
Management Science

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

# Putting Teams into the Gig Economy: A Field Experiment at a Ride-Sharing Platform

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

To cite this article:
Wei Ai, Yan Chen, Qiaozhu Mei, Jieping Ye, Lingyu Zhang (2023) Putting Teams into the Gig Economy: A Field Experiment at a Ride-Sharing Platform. Management Science 69(9):5336-5353. https://doi.org/10.1287/mnsc.2022.4624

informs
https://pubsonline.informs.org/journal/mnsc
MANAGEMENT SCIENCE
Vol. 69, No. 9, September 2023, pp. 5336–5353
ISSN 0025-1909 (print), ISSN 1526-5501 (online)



# Putting Teams into the Gig Economy: A Field Experiment at a Ride-Sharing Platform

**Wei Ai,[a] Yan Chen,[b,c,*] Qiaozhu Mei,[b] Jieping Ye,[d] Lingyu Zhang[e]**

[a] College of Information Studies, University of Maryland, College Park, Maryland 20742; [b] School of Information, University of Michigan, Ann Arbor, Michigan 48109; [c] Department of Economics, School of Economics and Management, Tsinghua University, Beijing 100084, China; [d] Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, Michigan 48109; [e] School of Computer Science and Technology, Shandong University, Qingdao 266237, China

*Corresponding author

**Contact:** aiwei@umd.edu, https://orcid.org/0000-0001-6271-9430 (WA); yanchen@umich.edu, https://orcid.org/0000-0003-1287-4525 (YC); qmei@umich.edu, https://orcid.org/0000-0002-8640-1942 (QM); jpye@umich.edu, https://orcid.org/0000-0001-8662-5818 (JY); zhanglingyu@mail.sdu.edu.cn, https://orcid.org/0000-0002-4651-7991 (LZ)

**Received:** June 14, 2020
**Revised:** March 21, 2022
**Accepted:** June 13, 2022
**Published Online in Articles in Advance:** February 2, 2023

https://doi.org/10.1287/mnsc.2022.4624



**Abstract.** The gig economy provides workers with the benefits of autonomy and flexibility but at the expense of work identity and coworker bonds. Among the many reasons why gig workers leave their platforms, one unexplored aspect is the lack of an organization identity. In this study, we develop a team formation and interteam contest field experiment at a ride-sharing platform. We assign drivers to teams either randomly or based on similarity in age, hometown location, or productivity. Having these teams compete for cash prizes, we find that (1) compared with those in the control condition, treated drivers work longer hours and earn 12% higher revenue during the contest; (2) the treatment effect persists two weeks postcontest, albeit with half of the effect size; and (3) drivers in hometown-similar teams are more likely to communicate with each other, whereas those in age-similar teams continue to work longer hours and earn higher revenue during the two weeks after the contest ends. Together, our results show that platform designers can leverage team identity and team contests to increase revenue and worker engagement in a gig economy.

**History:** Accepted by David Simchi-Levi, behavioral economics and decision analysis.

**Funding:** Financial support from the platform through the Michigan Institute for Data Science is gratefully acknowledged.
**Supplemental Material:** The e-companion are data are available at https://doi.org/10.1287/mnsc.2022.4624.

**Keywords:** team • contest • organization identity • gig economy • ride sharing

## 1. Introduction

As trends in work sourcing move the world toward a gig economy, this economy is widely considered to be the future face of work despite questions about its sustainability (Ravenelle 2019). Whereas workers in traditional sectors derive their identities from their work and share their experiences with coworkers, those whose livelihood relies on the gig economy often find that "these are jobs that don't lead to anything," citing a lack of work identity or bonds with coworkers as well as an inability to move upward even with strong performance (Heller 2017). Ride-sharing platforms, in particular, such as Uber, experience especially high attrition rates (Scheiber 2017). This study examines one reason why gig workers leave their platforms: the lack of an organization identity.

The Covid-19 pandemic created a work structure that placed many workers in a work-from-home scenario susceptible to the same issues related to the lack of in-person interaction for those in a gig economy. In September 2021, 45% of full-time U.S. employees worked from home either all or part of the time (Saad and Wigert 2021). This trend continues into 2022. Given that we expect at least some portion of this remote work to continue in the future, an important question is how organizations can help their workers create and maintain positive work-related social connections when working remotely.

