Statistical Association* **103**, 1481 (2008).

43. B. Efron, *Large-Scale Inference: Empirical Bayes Methods for Estimation, Testing and Prediction* (Stanford University Online Manuscript, 2010).

44. M. Shayo, *American Political Science Review* **103**, 147 (2009).

45. M. Bernard, F. Hett, M. Mechtel, *European Economic Review* **90**, 4 (2016). Social identity and discrimination.

46. S. DellaVigna, E. Linos, Rcts to scale: Comprehensive evidence from two nudge units, *Working Paper 27594*, National Bureau of Economic Research (2020).

47. B. Moldovanu, A. Sela, X. Shi, *Journal of Political Economy* **115**, 338 (2007).

48. D. Bergemann, S. Morris, *Econometrica* **81**, 1251 (2013).

49. C. Bellemare, L. Bissonnette, S. Kröger, *Journal of the Economic Science Association* **2**, 157 (2016).

Acknowledgments

We thank Subhasish Chowdhury, Alain Cohn, Jim Cox, Glenn Harrison, John Ledyard, Steve Leider, Yuqing Ren, Tanya Rosenblat, Katya Sherstyuk, and seminar participants at Georgia State University, University of Bath, University of California at Berkeley, University of Innsbruck, University of Michigan, University of Minnesota, Shanghai Jiao Tong University, Tsinghua University and the 2019 North America ESA Meetings (LA, CA) for helpful discussions and comments, and Miao Liang, Tao Song, Quanjiang Wan, Guobin Wu, and Lulu Zhang for their help in implementing the experiment. The research has been approved by the University of Michigan IRB (HUM00153090) and pre-registered at AEA RCT registry (AEARCTR-0003537). Financial support from DiDi Chuxing through the Michigan Institute for Data Science is gratefully acknowledged.

Supplementary materials

Experiment Design Details

Power Analysis

Randomization Check

Robustness Checks and Additional Results

Post-experiment Survey

Figures S1 to S3

Tables S1 to S17

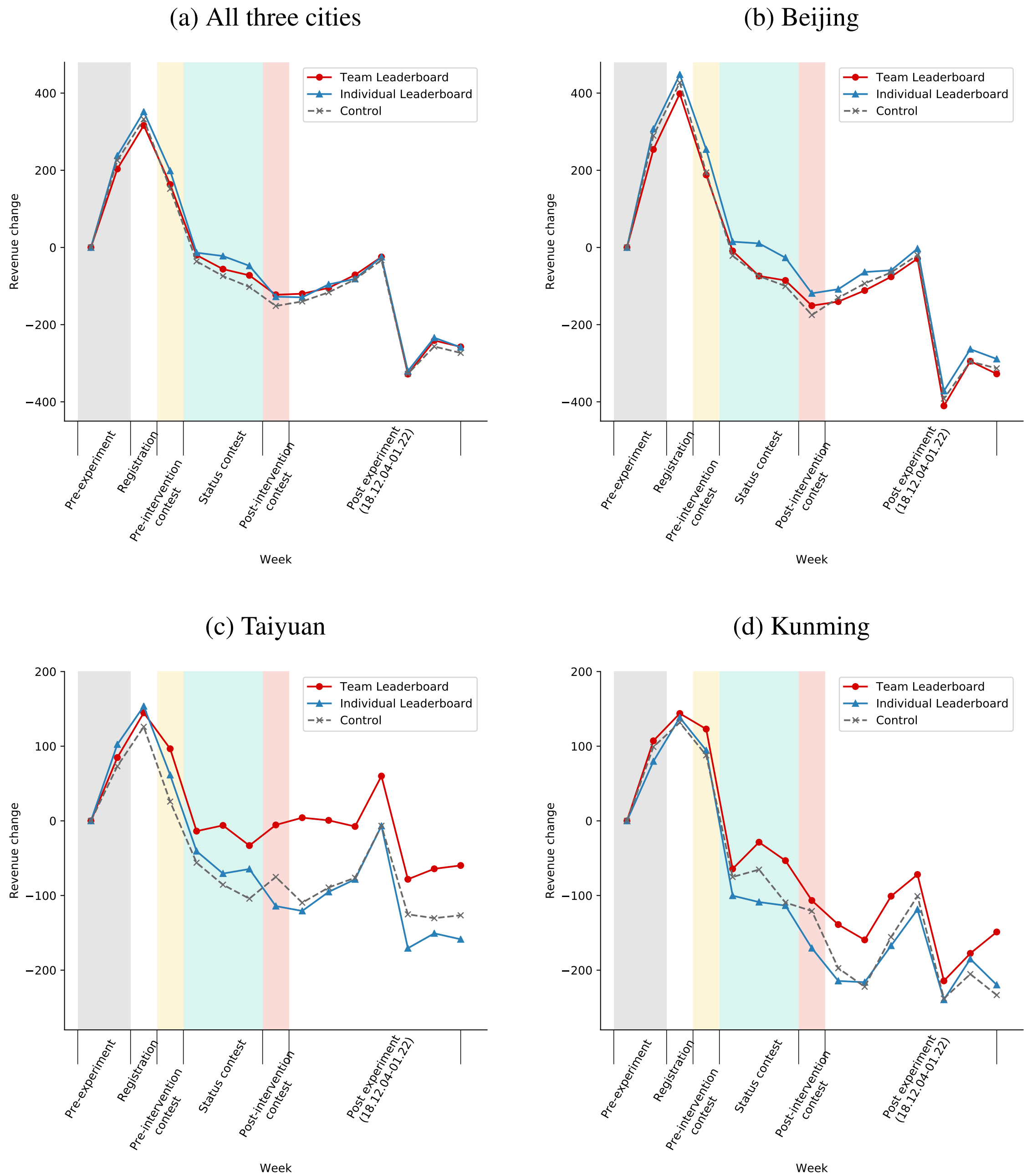

Figure 1: Average weekly driver revenue under each experimental condition. To better visualize the changes over time, we re-scale the revenue within each experimental condition with reference to its pre-experiment average weekly revenue from the week of October 8-14, i.e., two weeks before the start of the experiment. For example, each point represents the weekly average revenue per driver under that experimental condition minus the pre-experiment weekly average revenue per driver under the same experimental condition.

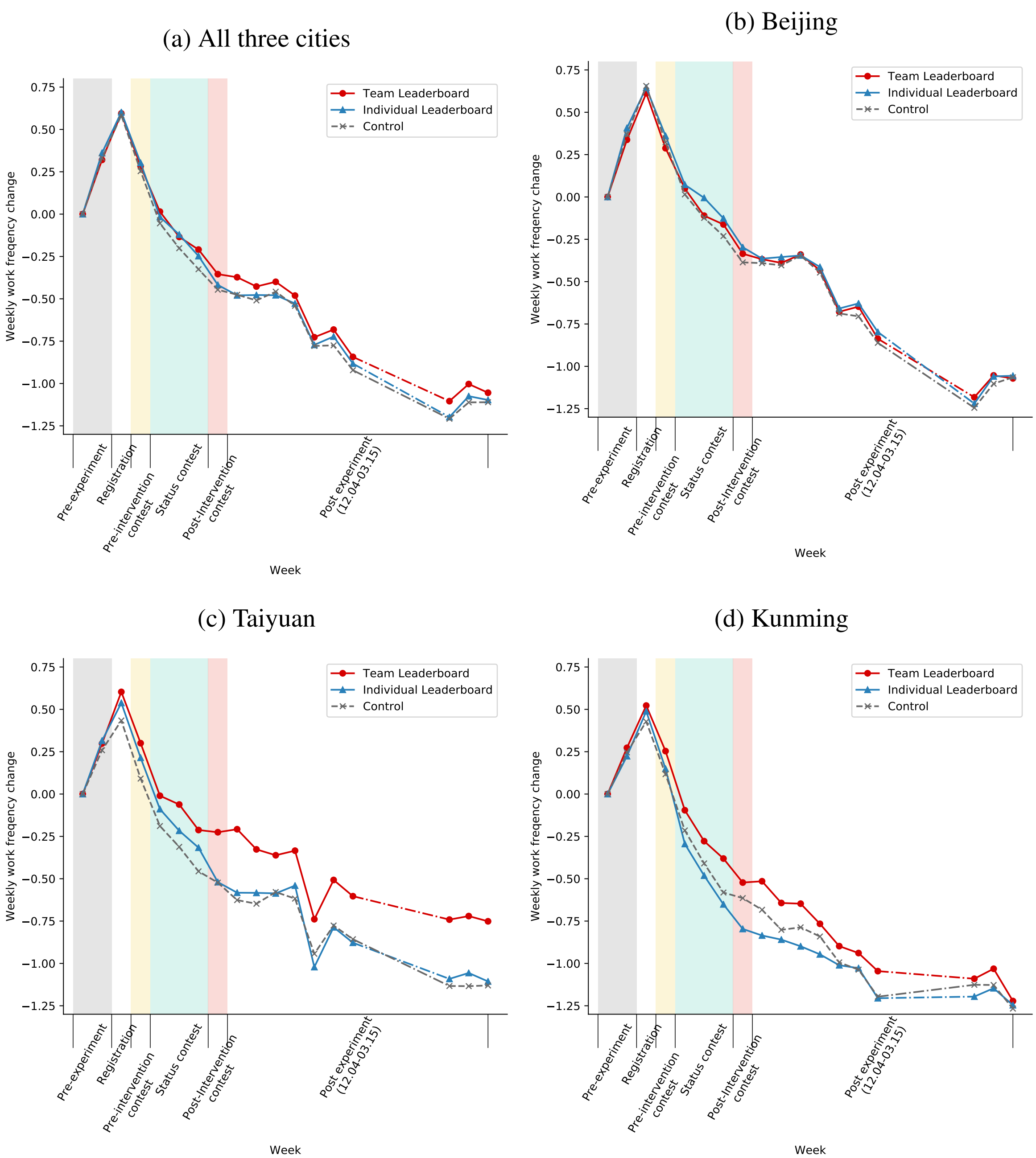

Figure 2: Average work frequency by week under each experimental condition. To better visualize the change over time, we re-scale the work frequency within each condition with reference to its pre-experiment work frequency in the week of October 8-14, i.e., two weeks before the start of the experiment. For example, each point represents the average weekly number of working days per driver of an experimental condition minus the pre-experiment weekly average working days per driver under the same experimental condition. The month of Spring Festival (a.k.a. Chinese New Year) is omitted as most migrant workers go back home during that month.