To analyze how work connections are impacted in the gig economy and remote work, we apply social identity theory (Tajfel et al. 1971, Tajfel and Turner 1979, Akerlof and Kranton 2000) to a large gig platform (the platform henceforth), on which individual drivers offer ride-sharing services in the Asia-Pacific region, Africa, and Latin America. Specifically, we design a field experiment to study team formation and interteam contests within the platform. In our experiment, we examine the impact of the creation of a team identity on individual driver revenue. Furthermore, we make use of the flat organization structure of the platform to investigate how different team formations impact team member communication and productivity.

Our research applies insights from the social psychology (Tajfel et al. 1971, Tajfel and Turner 1979, Brewer 1999) and behavioral economics literature on identity (Akerlof and Kranton 2000, 2010). This research shows that, when people feel a stronger sense of common identity with a group, they exert higher effort and make more contributions to public goods to reach a more efficient outcome in the laboratory. This result holds for identities that are induced (Eckel and Grossman 2005, Charness et al. 2007, Chen and Chen 2011) or natural (Goette et al. 2012, Chen et al. 2014). Moving from the laboratory to the field, results are mixed. In an early public goods field experiment, Erev et al. (1993) use team competition in fruit harvesting and find that team competition increases productivity. Likewise, Ai et al. (2016) report the results of a large-scale field experiment designed to test the hypothesis that team membership can increase prosocial lending in an online microfinance community, Kiva.org. They find that team recommendations increase the likelihood that a lender joins a team and joining a team substantially increases postintervention lending. In an American Red Cross field experiment, Kessler and Milkman (2018) show that individuals are more likely to donate when a facet of their identity associated with the generosity norm is primed, illustrating the practical use of identity primes to encourage public good provision. However, in another study, Gee et al. (2020) randomly assign potential donors into teams with varying within-team social distances and find no evidence that reduced social distance increases giving. In a recent survey of theoretical and experimental identity-economics research, Charness and Chen (2020) conclude that identity-based teams in the field might be a useful behavioral mechanism to increase prosocial behavior. Applying this framework to our setting, we anticipate that a driver with a stronger sense of team identity works harder to help the team get ahead compared with drivers who do not belong to any team (Brewer and Silver 2000).

In examining how different team formations may have different effects on communication and coordination, we use an algorithm that maximizes either similarity or diversity within a team. We conjecture that similarity might facilitate team-member communication and coordination, leading to intrateam bonding and team stability (Kim and Aldrich 2002, Ruef et al. 2003). Indeed, empirical network science studies provide evidence for the phenomenon of homophily or the tendency of people to associate with others whom they perceive as similar to themselves in some way (McPherson et al. 2001, Girvan and Newman 2002). By contrast, we conjecture that diversity might bolster team performance because of different perspectives in problem-solving and better complementarity among team members (Krishnan et al. 1997, Page 2007). Thus, the effect of team composition on performance is an empirical question.

In addition to examining different team-formation algorithms, we draw on insights from contest theory (Konrad 2009, Vojnović 2016) and experiments (Sheremeta 2018) to evaluate differences in contest outcomes based on contest structure as team contests are shown to be among the most effective ways to strengthen team identity in the laboratory (Eckel and Grossman 2005).

Finally, our work contributes to the rapidly growing literature on the gig economy and ride sharing in particular, which has uncovered important insights related to labor market outcomes (Hall and Krueger 2018, Jackson et al. 2017), the value of flexible work (Chen et al. 2019), consumer surplus (Cohen et al. 2016), and decentralized dynamic matching efficiency (Liu et al. 2019). Our findings contribute to this stream of research by showing that a team-based approach can significantly increase drivers' revenue, bonds with coworkers, and team identities. More broadly, our research demonstrates the value of a social-relational approach in integrating teams and social relationships into the gig economy.