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

	Outcome variable: Δ of Weekly Revenue (CNY)							
	Treatment effects				Control individual heterogeneity			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	All	Beijing	Taiyuan	Kunming	All	Beijing	Taiyuan	Kunming
Treated	34.53**	41.67**	33.99	8.25	39.08**	45.82**	38.40	14.53
(In a virtual team)	(15.37)	(21.01)	(23.86)	(24.97)	(15.31)	(20.93)	(23.69)	(24.81)
		[0.17]	[0.18]	[0.33]		[0.09]	[0.12]	[0.23]
Age					6.98***	7.47***	1.90	8.39***
(Year)					(0.83)	(1.17)	(1.37)	(1.27)
DiDi age					32.16***	40.85***	3.64	3.43
(Year)					(7.47)	(9.59)	(11.53)	(13.39)
Hometown distance					-0.02	-0.01	-0.12**	-0.03
to contest city (km)					(0.02)	(0.02)	(0.05)	(0.02)
Self-formed team					-45.25***	-60.09***	-24.18	-4.10
					(16.09)	(21.59)	(27.49)	(26.90)
City fixed effect	Yes	-	-	-	Yes	-	-	-
# of clusters	11,890	8,100	1,625	2,165	11,890	8,100	1,625	2,165
# of drivers	27,790	18,900	3,815	5,075	27,790	18,900	3,815	5,075

Notes: Standard errors in parentheses are clustered at the team (individual) level for treated (control) drivers. False Discovery Rate adjusted q-values calculated separately for individual cities (2-4) & (6-8) are reported in square brackets.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

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

	Outcome variable: Δ of Weekly Revenue (CNY)							
	Treatment effects				Control individual heterogeneity			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	All	Beijing	Taiyuan	Kunming	All	Beijing	Taiyuan	Kunming
Team leaderboard	32.12*	27.03	58.49**	30.54	36.70**	32.40	62.31**	33.81
(β_1)	(17.97)	(24.61)	(26.60)	(29.91)	(17.90)	(24.50)	(26.57)	(29.69)
	[0.08]	[0.44]	[0.09]	[0.44]	[0.04]	[0.33]	[0.06]	[0.34]
Individual leaderboard	36.96**	56.32**	8.81	-14.50	41.47**	59.24**	13.68	-5.18
(β_2)	(17.90)	(24.49)	(28.76)	(28.03)	(17.82)	(24.37)	(28.43)	(27.86)
	[0.08]	[0.09]	[0.86]	[0.86]	[0.04]	[0.06]	[0.61]	[0.62]
Age					6.98***	7.47***	1.91	8.35***
(Year)					(0.83)	(1.17)	(1.37)	(1.28)
DiDi age					32.15***	40.77***	3.57	3.29
(Year)					(7.46)	(9.59)	(11.57)	(13.39)
Hometown distance					-0.02	-0.01	-0.12**	-0.03
to contest city (km)					(0.02)	(0.02)	(0.05)	(0.02)
Self-formed team					-45.22***	-59.76***	-23.62	-3.96
					(16.10)	(21.59)	(27.40)	(26.91)
City fixed effect	Yes	-	-	-	Yes	-	-	-
H_0: $\beta_1 = \beta_2$ (p-value)	0.79	0.25	0.08*	0.13	0.80	0.29	0.08*	0.18
# of clusters	11,890	8,100	1,625	2,165	11,890	8,100	1,625	2,165
# of drivers	27,790	18,900	3,815	5,075	27,790	18,900	3,815	5,075

Notes: Standard errors 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 (1) & (5) and for individual cities (2-4) & (6-8) and are reported in square brackets.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

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

	Outcome variable: Δ of Weekly Revenue (CNY)							
	Treatment effects				Control individual heterogeneity			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	All	Beijing	Taiyuan	Kunming	All	Beijing	Taiyuan	Kunming
Team leaderboard	49.91**	59.89*	58.03	6.05	55.75**	67.20**	59.78	11.14
(β_1)	(23.80)	(32.49)	(37.50)	(39.57)	(23.44)	(31.92)	(36.92)	(39.00)
	[0.08]	[0.32]	[0.32]	[0.56]	[0.04]	[0.27]	[0.27]	[0.39]
Individual leaderboard	11.75	38.98	-68.26*	-30.36	17.55	42.82	-65.75*	-18.30
(β_2)	(24.30)	(33.12)	(39.25)	(39.52)	(23.84)	(32.42)	(38.27)	(39.01)
	[0.46]	[0.32]	[0.32]	[0.36]	[0.30]	[0.27]	[0.27]	[0.39]
Age					10.56***	11.31***	4.72***	11.57***
(Year)					(1.07)	(1.50)	(1.70)	(1.68)
DiDi age					84.14***	97.94***	38.20**	38.55**
(Year)					(9.62)	(12.33)	(15.49)	(17.20)
Hometown distance					-0.03	-0.04	-0.16**	0.02
to contest city (km)					(0.02)	(0.03)	(0.06)	(0.03)
Self-formed team					-20.55	-39.15	23.93	28.60
					(21.57)	(28.73)	(38.61)	(37.24)
City fixed effect	Yes	-	-	-	Yes	-	-	-
H_0: $\beta_1 = \beta_2$ (p-value)	0.11	0.53	0.00***	0.33	0.11	0.45	0.00***	0.42
# of clusters	3,970	2,700	545	725	3,970	2,700	545	725
# of drivers	27,790	18,900	3,815	5,075	27,790	18,900	3,815	5,075

Notes: Standard errors in parentheses are clustered at the team level. False Discovery Rate adjusted q-values are calculated separately for all cities (1) & (5) and for individual cities (2-4) & (6-8) and are reported in square brackets.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

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

	Outcome variable: Δ of weekly # of work days							
	Treatment effects				Control individual heterogeneity			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	All	Beijing	Taiyuan	Kunming	All	Beijing	Taiyuan	Kunming
Team leaderboard	0.10**	0.06	0.33***	0.05	0.11**	0.08	0.33***	0.06
(β_1)	(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	-0.01	-0.01	-0.02	0.01	0.01	-0.00	-0.01	0.04
(β_2)	(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					0.03***	0.03***	0.02***	0.03***
(Year)					(0.00)	(0.00)	(0.01)	(0.00)
DiDi age					0.22***	0.24***	0.08	0.18***
(Year)					(0.02)	(0.02)	(0.05)	(0.05)
Hometown distance					-0.00***	-0.00***	-0.00**	-0.00
to contest city (km)					(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 post-					0.66***	0.68***	0.63***	0.61***
intervention contest					(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
# of drivers	27,790	18,900	3,815	5,075	27,790	18,900	3,815	5,075

Notes: False Discovery Rate adjusted q-values are calculated separately for all cities (1) & (5) and for individual cities (2-4) & (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 post-intervention contest.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Supplementary Materials for

Virtual Teams in a Gig Economy

Teng Ye, Wei Ai, Yan Chen, Qiaozhu Mei, Jieping Ye

This PDF file includes:

1 Experiment Design Details

1.1 City characteristics

Table S1: City Characteristics

City	Location	# of historical contests	Order fulfillment rate	Population (million)	# of drivers registered the experiment	# of participants
Beijing	Northern	17	0.90	21.54	21,126	18,900
Taiyuan	Central	14	0.90	4.42	4,648	3,815
Kunming	Southwest	5	0.98	6.85	5,776	5,075

Of the three cities where we implemented our experiment, Beijing is the capital of China, located in northern China, with over 21.54 million residents. Taiyuan is the capital of Shanxi province, located in central China, with a population of 4.42 million. Kunming is the capital of Yunnan province, located in southwest China, with a population of 6.85 million.[3]

1.2 Experimental process

Figure S1 presents the experimental process.

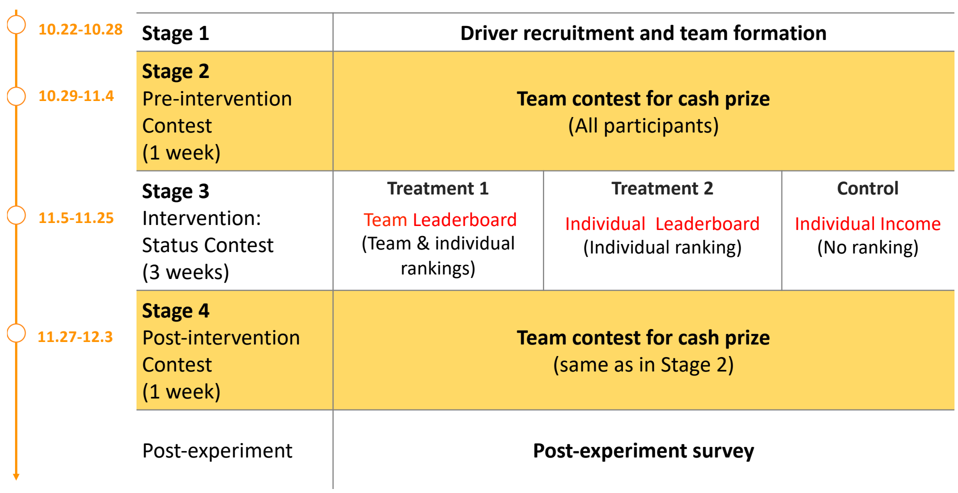

Figure S1: Experimental process

1.3 APP interface mock-ups for each experimental condition

Figure S2 presents the interface mock-ups for different experimental conditions during the intervention stage.