## 2. Experiment Design

To test the effectiveness of team formation and interteam competition on productivity, we design a multistage natural field experiment using the ride-sharing platform. Founded in 2012, the platform is the dominant ride-sharing company in Asia. The platform employs more than 31 million drivers globally and offers app-based transportation options for 550 million users across Asia, Latin America, and Australia, making it the largest ride-sharing platform in the world.[1] In China, the platform driver base is largely composed of workers laid off from their traditional jobs, veterans, migrant workers from rural areas, and workers who offer rides during their daily commute to their job. On the platform, drivers receive 81% of the revenue they generate and give the remaining 19% to the platform.

The platform is not immune to the low engagement problem faced by other gig economy platforms (Ravenelle 2019). To assess the effects of team contests on driver engagement levels, we collaborate with the AI Labs of the platform to conduct our field experiment. Our experiment consists of three stages: recruitment, team formation, and team contest. In what follows, we present our design choices in each stage.

### 2.1. Recruitment and Power Analysis

We conduct our experiment in the summer of 2017 in the southern city of Dongguan, which has 480,000 registered drivers. We select our pool of drivers based on their productivity in a two-week period (July 18–31, 2017) prior to the announcement of the contest, using the following two criteria to filter the drivers. First, the driver has finished one or more trips on at least five weekdays and two weekend days during the two-week period. Second, the driver finishes five or more trips on average on the days the driver works during the two-week period. This filtering process yields a total of 28,394 eligible drivers. From this pool, we randomly select 24,000 drivers to receive a text message invitation. The remaining 4,394 drivers compose our no-contact group.

To determine the number of drivers needed in our experiment, we use the observational data from 9,000 randomly selected drivers who participated in a team contest organized by the platform in Beijing in late January and early February 2017.[2] We find that drivers in that random sample complete, on average, 11.7 trips per day ($\sigma = 4.7$) during the contest period. Based on their performance before and during the contest, we expect an effect size of 10% ($\delta = 1.17$). Setting $\alpha = 0.05$ and $\beta = 0.10$ (90% power) requires us to have 340 drivers per experimental condition, assuming equal variance across experimental conditions (List et al. 2011). As each team has seven drivers, the number of drivers per treatment should be a multiple of seven. This leads us to selecting a sample size of 350 drivers per experimental condition.

We elicit participation interest by sending each potential participant a text message. In each text message, we ask if the driver would like to register for a team contest in which the driver might find new friends and earn 1,000 Chinese Yuan (CNY) or more together as a team if they win.[3] Additionally, we ask if a driver is interested in becoming a team captain and earning an additional 100 CNY upon fulfillment of the captain's duties. Our announcement received 2,343 positive responses, 531 of which were interested in being a team captain. These text messages are included in the Electronic Companion Section EC.1.

We implement a $5 \times 3$ factorial design in two stages. In the next two sections, we explain each factor and our randomization procedures. Figure 1 presents an overview of the experimental procedure.

### 2.2. Team Formation

Our experiment consists of two levels of randomization. First, we randomize our 2,343 positive responses into seven groups: 1,750 drivers are randomized into one of five team-formation algorithms, each of which contains 350 drivers (groups 1–5). These drivers are subsequently partitioned into teams of seven (250 teams in total). Note that 350 drivers are randomized into the control group (group 6). These drivers are not placed in a team. During the contest period, they continue to earn piece rate. The remaining 243 drivers serve as backups in case drivers in the treatments drop out before the start of the contest (group 7). Indeed, in our experiment, 15 drivers were reported by their captains as not responsive or no longer available for the contest. We mark these 15 drivers as "dropouts" and replace them with similar drivers from the backup group.[4]

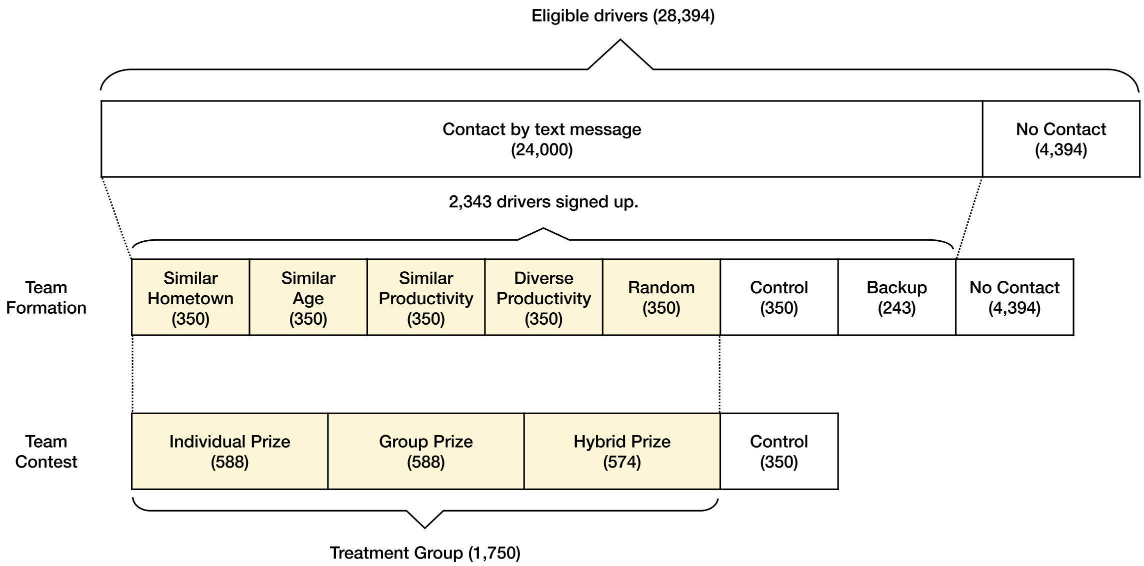
The pseudocode for our team-formation algorithm is presented as Algorithm 1 in Section EC.4 and briefly described as follows. We first randomly assign the 2,343 drivers into seven groups (five team-formation algorithms, the control group, and the backup group). We repeat the random assignment until the following conditions are satisfied: (1) covariates are balanced across conditions, (2) each team-formation algorithm contains enough volunteers willing to serve as captains, and (3) the demographic distribution meets the requirement of each team-formation algorithm, for example, age similarity. We then group drivers in each treatment into 50 teams. We choose five team-formation algorithms based on similarity or diversity considerations. In what follows, we describe the rationale and implementation for selecting each team-formation method.

Our first condition, hometown similarity, is based on previous findings that location similarity is the most effective characteristic in motivating a microfinance community member to join a specific lending team (Ai et al. 2016). In our study, we use hometown similarity, a form of location similarity, by assigning drivers from the same (or a nearby) province to the same team. Prior studies indicate that hometown location is a salient identity among migrant workers in China (Zhang and Xie 2013). We partition the drivers into teams to ensure the seven drivers in the same team are all from the same (or a nearby) province.

Our second condition, age similarity, is based on prior research illustrating that age cohorts, such as the generation born in the 1980s, are meaningful identity groups in China. Those within these age cohorts are linked by their common socialization experiences (Harmel and Yeh 2015). Therefore, we form our age-similar teams to reflect an age span of five years, for example, 1980–1984, 1985–1989, with the exception of the 1960–1969 cohort, which spans a decade because of the relatively lower number of drivers born in the 1960s in our sample. Within each age cohort, we randomize the drivers into teams of seven. We randomize several times until we reach a partition that ensures at least one driver in each team has volunteered to be a team captain.

Third, we include productivity similarity as one of our algorithms as it is the preferred team formation algorithm

**Figure 1.** (Color online) Experimental Procedure




by the platform. For this condition, we partition the 350 drivers into 10 buckets (35 each) based on their number of trips completed in the two weeks prior to the announcement of the team contest. Within each bucket, we randomize the drivers into teams of seven. We randomize several times until we reach a partition that ensures at least one driver in each team has volunteered to be a team captain.

Fourth, we draw on recent scholarly research supporting the advantages of diversity (Page 2007) and use two strategies to create diverse teams. To achieve productivity diversity in our teams, we partition the 350 drivers into seven buckets based on their number of trips completed in the two weeks prior to the announcement of the team contest in our experiment. We then randomly select one driver from each bucket to form a team. We randomize several times until we reach a partition that ensures at least one driver in each team has volunteered to be a team captain. Therefore, each team consists of drivers from all seven buckets.