[3]Population data at the end of the year of 2018 are retrieved from:
the National Bureau of Statistics of China (http://www.stats.gov.cn/tjsj/ndsj/2019/indexch.htm),
the Province Bureau of Statistics of Yunnan (http://stats.yn.gov.cn/tjsj/tjnj/201912/t20191202_908222.html),
and the City Bureau of Statistics of Taiyuan (http://stats.taiyuan.gov.cn/doc/2019/05/14/845586.shtml).

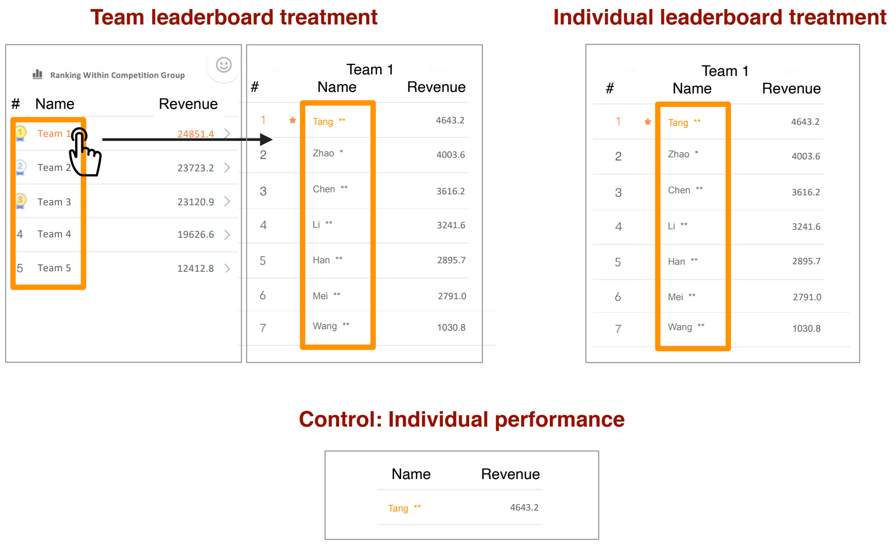

Figure S2: APP interface mock-ups for drivers in the team leaderboard (top left), individual leaderboard (top right), and control (bottom panel) conditions

1.4 Messages sent in each stage

Stage 1: Driver recruitment and team formation

Our collaborators at DiDi send out both text messages and in-app push notifications[4] to inform all drivers in the experimental cities about the team contest.

The English translation of the call for participation to the drivers reads as follows:

> The DiDi driver team contest is about to start soon! Say goodbye to the lonely driving work on your own. Get to know new driver friends and compete for rewards with your teammates! Click here to register for the contest. Please keep up your good service and drive safely.

[4]Text message refers to the normal message sent out by DiDi. In-app push notification refers to the message popping up within the DiDi app.

Stage 2: Pre-intervention contest

We send the following reminder by text message and in-app push notification to every driver every evening during the one-week period of the pre-intervention contest.

> The driver team contest has become more intense! Want to know your team's ranking? Want to check your teammates' performance? Want to know your competitors' performance? Click this link and you can access all the above information. Please keep up your good service and drive safely.

Stage 3: Status contest (the intervention)

The notifications and reminders for the three experiment conditions during the three-week status contest include the following:

1. **Team leaderboard treatment**

   At the beginning of this stage, we send the following text message to notify drivers in the team leaderboard treatment that:

   > The team contest is over. The ranking information will continue to be updated during November. Please pay attention to the performance of your team and your teammates. DiDi is amazing because of you!

   The following reminder is sent by text message and in-app push notification once a day during the evening:

   > The latest performance just came out! Want to know your team's and teammates' performance? Click this link and you can access all the information. Please keep up your good service and drive safely.

2. **Individual leaderboard treatment**

   At the beginning of this stage, we send the following text message to notify drivers in the individual leaderboard treatment that:

   > The team contest is over. The ranking information will continue to be updated during November. Please pay attention to the performance of your teammates. DiDi is amazing because of you!

   The following individual performance reminders are sent every evening by both text message and in-app push notification:

   > The latest performance just came out! Want to know your teammates' performance? Click this link and you can access all the information. Please keep up your good service and drive safely.

3. **Control condition**

   At the beginning of this stage, we send the following text message to drivers in the control group that:

   > The team contest is over. Please pay attention to your performance. DiDi is amazing because of you!

   An individual performance update reminder is sent every evening by a text message and an in-app push notification as follows:

   > The latest performance just came out! Want to know the your performance? Click this link and you can go to the your revenue page. Please keep up your good service and drive safely.

Stage 4: Post-intervention contest

The following text message announcement is sent to all drivers in all three conditions on the day before the post-intervention contest to notify them of the contest:

> Here comes the driver team contest again (from November 27 to December 3, 2018)! You don't need to form a team again. Team members and competitor teams will remain the same as in the last contest. Please contact your team members and get ready to compete for the cash prize!

1.5 Prize determination across cities

To make the experiment in each city most comparable, we determine the bonus volume for the winner team by keeping the rate of the bonus above the city-specific drivers' hourly earnings. Specifically, we first calculated the average hourly pay using the 30-day data from DiDi prior to the experiment. We carefully excluded the national holiday period (2018/10/01 - 2018/10/07) from our calculations to obtain a better indication of the average hourly earnings. As a result, we calculated the average hourly pay based on data from 2018/09/10 - 2018/09/29 and 2018/10/08 - 2018/10/17. The details of the financial reward for each city are reported in Table S2.

Table S2: Details of prize in each city (money in CNY)

City	Calculated team prize	Rounded team prize	Team leader extra prize
Beijing	1,000	1,000	10
Taiyuan	654.21	650	10
Kunming	663.02	650	10

2 Power analysis

We use a subset of the experimental data from our 2017 field experiment conducted among DiDi drivers in the city of Dongguan to generate an estimated effect size and variance parameters for our power analysis and sample size calculation. For our experiment, we would like to have a sample size large enough to obtain 90% power.

In the 2017 experiment, drivers are randomized into treatment and control conditions. Among the treated drivers, we deem teams for which the captain submits a pre-contest questionnaire as responsive and those who do not submit a questionnaire as unresponsive. In our power analysis, we use the responsive teams as our treatment condition and the unresponsive teams as our control condition since the 2017 placebo control drivers are not formed into teams. We use the five contest days as five periods. With this setting, we run the following fixed effects panel regression:

Table S3: Panel analysis with 2017 experiment data by fixed-effects (within-subject) regression

	Δ of Daily Orders
Game day	-1.35**
	(0.29)
Responsive	2.81 **
	(0.37)

# of observations = 17,500; # of groups = 250;
σ_u = 4.01; σ_e = 12.10; ρ = 0.10;
Standard errors in parentheses.
$^{*}\ p < 0.1$, $^{**}\ p < 0.05$, $^{***}\ p < 0.01$

According to the results in Table S3, we use the PowerBBK package (*49*) to compute the power of the new design, assuming similar behavioral responses as in the 2017 experiments.

The parameters are determined based on the following considerations (see Table S3 for statistics):

- budget = 125 teams per condition $\times$ 2 experimental conditions $\times$ 5 contest periods =

1250.

- beta = $(15.24, 2.8)$ since (1) $15.24 = 16.582 - 1.347$ is the daily number of trips of the unresponsive teams during the contest, (2) whereas 2.8 is the treatment effect of responsiveness.

- muvar = $\sigma_u^2 = 16$.

- espva = $\sigma_e^2 = 144$.

- panel allocation = 0.4 since 40% of the teams were unresponsive.

This command yields a power of 0.896. As we have three experimental conditions in our main analyses, we need 375 teams.

Increasing the budget by 1.5 (from 250 to 375 teams in two conditions) would give us a power of 0.982. In this case, having 564 teams (4,000 drivers) would be sufficient for our analysis. The caveat is that we do not know the potential treatment effect in the leader board phase, and therefore, cannot account for this effect in our power calculation.

3 Randomization check

We conduct randomization check for pre-experiment revenue, age, gender, DiDi age, team formation methods, and hometown distance to the contest city using pairwise Kolmogorov-Smirnov tests across experimental conditions for each city, and find that none of the pairwise tests is significant at the conventional level ($p > 0.10$). Results are reported in Table S4.