Our final algorithm, random formation, reflects the diversity achieved from a random grouping of drivers. For this condition, we randomly partition the drivers into teams of seven and repeat the randomization until at least one driver in each team has volunteered to be a team captain.

In sum, our team formation stage yields a total of 1,750 treatment drivers formed into 250 teams with seven drivers in each team and 50 teams in each treatment. As we have 531 volunteers for 250 captain positions, we randomly select one volunteer to be the captain whenever a team has more than one volunteer. This randomization enables us to estimate the effect of being a captain on productivity and other metrics.

Table 1 provides the summary statistics by experimental condition and reports our randomization checks. From Table 1, we see that our drivers, on average, are 35 years old, have been driving for the platform for 10 months, are predominantly male (98%), and are from different regions with a quarter of them from Dongguan. Joint orthogonality tests between the treatment and control conditions indicate that this level of randomization yields balanced experimental groups ($p > 0.10$). Note that the no-contact condition is not part of the experiment. Comparing our no-contact group with both our control and treatment groups, we find that the no-contact drivers earn significantly less revenue than the participants in our experiments ($p < 0.001$, $t$-test), indicating that the drivers who sign up for our contests are highly selected.

## 2.3. Team Contest Rules

Our contest rules are based on the structures of team contests within inherently individualistic sports, such as tennis and chess, in which team outcomes are determined by multiple pairwise battles. Specifically, we set up a contest in which drivers from two rival teams form pairwise matches to engage in distinct component battles. Within each team, we use an algorithm to automatically rank drivers by their number of trips completed in the two weeks prior to the announcement of the contest and pair the most productive driver in team A with the most productive one in team B and so on. In this contest, a team wins if and only if its drivers win a majority of their battles. This team contest format is often deployed in table tennis and badminton competitions, both of which are sports that are popular in China and with which our drivers are familiar. Fu et al. (2015) provide a theoretical analysis of this type of team contest under complete

**Table 1.** Summary Statistics and Randomization Check by Team Formation Algorithm

| | No contact | Control | Hometown similarity | Age similarity | Productivity similarity | Productivity diversity | Random | $p$-value |
|---|---|---|---|---|---|---|---|---|
| Daily revenue | 223.8 | 269.0 | 264.4 | 266.6 | 255.9 | 265.9 | 260.9 | 0.691 |
| (CNY) | (109.7) | (116.0) | (117.2) | (111.8) | (107.8) | (116.1) | (105.2) | |
| Local | 0.307 | 0.24 | 0.231 | 0.254 | 0.3 | 0.254 | 0.231 | 0.293 |
| | (0.461) | (0.428) | (0.422) | (0.436) | (0.459) | (0.436) | (0.422) | |
| Age | 36.866 | 35.16 | 35.483 | 34.786 | 35.057 | 34.609 | 35.506 | 0.524 |
| | (7.775) | (7.637) | (7.273) | (7.277) | (7.620) | (7.201) | (7.462) | |
| Platform age | 0.806 | 0.869 | 0.868 | 0.879 | 0.868 | 0.850 | 0.849 | 0.982 |
| (Year) | (0.560) | (0.559) | (0.587) | (0.582) | (0.565) | (0.595) | (0.557) | |
| Male | 0.976 | 0.983 | 0.983 | 0.963 | 0.991 | 0.977 | 0.971 | 0.147 |
| | (0.152) | (0.130) | (0.130) | (0.189) | (0.092) | (0.150) | (0.167) | |
| Number of drivers | 4,397 | 350 | 350 | 350 | 350 | 350 | 350 | |

*Notes.* Standard deviations appear in parentheses. $p$-values are from joint orthogonality tests between the control and the treatment groups. "No contact" is not part of either the control or treatment condition.

information and sequential moves. In our ride-sharing context, because we conduct a field experiment, our drivers also earn their piece rate in addition to any prize money. Furthermore, each driver finds out about the outcome at the end of the contest day, making our contest an incomplete-information, simultaneous-move team contest. Finally, as there is some randomness in which revenue depends on effort, we model the contest as a rank order tournament (Lazear and Rosen 1981) in our theoretical framework in Section EC.5. While these features differ from the settings in Fu et al. (2015), their theoretical analysis is one of the reasons for us to choose this contest format.

Table 2 illustrates the prize structure in each of the individual, group, and hybrid prize allocation conditions. Under the individual prize condition, the driver who wins the contest receives a 30 CNY prize regardless of team performance. Under the group prize condition, each driver in a team that wins a majority (four or more) of its contests receives a 30 CNY prize. Under the hybrid prize condition, drivers receive both individual and group prizes of 15 CNY each. The prizes are set such that the expected reward per driver remains the same across treatments, which is 15 CNY under the symmetry assumption. The allocation rules are explained in the newsletter released to the drivers prior to the contest (see Section EC.3 for the full-text translation of the newsletters). Finally, each contest consists of seven component battles, in which the drivers compete based on the number of trips they finish in one day of competition. Whereas revenue is the outcome that both drivers and the platform seek to maximize, prior to our experiment, the platform was also using the number of trips as the outcome metric for any additional incentives or bonuses, primarily for simplicity. Therefore, we use the number of trips as our outcome measure in the experiment, which is highly correlated with revenue.

Our second level of randomization assigns teams to prize structures. The pseudocode for this algorithm is presented as Algorithm 2 in Section EC.4. We first randomly assign each of the 250 teams into one of three prize structures, checking that each prize structure has an even number of teams and the covariates are balanced. The process repeats until both conditions are satisfied. Next, for each prize structure, we randomly select two teams from the same team formation algorithm and pair them together. We repeat this process until there are no more remaining teams constructed via the same team-formation algorithm. Of the 250 teams, 222 (89%) are paired with a team from the same team-formation algorithm. We randomly match the remaining 28 teams (11%) into pairs.

Table 3 presents the summary statistics and balance tests by prize structure. From Table 3, we see that the covariates are balanced across control and prize structure conditions. We present the 5 (team formation algorithms) × 3 (prize structures) design, summary statistics, and balance tests in Table EC.1, Section EC.6. These statistics show that covariates are balanced except for the proportion of local drivers from Dongguan ("Local"). In all subsequent analyses, we control for this and other demographic variables.

## 2.4. Within-Team Communication

Within each team, we identify a team captain who is notified of this position, given the phone number of each team member, and asked to complete a precontest survey. The

**Table 2.** Prize Structure

| Prize structure | Individual wins | Team wins |
|---|---|---|
| Individual-prize treatment | 30 | — |
| Group-prize treatment | — | 30 |
| Hybrid-prize treatment | 15 | 15 |

*Notes.* This table indicates the prize that drivers get if they win the individual contests (individual wins), if their team wins a majority of the contests (team wins), or both. The prize is calculated for each contest based on the number of trips a matched pair of drivers make on that day.

**Table 3.** Summary Statistics and Randomization Check by Prize Structure

| | No contact | Control | Individual prize | Group prize | Hybrid prize | *p*-value |
|---|---|---|---|---|---|---|
| Daily revenue | 223.8 | 269.0 | 263.6 | 266.0 | 258.5 | 0.527 |
| (CNY) | (109.7) | (116.0) | (112.4) | (110.6) | (112.0) | |
| Local | 0.307 | 0.24 | 0.248 | 0.238 | 0.277 | 0.421 |
| | (0.461) | (0.428) | (0.432) | (0.426) | (0.448) | |
| Age | 36.9 | 35.2 | 35.6 | 34.9 | 34.7 | 0.252 |
| | (7.8) | (7.6) | (7.2) | (7.3) | (7.6) | |
| Platform age | 0.806 | 0.869 | 0.892 | 0.834 | 0.863 | 0.377 |
| | (0.560) | (0.559) | (0.570) | (0.568) | (0.592) | |
| Male | 0.976 | 0.983 | 0.980 | 0.986 | 0.965 | 0.078 |
| | (0.152) | (0.130) | (0.142) | (0.116) | (0.184) | |
| Number of drivers | 4,397 | 350 | 588 | 588 | 574 | |

*Notes.* Standard deviations appear in parentheses. *p*-values are from joint orthogonality tests between the control and the treatment groups. "No contact" is not part of either the control or treatment condition.

survey requires captains to communicate individually with each driver in the team to obtain the last three digits of the driver's license plate number as well as several key pieces of demographic information (see Section EC.2 for survey questions). Meanwhile, team members receive a text containing the captain's phone number that informs them the captain might call them. Each team captain earns 100 CNY, which is public information announced in the initial text message.