Table S4: Summary statistics and randomization check

	Beijing			Taiyuan			Kunming		
	Team	Individual	Control	Team	Individual	Control	Team	Individual	Control
Daily revenue before	381.97	381.71	381.70	171.64	180.54	176.09	212.71	214.21	218.03
experiment (CNY)	(215.35)	(216.15)	(213.83)	(126.25)	(129.99)	(125.88)	(144.30)	(143.99)	(144.56)
Age (Year)	36.82	36.91	37.35	36.53	36.63	36.86	36.49	35.91	37.02
	(8.12)	(8.09)	(8.28)	(8.26)	(8.22)	(8.34)	(8.50)	(8.58)	(8.81)
Male	0.97	0.97	0.97	0.97	0.95	0.96	0.93	0.92	0.93
	(0.17)	(0.16)	(0.17)	(0.18)	(0.22)	(0.19)	(0.26)	(0.26)	(0.26)
DiDi age (month)	24.69	24.97	24.86	24.08	23.88	24.55	15.06	14.70	14.51
	(13.19)	(13.11)	(13.01)	(11.13)	(11.62)	(11.04)	(11.04)	(11.01)	(10.98)
Self-formed teams	0.38	0.37	0.37	0.36	0.38	0.26	0.31	0.31	0.30
	(0.49)	(0.48)	(0.48)	(0.48)	(0.49)	(0.44)	(0.46)	(0.46)	(0.47)
Hometown distance	451.93	465.04	463.66	114.50	121.68	109.06	249.72	293.28	289.13
to the contest city (km)	(281.53)	(283.85)	(285.47)	(117.96)	(135.84)	(100.40)	(219.29)	(326.88)	(343.39)
# of leaderboards	180	180	180	37	36	36	49	48	48
# of teams	900	900	900	185	180	180	245	240	240
# of drivers	6,300	6,300	6,300	1,295	1,260	1,260	1,715	1,680	1,680

Notes: Standard deviations are in parentheses.

4 Robustness Checks: Treatment effects on driver revenue after excluding team captains

Table S5: Average and heterogeneous treatment effects on weekly revenue during the intervention (status contest) after excluding team captains: Difference-in-differences panel regressions investigating the two treatments separately.

	Outcome variable: Δ of Weekly Revenue (CNY)							
	Treatment effects				Control individual heterogeneity			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	All	Beijing	Taiyuan	Kunming	All	Beijing	Taiyuan	Kunming
Team leaderboard	35.90*	26.81	69.93**	43.43	41.13**	32.82	73.10**	47.72
(β_1)	(19.30)	(26.54)	(28.34)	(30.87)	(19.24)	(26.47)	(28.36)	(30.58)
	[0.03]	[0.46]	[0.04]	[0.27]	[0.02]	[0.27]	[0.03]	[0.19]
Individual leaderboard	48.39**	65.11**	25.95	2.55	52.92***	68.21***	30.54	11.61
(β_2)	(19.12)	(26.14)	(31.12)	(29.97)	(19.07)	(26.06)	(30.96)	(29.89)
	[0.02]	[0.04]	[0.47]	[0.87]	[0.01]	[0.03]	[0.31]	[0.48]
Age					6.54***	7.13***	1.31	7.75***
(Year)					(0.90)	(1.26)	(1.54)	(1.37)
DiDi age					30.20***	38.62***	7.69	-1.25
(Year)					(8.05)	(10.32)	(13.22)	(14.06)
Hometown distance					-0.01	-0.00	-0.13**	-0.02
to contest city (km)					(0.02)	(0.03)	(0.05)	(0.03)
Self-formed team					-43.83**	-54.71**	-19.99	-21.97
					(17.19)	(23.11)	(29.20)	(28.05)
City fixed effect	Yes	-	-	-	Yes	-	-	-
H_0: $\beta_1 = \beta_2$ (p-value)	0.52	0.16	0.14	0.18	0.55	0.19	0.15	0.23
# of clusters	10,570	7,200	1,445	1,925	10,570	7,200	1,445	1,925
# of drivers	23,820	16,200	3,270	4,350	23,820	16,200	3,270	4,350

Notes: Standard errors in parentheses are clustered at the team (individual) level for ranking (control) conditions. False Discovery Rate adjusted q-values are calculated separately for all cities (1) & (5) and for individual cities (2-4) & (6-8) and are reported in square brackets.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table S6: Average and heterogeneous treatment effects on weekly revenue in the post-intervention contest after excluding team captains: Difference-in-differences panel regressions.

	Outcome variable: Δ of Weekly Revenue (CNY)							
	Treatment effects				Control individual heterogeneity			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	All	Beijing	Taiyuan	Kunming	All	Beijing	Taiyuan	Kunming
Team leaderboard	56.43**	61.42*	72.75*	24.86	64.07***	70.28**	74.40**	32.40
(β_1)	(24.94)	(34.15)	(38.10)	(40.85)	(24.58)	(33.61)	(37.54)	(40.17)
	[0.05]	[0.28]	[0.28]	[0.35]	[0.02]	[0.17]	[0.17]	[0.34]
Individual leaderboard	21.20	47.30	-55.35	-19.28	27.71	51.81	-52.66	-7.36
(β_2)	(25.36)	(34.63)	(39.99)	(40.84)	(24.95)	(34.02)	(39.08)	(40.50)
	[0.25]	[0.28]	[0.28]	[0.35]	[0.15]	[0.21]	[0.22]	[0.40]
Age					10.60***	11.55***	4.08**	11.27***
(Year)					(1.14)	(1.60)	(1.77)	(1.79)
DiDi age					81.98***	95.10***	29.31*	45.30**
(Year)					(10.34)	(13.22)	(16.78)	(18.50)
Hometown distance					-0.02	-0.03	-0.15**	0.04
to contest city (km)					(0.02)	(0.03)	(0.07)	(0.03)
Self-formed team					-28.48	-45.97	18.55	9.83
					(22.57)	(30.18)	(39.59)	(38.17)
City fixed effect	Yes	-	-	-	Yes	-	-	-
H_0: $\beta_1 = \beta_2$ (p-value)	0.16	0.68	0.00***	0.26	0.14	0.59	0.00***	0.30
# of clusters	3,970	2,700	545	725	3,970	2,700	545	725
# of drivers	23,820	16,200	3,270	4,350	23,820	16,200	3,270	4,350

Notes: Standard errors in parentheses are clustered at the team level. False Discovery Rate adjusted q-values are calculated separately for all cities (1) & (5) and for individual cities (2-4) & (6-8) and are reported in square brackets.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

5 Robustness Checks: Treatment effects on driver retention after excluding team captains or using different time windows

Table S7: Average and heterogeneous treatment effects on weekly number of working days during the week after the experiment ended (December 5-11, 2018): Difference-in-differences panel regressions.

	Outcome variable: Δ of weekly # of work days							
	Treatment effects				Control individual heterogeneity			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	All	Beijing	Taiyuan	Kunming	All	Beijing	Taiyuan	Kunming
Team leaderboard	0.11***	0.05	0.39***	0.14	0.12***	0.06	0.41***	0.15
(β_1)	(0.04)	(0.05)	(0.11)	(0.09)	(0.04)	(0.05)	(0.11)	(0.09)
	[0.01]	[0.61]	[0.002]	[0.46]	[0.004]	[0.43]	[0.001]	[0.34]
Individual leaderboard	-0.03	-0.01	-0.01	-0.12	-0.02	-0.00	0.02	-0.09
(β_2)	(0.04)	(0.05)	(0.11)	(0.09)	(0.04)	(0.05)	(0.11)	(0.09)
	[0.30]	[0.90]	[0.90]	[0.55]	[0.48]	[0.86]	[0.86]	[0.51]
Age					0.02***	0.02***	0.02***	0.03***
(Year)					(0.00)	(0.00)	(0.01)	(0.00)
DiDi age					0.14***	0.15***	0.01	0.15***
(Year)					(0.02)	(0.02)	(0.05)	(0.04)
Hometown distance					-0.00	-0.00	-0.00	0.00
to contest city (km)					(0.00)	(0.00)	(0.00)	(0.00)
Self-formed team					-0.08**	-0.13***	-0.16*	0.19**
					(0.03)	(0.04)	(0.09)	(0.08)
Team won in post-					0.86***	0.91***	0.77***	0.71***
intervention contest					(0.04)	(0.05)	(0.11)	(0.10)
City fixed effect	Yes	-	-	-	Yes	-	-	-
H_0: $\beta_1 = \beta_2$ (p-value)	0.00***	0.24	0.00***	0.01***	0.00***	0.18	0.00***	0.01**
# of drivers	27,790	18,900	3,815	5,075	27,790	18,900	3,815	5,075

Notes: False Discovery Rate adjusted q-values are calculated separately for all cities (1) & (5) and for individual cities (2-4) & (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 winning team in the post-intervention contest.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table S8: Average and heterogeneous treatment effects on weekly number of working days during the week after the contest (December 5-11, 2018) after excluding team captains: Difference-in-differences panel regressions.

	Outcome variable: Δ of weekly # of work days							
	Treatment effects				Control individual heterogeneity			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	All	Beijing	Taiyuan	Kunming	All	Beijing	Taiyuan	Kunming
Team leaderboard	0.13^{***}	0.04	0.47^{***}	0.22^{**}	0.15^{***}	0.05	0.49^{***}	0.23^{**}
(β_1)	(0.04)	(0.05)	(0.11)	(0.10)	(0.04)	(0.05)	(0.11)	(0.10)
	[0.01]	[0.94]	[0.001]	[0.09]	[0.002]	[0.59]	[0.001]	[0.06]
Individual leaderboard	-0.00	-0.01	0.09	-0.05	0.01	0.00	0.10	-0.02
(β_2)	(0.04)	(0.05)	(0.11)	(0.10)	(0.04)	(0.05)	(0.11)	(0.10)
	[0.95]	[1.00]	[0.94]	[1.00]	[0.66]	[1.00]	[0.59]	[1.00]
Age					0.02^{***}	0.02^{***}	0.02^{***}	0.03^{***}
(Year)					(0.00)	(0.00)	(0.01)	(0.00)
DiDi age					0.13^{***}	0.15^{***}	-0.02	0.15^{***}
(Year)					(0.02)	(0.02)	(0.05)	(0.05)
Hometown distance					-0.00	-0.00	-0.00^{*}	0.00
to contest city (km)					(0.00)	(0.00)	(0.00)	(0.00)
Self-formed team					-0.07^{*}	-0.13^{***}	-0.09	0.16^{*}
					(0.04)	(0.04)	(0.10)	(0.09)
Team won in post-					0.87^{***}	0.95^{***}	0.74^{***}	0.71^{***}
intervention contest					(0.04)	(0.05)	(0.12)	(0.10)
City fixed effect	Yes	-	-	-	Yes	-	-	-
H_0: $\beta_1 = \beta_2$ (p-value)	0.00^{***}	0.41	0.00^{***}	0.01^{***}	0.00^{***}	0.32	0.00^{***}	0.01^{**}
# of drivers	23,820	16,200	3,270	4,350	23,820	16,200	3,270	4,350

Notes: False Discovery Rate adjusted q-values are calculated separately for all cities (1) & (5) and for individual cities (2-4) & (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 winning team in the post-intervention contest.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table S9: Average and heterogeneous treatment effects on weekly number of working days during the week of January 12-18, 2019, about one month after the experiment ended: Difference-in-differences panel regressions.

	Outcome variable: Δ of weekly # of work days							
	Treatment effects				Control individual heterogeneity			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	All	Beijing	Taiyuan	Kunming	All	Beijing	Taiyuan	Kunming
Team leaderboard	0.11^{**}	0.08	0.24^{**}	0.11	0.12^{***}	0.10^{*}	0.26^{**}	0.12
(β_1)	(0.04)	(0.05)	(0.11)	(0.10)	(0.04)	(0.05)	(0.11)	(0.10)
	[0.02]	[0.43]	[0.18]	[0.56]	[0.01]	[0.18]	[0.11]	[0.39]
Individual leaderboard	0.03	0.05	-0.08	0.03	0.04	0.06	-0.05	0.06
(β_2)	(0.04)	(0.05)	(0.11)	(0.10)	(0.04)	(0.05)	(0.11)	(0.10)
	[0.36]	[0.57]	[0.63]	[0.98]	[0.20]	[0.39]	[0.71]	[0.71]
Age					0.03^{***}	0.03^{***}	0.01^{***}	0.03^{***}
(Year)					(0.00)	(0.00)	(0.01)	(0.00)
DiDi age					0.19^{***}	0.22^{***}	0.03	0.18^{***}
(Year)					(0.02)	(0.02)	(0.05)	(0.04)
Hometown distance					-0.00^{***}	-0.00^{**}	-0.00^{**}	-0.00
to contest city (km)					(0.00)	(0.00)	(0.00)	(0.00)
Self-formed team					-0.05	-0.10^{**}	-0.09	0.18^{**}
					(0.04)	(0.05)	(0.10)	(0.09)
Team won in post-					0.69^{***}	0.73^{***}	0.58^{***}	0.60^{***}
intervention contest					(0.04)	(0.05)	(0.11)	(0.10)
City fixed effect	Yes	-	-	-	Yes	-	-	-
H_0: $\beta_1 = \beta_2$ (p-value)	0.05^{*}	0.52	0.00^{***}	0.37	0.05^{*}	0.43	0.00^{***}	0.53
# of drivers	27,790	18,900	3,815	5,075	27,790	18,900	3,815	5,075

Notes: False Discovery Rate adjusted q-values are calculated separately for all cities (1) & (5) and for individual cities (2-4) & (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 winning team in the post-intervention contest.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table S10: Average and heterogeneous treatment effects on weekly number of working days during the week of January 12-18, 2019, about one month after the experiment ended, after excluding team captains: Difference-in-differences panel regressions.

	Outcome variable: Δ of weekly # of work days							
	Treatment effects				Control individual heterogeneity			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	All	Beijing	Taiyuan	Kunming	All	Beijing	Taiyuan	Kunming
Team leaderboard	0.12**	0.07	0.31***	0.14	0.14***	0.09*	0.32***	0.16
(β_1)	(0.05)	(0.06)	(0.12)	(0.11)	(0.05)	(0.06)	(0.12)	(0.11)
	[0.02]	[0.49]	[0.05]	[0.49]	[0.01]	[0.30]	[0.05]	[0.30]
Individual leaderboard	0.04	0.06	-0.01	0.02	0.06	0.07	0.00	0.05
(β_2)	(0.05)	(0.06)	(0.12)	(0.11)	(0.05)	(0.06)	(0.12)	(0.11)
	[0.24]	[0.65]	[0.88]	[0.88]	[0.12]	[0.38]	[0.53]	[0.53]
Age					0.03***	0.03***	0.01**	0.04***
(Year)					(0.00)	(0.00)	(0.01)	(0.01)
DiDi age					0.19***	0.22***	0.02	0.16***
(Year)					(0.02)	(0.02)	(0.05)	(0.05)
Hometown distance					-0.00***	-0.00**	-0.00**	-0.00
to contest city (km)					(0.00)	(0.00)	(0.00)	(0.00)
Self-formed team					-0.05	-0.12**	-0.00	0.19**
					(0.04)	(0.05)	(0.11)	(0.09)
Team won in					0.69***	0.76***	0.54***	0.58***
surprise short contest					(0.05)	(0.06)	(0.12)	(0.11)
City fixed effect	yes	-	-	-	yes	-	-	-
H_0: $\beta_1 = \beta_2$ (p-value)	0.09*	0.77	0.01***	0.24	0.09*	0.65	0.01***	0.32
# of drivers	23,820	16,200	3,270	4,350	23,820	16,200	3,270	4,350

Notes: False Discovery Rate adjusted q-values are calculated separately for all cities (1) & (5) and for individual cities (2-4) & (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 winning team in the post-intervention contest.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table S11: Average and heterogeneous treatment effects on weekly number of working days during the second week of March (March 4-10, 2019), about three months after the experiment ended, after excluding team captains: Difference-in-differences panel regressions.

	Outcome variable: Δ of weekly # of work days							
	Treatment effects				Control individual heterogeneity			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	All	Beijing	Taiyuan	Kunming	All	Beijing	Taiyuan	Kunming
Team leaderboard	0.09*	0.04	0.38***	0.06	0.11**	0.06	0.37***	0.07
(β_1)	(0.05)	(0.06)	(0.12)	(0.11)	(0.05)	(0.06)	(0.12)	(0.11)
	[0.15]	[1.00]	[0.01]	[1.00]	[0.06]	[1.00]	[0.02]	[1.00]
Individual leaderboard	-0.03	-0.05	0.04	0.00	-0.01	-0.03	0.03	0.03
(β_2)	(0.05)	(0.06)	(0.13)	(0.11)	(0.05)	(0.06)	(0.13)	(0.11)
	[0.42]	[1.00]	[1.00]	[1.00]	[0.78]	[1.00]	[1.00]	[1.00]
Age					0.03***	0.03***	0.02***	0.03***
(Year)					(0.00)	(0.00)	(0.01)	(0.01)
DiDi age					0.21***	0.24***	0.08	0.16***
(Year)					(0.02)	(0.02)	(0.06)	(0.05)
Hometown distance					-0.00***	-0.00***	-0.00**	-0.00
to contest city (km)					(0.00)	(0.00)	(0.00)	(0.00)
Self-formed team					-0.07	-0.17***	0.19*	0.15
					(0.04)	(0.05)	(0.11)	(0.10)
Team won in post-					0.64***	0.68***	0.56***	0.54***
intervention contest					(0.05)	(0.06)	(0.13)	(0.11)
City fixed effect	Yes	-	-	-	Yes	-	-	-
H_0: $\beta_1 = \beta_2$ (p-value)	0.02**	0.12	0.01***	0.61	0.02**	0.11	0.01***	0.74
# of drivers	23,820	16,200	3,270	4,350	23,820	16,200	3,270	4,350

Notes: False Discovery Rate adjusted q-values are calculated separately for all cities (1) & (5) and for individual cities (2-4) & (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 winning team in the post-intervention contest.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

6 Preference for being a captain

To better understand driver incentives, we conduct an additional analysis on driver preferences to be a team captain. We use a Logistic Regression model (eq. S1) to understand how past experience on DiDi affects a driver's choice to be a team captain (H4), where V refers to the indicator function which equals 1 if a driver volunteers to be a team captain, and *Pre-Experiment Productivity* is operationalized as driver revenue in the two weeks before our experiment. *Served as Captain Before* is a binary variable that shows whether the driver had been a captain before he participated in the current team contest. We include γ_c to control for city-specific characteristics.

$$Pr(V=1) = \Phi(\beta_0 + \beta_1 \text{Pre-Experiment Productivity} + \beta_2 \text{Served as Captain before} + \beta_3 \text{Didi Age} + \gamma_c) \tag{S1}$$

The results (Table S12) show that drivers with higher performance prior to the experiment and who have served as captains before are significantly more likely to volunteer to be a captain overall and at the city level. However, the effects of DiDi age are more complicated. DiDi age is positively correlated with captain preference overall and in Beijing (with $\beta = 0.01$, $p < .01$ and $\beta = 0.02$, $p < .01$, respectively), while it is negatively related to captain preference in Taiyuan (with $\beta = -0.01$, $p < .05$) and has no significant relationship with captain preference in Kunming.