The team communication task is designed to nudge the captains to communicate and collaborate with their team members as several laboratory experiments demonstrate that taking part in a collaborative problem-solving task involving group communications can strengthen both group identity (Chen and Li 2009, Chen and Chen 2011, Chen et al. 2020) and within-group coordination (Cason et al. 2012). As the platform does not provide the option for team communication, our postexperiment survey indicates that most teams communicate by phone or WeChat.[5] Captains who submit the survey through an online form are given 50 CNY as a bonus regardless of the correctness of their answers.

If a captain submits the survey, we mark the team as *responsive* because a complete survey indicates captain communication to the team and team member communication to the captain. In comparison, a nonresponsive team is defined as one for which the captain does not submit the survey. A captain might not submit a survey if (1) the captain does not contact team members, (2) team members do not respond, or (3) the captain and the team communicate but the captain does not submit the survey. The first two cases indicate either the captain or team members are nonresponsive, whereas the last case simply indicates a submission lapse. Whereas our design cannot disentangle these scenarios, nonsubmissions resulting from a lapse only bias our estimates of any communication effects downward. In our sample, 60.8% of our captains submit their survey. Conditional on submitting the survey, 81.1% of the answers are correct, indicating that responsive team members communicate with their captain and each other.

Note that our precontest survey encourages team members to learn their teammates' hometowns (question 3) and ages (question 4). Thus, we assume that only members of responsive teams are likely to identify the degree of potential homogeneity within their teams.

The contest was implemented between August 13 and 21, 2017, with one day off between every two contest days. Before each contest day, we reset the contest. We then repeat it five times with the same pairing of teams. The contest results are calculated at the end of each contest day and communicated to each driver on the following day. In addition to the contest days, we also obtain data on drivers two weeks prior to and four weeks after the contest.

## 3. Hypotheses

To motivate our hypotheses, we begin with a simple theoretical framework of team contests, with and without team identity, characterizing the solutions in Section EC.5.

Following Lazear and Rosen (1981), we model our contest as a rank order tournament as there is some randomness in which revenue depends on effort.[6] For driver $i$, let $e_i \in [0, \bar{e}]$ denote effort, which can be approximated by the number of hours $i$ drives each day. The platform imposes a maximum of 10 hours of driving per day, which justifies the upper bound, $\bar{e}$. Let $x_i = e_i + \epsilon_i$ be driver $i$'s output, such as the number of trips completed, where $\epsilon_i$ is the random component drawn from a known distribution with zero mean and variance $\sigma^2$. Let $c(\cdot)$ denote the cost function, which is assumed to be convex, that is, $c'(\cdot) > 0$, and $c''(\cdot) > 0$. Let $w > 0$ be the piece rate for all drivers.

In the control condition, in which a driver earns piece rate, a risk-neutral driver chooses an effort level to maximize expected income, $wE(x_i) - c_i(e_i)$. The following first order condition characterizes the interior solution of the driver's optimal effort level: $c'(e) = w$. In other words, a driver equates marginal cost of effort with the wage rate.

Under the individual prize contest rule, a driver wins a cash prize, $V$, if the driver completes more trips than a matched driver regardless of whether the driver's team wins. In our model without team identity, this reduces to a modified two-player rank order tournament with the extra component of piece rate, $wx_i$. Here, driver $i$ chooses effort level $e_i$ to maximize the following expected utility function:

$$EU_i = P(x_i \geq x_j)V + wE(x_i) - c_i(e_i). \qquad (1)$$

We characterize the solution to Equation (1) as well as those to the other two prize structures in Section EC.5. Based on our theoretical analysis, we formulate the following hypotheses.

**Hypothesis 1** (Contest Effect). *Drivers in a team contest exert greater effort than those in the control condition.*

Hypothesis 1 is based on Observation 1 in Section EC.5, namely, that a contest prize represents a sufficient monetary incentive for drivers to increase their effort even when there is no team identity component.