Table S12: Results of preference for being team captains: Logistic regression with all participants.

	Outcome: Whether drivers volunteer to be captains			
	(1)	(2)	(3)	(4)
	All	Beijing	Taiyuan	Kunming
Pre Experiment Revenue	0.04***	0.04***	0.07**	0.08***
(in 10,000 RMB)	(0.01)	(0.01)	(0.03)	(0.02)
	[0.00]	[0.00]	[0.01]	[0.00]
Served as captain before	0.22***	0.22***	0.23***	0.22***
(Binary indicator)	(0.00)	(0.00)	(0.02)	(0.01)
	[0.00]	[0.00]	[0.00]	[0.00]
DiDi age	0.01***	0.02***	-0.01**	0.00
(Year)	(0.00)	(0.00)	(0.01)	(0.01)
	[0.00]	[0.00]	[0.01]	[0.10]
City fixed effect	Yes	-	-	-
# of drivers	27,790	18,900	3,815	5,075

Notes: Average marginal effect with *delta-method* SE in parentheses.False Discovery Rate adjusted q-values are calculated separately for all cities (1) and for individual cities (2-4) and are reported in square brackets.

$^{*}p < 0.1, ^{**}p < 0.05, ^{***}p < 0.01$

7 Who benefits more from team contests? Below- versus above-median drivers

In this analysis, we examine who benefits more from the team identity and social information by testing the heterogeneous treatment effects on drivers with different levels of pre-experiment revenue. We differentiate drivers by whether their pre-experiment revenue is below the city median. As shown in Figure S3, drivers whose revenue falls in the lower half in their city consistently exhibit a greater revenue increase than their counterparts, in the status contest across all cities.

Specifically during the longer-term contest, pooling drivers in all cities (table S14 (1)), we find that drivers whose pre-experiment revenue is below the city median generate 782.07 Yuan more than drivers whose pre-experiment revenue is above the city median ($p < .01$), accounting for about 37.53% of the average weekly revenue. This pattern is consistent in each of the three cities, with a revenue increase of 943.36 Yuan in Beijing (38.32% of Beijing average weekly revenue, $p < .01$), 401.99 Yuan in Taiyuan (36.08% of Taiyuan average weekly revenue, $p < .01$), and 462.37 Yuan in Kunming (33.19% of Kunming average weekly revenue, $p < .01$). No interaction effect is identified across cities and treatments. Additional tests show that drivers with below-median revenue in the team (H_0: $\beta_3 + \beta_4 = 0$) and individual (H_0: $\beta_3 + \beta_5 = 0$) leaderboard conditions exhibit a greater revenue increase during the status competition overall and in each of the three cities.

From Table S15, we see that drivers with below-median revenue also benefit more in the rewarded post-intervention contest: they generate a higher revenue of 837.31 Yuan ($p < .01$) than the above-median drivers overall, which accounts for 41.80% of the average weekly revenue of all drivers in the control groups in the three cities. Among these drivers, below-median drivers in Beijing exhibit a higher increase of 1013.26 Yuan (43.08% of Beijing average weekly revenue, $p < .01$), while drivers in Taiyuan and Kunming generate 386.66 Yuan (34.50% of Taiyuan average weekly revenue, $p < .01$) and 517.18 Yuan (38.15% of Kunming average weekly revenue, $p < .01$), respectively, compared to the above-median drivers. Results of additional tests (H_0: $\beta_3 + \beta_4 = 0$ and H_0: $\beta_3 + \beta_5 = 0$) confirm that the below-median drivers in both the team and individual leaderboard conditions exhibit a greater revenue increase during the post-intervention contest period overall and in each of the three cities.

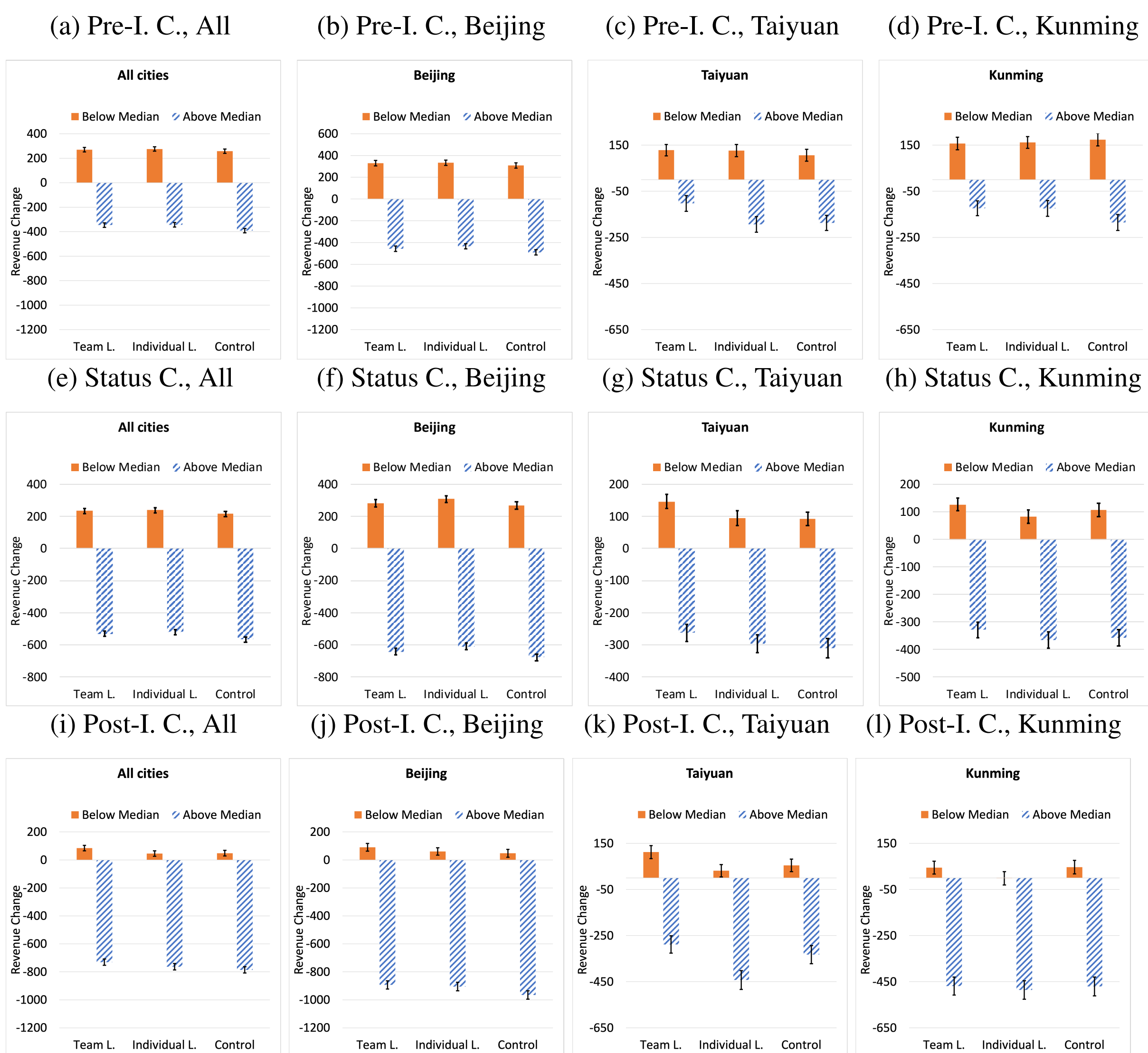

Figure S3: The effect of team and individual leaderboards for drivers with below and above median pre-contest revenue with standard error as error bars. (Pre-I. C.: Pre-intervention contest; Status C.: Status contest; Post-I. C.: Post-intervention contest.)

Table S13: Below- versus above-median drivers: Difference-in-differences regressions during the pre-intervention contest.

	Outcome: Δ of Weekly Revenue (CNY)			
	(1)	(2)	(3)	(4)
	All	Beijing	Taiyuan	Kunming
Below median	628.50***	784.24***	281.89***	308.65***
	(15.95)	(20.65)	(26.39)	(26.39)
City fixed effect	Yes	-	-	-
# of drivers	27,790	18,900	3,815	5,075

Notes: Standard errors in parentheses are clustered at the team level.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table S14: Below- versus above-median drivers: Difference-in-differences regressions during the intervention.

	Outcome: Δ of Weekly Revenue (CNY)			
	(1)	(2)	(3)	(4)
	All	Beijing	Taiyuan	Kunming
Team leaderboard	35.43	34.86	47.50	28.68
(β_1)	(23.86)	(31.53)	(40.82)	(43.01)
	[0.11]	[0.81]	[0.81]	[1.00]
Individual leaderboard	46.50**	67.87**	14.69	-8.29
(β_2)	(23.59)	(31.28)	(42.13)	(42.99)
	[0.11]	[0.22]	[1.00]	[1.00]
Below median	782.07***	943.36***	401.99***	462.37***
(β_3)	(23.15)	(31.32)	(36.21)	(38.36)
Team leaderboard * Below median	-17.03	-20.60	6.94	-7.68
(β_4)	(34.34)	(46.24)	(51.32)	(53.54)
Individual leaderboard * Below median	-22.48	-27.47	-13.00	-14.47
(β_5)	(34.15)	(45.60)	(52.96)	(54.99)
City fixed effect	Yes	-	-	-
H_0: $\beta_1 = \beta_2$ (p-value)	0.65	0.31	0.43	0.40
H_0: $\beta_3 + \beta_4 = 0$ (p-value)	0.00***	0.00***	0.00***	0.00***
H_0: $\beta_3 + \beta_5 = 0$ (p-value)	0.00***	0.00***	0.00***	0.00***
H_0: $\beta_1 + \beta_4 = 0$ (p-value)	0.46	0.68	0.09*	0.56
H_0: $\beta_2 + \beta_5 = 0$ (p-value)	0.33	0.23	0.96	0.50
H_0: $\beta_1 + \beta_4 = \beta_2 + \beta_5$ (p-value)	0.83	0.46	0.13	0.21
# of drivers	27,790	18,900	3,815	5,075

Notes: Standard errors in parentheses are clustered at the team (individual) level for the leaderboard (control) conditions. False Discovery Rate adjusted q-values are calculated separately for all cities (1) & (5) and for individual cities (2-4) & (6-8) and are reported in square brackets.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table S15: Below- versus above-median drivers: Difference-in-differences regressions during the post-intervention contest.

	Outcome: Δ of Weekly Revenue (CNY)			
	(1)	(2)	(3)	(4)
	All	Beijing	Taiyuan	Kunming
Team leaderboard	55.11	72.12	43.92	1.54
(β_1)	(34.27)	(45.54)	(59.38)	(59.27)
	[0.28]	[0.52]	[0.61]	[0.96]
Individual leaderboard	22.57	60.18	-109.98*	-15.16
(β_2)	(34.60)	(45.88)	(62.99)	(61.09)
	[0.35]	[0.52]	[0.52]	[0.93]
Below median	837.31***	1013.26***	386.66***	517.18***
(β_3)	(32.26)	(42.87)	(52.94)	(49.32)
Team leaderboard * Below median	-21.54	-29.80	13.82	-3.69
(β_4)	(45.47)	(60.95)	(71.06)	(68.99)
Individual leaderboard * Below median	-25.31	-47.23	86.81	-33.15
(β_5)	(45.43)	(60.55)	(74.45)	(70.54)
City fixed effect	Yes	-	-	-
H_0: $\beta_1 = \beta_2$ (p-value)	0.35	0.80	0.01**	0.77
H_0: $\beta_3 + \beta_4 = 0$ (p-value)	0.00***	0.00***	0.00***	0.00***
H_0: $\beta_3 + \beta_5 = 0$ (p-value)	0.00***	0.00***	0.00***	0.00***
H_0: $\beta_1 + \beta_4 = 0$ (p-value)	0.27	0.31	0.17	0.96
H_0: $\beta_2 + \beta_5 = 0$ (p-value)	0.93	0.76	0.56	0.25
H_0: $\beta_1 + \beta_4 = \beta_2 + \beta_5$ (p-value)	0.23	0.48	0.05**	0.28
# of drivers	27,790	18,900	3,815	5,075

Notes: Standard errors in parentheses are clustered at the team level. False Discovery Rate adjusted q-values are calculated separately for all cities (1) & (5) and for individual cities (2-4) & (6-8) and are reported in square brackets.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

8 The effect of being treated on driver revenue change

To examine the general effect of having a leaderboard, we code the binary variable treated as 0 if the driver is in the control group and as 1 if the driver is in the team or individual leaderboard condition. We use models represented by Equation 1 to capture the effect.

We have discussed the effects of being treated in the main text and Table 1. Here we examine the persistent effects of being treated during the post-intervention contest. We see from the results in Table S16 that the treatment of having a leaderboard marginally significantly improves drivers revenue by 49.44 RMB ($p < .10$, 2.10% of average weekly revenue) in Beijing, but has no significant effect overall, or in Taiyuan or Kunming. Controlling for individual heterogeneity, we find that having a leaderboard increases drivers revenue by 36.72 RMB (1.83% of average weekly revenue) with marginal significance ($p < .10$) and by 55.00 RMB (2.34% of average weekly revenue) with significance ($p < .05$) overall.

Table S16: Average and heterogeneous treatment effects on weekly revenue during the post-intervention contest: Difference-in-differences panel regressions.

	Outcome variable: Δ of Weekly Revenue (CNY)							
	Treatment effects				Control individual heterogeneity			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	All	Beijing	Taiyuan	Kunming	All	Beijing	Taiyuan	Kunming
Treated	30.90	49.44*	-4.25	-11.97	36.72*	55.00**	-1.94	-3.42
(In a virtual team)	(20.81)	(28.32)	(33.04)	(34.82)	(20.45)	(27.77)	(32.22)	(34.38)
		[0.32]	[1.00]	[1.00]		[0.17]	[1.00]	[1.00]
Age					10.57***	11.31***	4.67***	11.61***
(Year)					(1.07)	(1.50)	(1.70)	(1.68)
DiDi age					84.07***	97.86***	38.37**	38.65**
(Year)					(9.62)	(12.33)	(15.43)	(17.18)
Hometown distance					-0.03	-0.04	-0.16**	0.02
to contest city (km)					(0.02)	(0.03)	(0.06)	(0.03)
Self-formed team					-20.27	-38.85	22.50	28.50
					(21.57)	(28.71)	(39.11)	(37.24)
City fixed effect	Yes	-	-	-	Yes	-	-	-
# of clusters	3,970	2,700	545	725	3,970	2,700	545	725
# of drivers	27,790	18,900	3,815	5,075	27,790	18,900	3,815	5,075

Notes: Standard errors in parentheses are clustered at the team level. False Discovery Rate adjusted q-values are calculated separately for individual cities (2-4) & (6-8) and are reported in square brackets.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table S17: Logistic regression results of driver tendency to complete the survey.

	Outcome: Whether driver completes survey			
	(1)	(2)	(3)	(4)
	All	Beijing	Taiyuan	Kunming
Is captain	0.08***	0.09***	0.06***	0.06***
(Binary)	(0.01)	(0.01)	(0.02)	(0.01)
Team won in post-intervention contest	0.12***	0.11***	0.13***	0.13***
(Binary)	(0.00)	(0.01)	(0.01)	(0.01)
Pre-contest average daily revenue	0.07***	0.00	0.22***	0.21***
(in 1000 Yuan)	(0.01)	(0.01)	(0.05)	(0.04)
Male	0.06***	0.03*	0.11***	0.10***
	(0.01)	(0.02)	(0.04)	(0.02)
Hometown distance to contest city	-0.03***	-0.02***	-0.01	-0.02
(in 1000 km)	(0.01)	(0.01)	(0.03)	(0.01)
Age	0.03***	0.03***	0.03***	0.02***
(in 10 years)	(0.00)	(0.00)	(0.01)	(0.01)
DiDi age	0.01***	0.01**	-0.01	0.00
(Year)	(0.00)	(0.00)	(0.01)	(0.01)
# of drivers	34,335	18,900	3,815	5,075

Notes: Average marginal effect with *delta-method* SE in parentheses. The results hold if we alternatively control for the number of wins of the two short-term contests.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

9 Post-experiment Survey and Responses

After the experiment, we sent a survey to every driver in the contest, among whom 4,295 drivers completed our survey constituting 15.46% of our sample.

To examine the tendency of drivers to complete the survey, we conduct logistic regression analysis with results presented in Table S17.

The survey questions and aggregated responses are listed below.

1. To what extent do you like the recent team contest from October 29, 2018 to December 3, 2018? Please rate on a scale between 0 (I don't like it at all) and 6 (I like very much). Depending on your answer, choose either question #3 or #4.

   0 - I don't like it at all. (*201 out of 4,295, 4.68%*)

   1 - I don't like it a moderate amount . (*53 out of 4,295, 1.23%*)

   2 - I don't like it a little. (*75 out of 4,295, 1.75%*)

   3 - Neither like nor dislike. (*196 out of 4,295, 4.56%*)

   4 - I like it a little. (*152 out of 4,295, 3.53%*)

   5 - I like it a moderate amount. (*245 out of 4,295, 5.70%*

   6 - I like it very much. (*3,373 out of 4,295, 78.53%*)

   [Branch: for those who choose like]

2. Why do you like this team contest? (Please check all that apply.)

   (a) Because I like the sense of team belonging. (*2,601 out of 3,966, 65.58%*)

   (b) Because I like the fun and excitement of the contest. (*2,025 out of 3,966, 51.06%*)

   (c) Because I got to know more friends during the contest. (*2,025 out of 3,966, 51.06%*)

   (d) Because winning the contest gave me a sense of honor. (*2,417 out of 3,966, 60.94%*)

   (e) Because I won the monetary bonus. (*2,196 out of 3,966, 55.37%*)

   (f) Other reasons. Please specify ___.

   [Branch: for those who choose dislike]

3. Why do you dislike this team contest? (Please check all that apply.)

   (a) Because my team members were not collaborative or united enough. (*118 out of 330, 35.76%*)

   (b) Because my team was not active enough to justify its existence. (*121 out of 330, 36.67%*)

   (c) Because the captain did not have good leadership or management skills. (*83 out of 330, 25.15%*)

   (d) Because the contest rules were too complicated to understand. (*77 out of 330, 23.33%*)

   (e) Because the contest rules were unfair. (*106 out of 330, 32.12%*)

   (f) Because the financial bonus was not large enough to attract me. (*172 out of 330, 52.12%*)

   (g) Other reasons. Please specify ___.

4. As a team member, what did you get from this team contest? (Please check all that apply.)

   (a) I got to know more friends. (*2,749 out of 4,295, 64.00%*)

   (b) I improved my leadership skills. (*1,443 out of 4,295, 33.60%*)

   (c) I improved my communication skills. (*2,067 out of 4,295, 48.13%*)

   (d) I improved my collaboration skills with other drivers. (*2,541 out of 4,295, 59.16%*)

   (e) I became more experienced and skillful about taking DiDi orders. (*2,452 out of 4,295, 57.09%*)

   (f) I received emotional support from my teammates when I was down. (*1,516 out of 4,295, 35.30%*)

(g) Other reasons. Please specify ___.

5. During this event, which option best describes how your team members got along with each other?

    (a) Our team shared commonalities and common interests. (*586 out of 4,295, 13.64%*)

    (b) Although team members each had our own personalities, we got along well. (*683 out of 4,295, 15.90%*)

    (c) Everyone contributed for our team honor during the contest. (*2,377 out of 4,295, 55.34%*)

    (d) Inactive team members influenced others' enthusiasm for the contest. (*649 out of 4,295, 15.11%*)

    (e) Other reasons. Please specify ___. (0)

6. To what extent do you agree that you have developed deep friendship with your teammates? (from 0 being strongly disagree to 6 being strongly agree)

    0 - Strongly disagree. (*288 out of 4,295, 6.71%*)

    1 - Disagree. (*49 out of 4,295, 1.14%*)

    2 - Somewhat disagree. (*100 out of 4,295, 2.33%*)

    3 - Neither agree nor disagree. (*268 out of 4,295, 6.24%*)

    4 - Somewhat agree. (*203 out of 4,295, 4.73%*)

    5 - Agree. (*264 out of 4,295, 6.15%*)

    6 - Strongly agree. (*3,123 out of 4,295, 72.71%*)

7. (A reverse coding question) To what extent do you not believe that you are a part of your team? (from 0 being not agree at all to 6 being agree very much)

0 - Strongly disagree. (*1,481 out of 4,295, 34.48%*)

1 - Disagree. (*312 out of 4,295, 7.26%*)

2 - Somewhat disagree. (*236 out of 4,295, 5.49%*)

3 - Neither agree nor disagree. (*255 out of 4,295, 5.94%*)

4 - Somewhat agree. (*177 out of 4,295, 4.12%*)

5 - Agree. (*94 out of 4,295, 2.19%*)

6 - Strongly agree. (*1,740 out of 4,295, 40.51%*)

8. Which option do you prefer if you participate in a team contest again?

   (a) I prefer to be a team captain. (*2,648 out of 4,295, 61.65%*)

   (b) I prefer to be a team member. (*1,647 out of 4,295, 38.35%*)

   [Branch: if choose team member]

9. Why did you choose NOT to be a team captain? (Please check all boxes that apply.)

   (a) I don't want to initiate communications with strangers. (*146 out of 1,647, 8.86%*)

   (b) I don't know how to lead a team. (*519 out of 1,647, 31.51%*)

   (c) The extra bonus for a captain was not enough. (*196 out of 1,647, 11.90%*)

   (d) I was concerned that being a captain would entail a lot of extra work. (*257 out of 1,647, 15.60%*)

   (e) I was inexperienced with team leadership and needed more practice in the first place. (*1,053 out of 1,647, 63.93%*)

   (f) Other reasons. Please specify ___.

[Branch: if choose team captain]

10. What do you think a team captain should do? (Please check all boxes that apply.)

    (a) A captain should be a good example for other teammates. (*2,351 out of 2,648, 88.78%*)

    (b) A captain should be positive and energetic. (*2,093 out of 2,648, 79.04%*)

    (c) A captain should help his teammates to become more active. (*2,108 out of 2,648, 79.61%*)

    (d) A captain should help his team win the contest. (*1,940 out of 2,648, 73.26%*)

    (e) A captain should provide feedback and suggestions to the DiDi platform on behalf of team members. (*1,621 out of 2,648, 61.22%*)

    (f) Other. Please specify ___.

11. Through which approach do you prefer to build your team?

    (a) I prefer to wait for others' phone calls to invite me to join a team. (*480 out of 4,295, 11.18%*)

    (b) I prefer to call other people and ask if I can join their team. (*2,983 out of 4,295, 69.45%*)

    (c) I prefer to join a team without prior communication and then contact teammates online. (*832 out of 4,295, 19.37%*)

    (d) Other. Please specify ___.

12. What do you hope would happen to your team?

(a) I hope it was a temporary team and I might be able to join a different team next time. (*3,457 out of 4,295, 80.49%*)

(b) I hope it is a long-lasting team and team members will keep in touch after the contest. (*838 out of 4,295, 19.51%*)

13. How do you communicate with your teammates during the contests?

    (a) WeChat (*3,372 out of 4,295, 78.51%*)

    (b) phone calls (*2,300 out of 4,295, 53.55%*)

    (c) text messages (*1,363 out of 4,295, 31.73%*)

    (d) face-to-face (*966 out of 4,295, 22.49%*)

14. How often do you communicate with your teammates during the first-week contest? During the three weeks in between the contests and during the last contest?

    (a) Never (*First short term: 712 out of 4295, 16.58%; Longer-term: 717 out of 4,295, 16.69%; Post-intervention contest: 755 out of 4,295, 17.58%*)

    (b) Once a week (*First short term: 725 out of 4295, 16.88%; Longer-term: 796 out of 4,295, 18.53%; Post-intervention contest: 757 out of 4,295, 17.63%*)

    (c) Multiple times a week, but not every day (*First short term: 1,142 out of 4295, 26.59%; Longer-term: 1,153 out of 4,295, 26.85%; Post-intervention contest: 1,097 out of 4,295, 25.54%*)

    (d) At least once per day (*First short term: 1,716 out of 4295, 39.95%; Longer-term: 1,629 out of 4,295, 37.93%; Post-intervention contest: 1,686 out of 4,295, 39.25%*)

15. (Treated drivers only.) During the three-week contest (November 5-25, 2018), do you hope to see your team ranking on top? (from 0 being not at all to 6 being very much so)

0 - Not hope so at all (*47 out of 2,824, 1.66%*)

1 - Not hope so (*10 out of 2,824, 0.35%*)

2 - Somewhat not hope so (*28 out of 2,824, 0.99%*)

3 - Neither hope nor not hope (*73 out of 2,824, 2.58%*)

4 - Somewhat hope so (*50 out of 2,824, 1.77%*)

5 - Hope so (*73 out of 2,824, 2.58%*)

6 - Hope so very much (*2,543 out of 2,824, 90.05%*)

16. (Treated drivers only.) During the three-week contest (November 5-25, 2018), which statement(s) about the leaderboard do you agree with? Please check all that apply.

    (a) Although there was no team bonus, keeping the team relationship makes me feel not lonely anymore. (*1,813 out of 2,824, 64.20%*)

    (b) Although there was no team bonus, I was curious about my ranking within my team members. (*1,459 out of 2,824, 51.66%*)

    (c) (Team-leaderboard drivers only.) Although there was no team bonus, I was curious about my team ranking among our competitor teams. (*694 out of 1,390, 49.93%*)

    (d) The ranking was meaningless since there was no monetary bonus, so I didn't care about the ranking and team. (*561 out of 2,824, 19.87%*)

17. On a scale of 0 to 6, 0 being not at all, and 6 being very much so, how would you evaluate your sense of belonging to your team?

    0 - Very not strong (*207 out of 4,295, 4.82%*)

    1 - Not strong (*70 out of 4,295, 1.63%*)

2 - Somewhat not strong (*92 out of 4,295, 2.14%*)

3 - Moderate (*257 out of 4,295, 5.98%*)

4 - Somewhat strong (*205 out of 4,295, 4.77%*)

5 - Strong (*296 out of 4,295, 6.89%*)

6 - Very strong (*3,168 out of 4,295, 73.76%*)

18. On a scale of 0 to 6, 0 being not at all, and 6 being very much so, how would you evaluate your sense of belonging to DiDi?

    0 - Very not strong (*237 out of 4,295, 5.52%*)

    1 - Not strong (*74 out of 4,295, 1.72%*)

    2 - Somewhat not strong (*91 out of 4,295, 2.12%*)

    3 - Moderate (*237 out of 4,295, 5.52%*)

    4 - Somewhat strong (*187 out of 4,295, 4.35%*)

    5 - Strong (*256 out of 4,295, 5.96%*)

    6 - Very strong (*3,213 out of 4,295, 74.81%*)

19. To what extent do you believe that your DiDi income is the primary source of income for your household?

    (a) Yes, it's the only source of income for our household. (*2,076 out of 4,295, 48.34%*)

    (b) It's the primary source of income, but not the only one. (*1,110 out of 4,295, 25.84%*)

    (c) It's a good amount of income, but not the primary income of the household. (*660 out of 4,295, 15.37%*)

(d) It's just an additional source of income. We don't depend on DiDi's income to live a life at all. (*449 out of 4,295, 10.45%*)

20. Why do you want to be a DiDi driver?

    (a) I would like to be a full-time DiDi driver for a long time. (*3,188 out of 4,295, 74.23%*)

    (b) I am and will be a full-time DiDi driver until I find the next job. (*406 out of 4,295, 9.45%*)

    (c) I have another job. I regard DiDi revenue as my extra pocket money in addition to my job. (*375 out of 4,295, 8.73%*)

    (d) I want to kill time by driving. It doesn't matter too much for me whether I make money from it. (*77 out of 4,295, 1.79%*)

    (e) Simply driving is my habit. I like driving. (*249 out of 4,295, 5.80%*)

21. What suggestions do you have for future team activities?

22. Please fill out the phone number which you use to log into the DiDi driver APP: ___.