We now consider the effects of team identity on driver effort. Eckel and Grossman (2005) demonstrate that interteam competition is among the strongest methods to induce team identity in the laboratory. According to Tajfel and Turner (1979), an important part of the social identification process is social comparison. In our context, drivers who are put into teams with which they subsequently identify tend to compare their team with a rival team and to maintain their self-esteem by comparing their team's performance favorably with that of the rival team. Based on this theory, we use a simple reduced-form method to incorporate team identity into our contest framework. Specifically, we use $\alpha_r \geq 1$, $r \in \{I, G, H\}$, representing the individual, group, and hybrid prize, respectively, to denote the strength of a player's team identity under contest rule $r$ or the weight the player puts on the player's team winning the contest.

Under the individual prize rule, even though a player wins the prize based on only the player's performance, this performance contributes to the team's performance. Therefore, the player might care more about winning the individual battle. In this case, the objective function (1) becomes $EU_i = \alpha_I P(x_i \geq x_j)V + we_i - c_i(e_i)$.

Based on prior field experiments on team competition (Ai et al. 2016), we expect that teams with strong natural identities perform better than those without. In our experiment, hometown-similar teams comprise drivers from the same (or a nearby) province, whereas age-similar teams share similar socialization experiences—each the basis for a meaningful and salient team identity. Again, based on Observation 3 in EC.5, we expect members of these teams to exert greater effort than those from teams with weaker identities, such as randomly formed teams. This leads to our next hypothesis.

**Hypothesis 2** (Team Composition). *Teams based on strong and salient identities, such as hometown similarity, exert greater effort than those based on weaker identities, such as randomly formed teams.*

Finally, Observation 2 in EC.5 suggests that drivers focused solely on the monetary prize should show the greatest effort under the individual versus other prize treatments. However, when team identity is incorporated into the contest framework, the effort ranking might be different. Specifically, as both the individual and, to a lesser extent, hybrid prize rules prime the importance of the individual, whereas the group prize rule primes the importance of the team, we expect that drivers exhibit the strongest team identity under the group prize rule. Observation 4 postulates that, under the group prize rule, a sufficiently strong team identity can lead to greater effort compared with that under either the individual or hybrid prize rule. Under the assumption of a normal distribution of the noise term, $\epsilon \sim N(0, \sigma^2)$, if $\alpha_G > \frac{16}{5}\alpha_I$, we have $e_G > e_I$. Similarly, when $\alpha_G > \frac{21}{10}\alpha_H$, we have $e_G > e_H$. Therefore, compared with the case without any team identity (Observation 2), the effort ranking with team identity might be different. In particular, based on Observation 4, we formulate the following hypothesis.

**Hypothesis 3** (Prize Structure). *When team identity is sufficiently strong, drivers under the group prize rule exert greater effort compared with those under either the individual or hybrid prize rule.*

## 4. Results

In this section, we present the results from our field experiment. We first examine the effect of our contest on driver working hours, number of trips completed, and revenue. We then examine the impact of team formation on team efficacy and performance, followed by the effects of the prize-allocation conditions and then the effects of leadership experience. Finally, we end with a discussion of the contest dynamics and the platform's return on investment.

### 4.1. Average Treatment Effects

We first investigate the average treatment effect, that is, the effect of the team contest on each driver's number of hours worked, number of completed trips, and revenue. Figure 2 presents our results for driver daily revenue before, during, and after the contest period by experimental condition.[7] The lines correspond to the three experimental conditions: drivers who were never contacted (no contact, light dashed line), those who expressed interest but were not assigned to a team (control, black dashed line), and those who expressed interest and were assigned to a team (treatment, solid line).

We refer to the five days of our team contest as *contest days* and the 14 days prior to (post) the contest as the *pre* (*post*) *contest* periods. Finally, to investigate whether our effects last more than two weeks, we create a four-week postcontest period. Our choice of windows ensures that we always compare the same day of the week precontest, contest, and postcontest. During our experiment, we record daily data on each driver, including the number of hours worked, the number of completed trips, and the revenue generated. On the platform, drivers receive 81% of the revenue they generate and give the remaining 19% to the platform.[8]

Returning to Figure 2, we see that those who sign up to join a team, regardless of whether they are assigned to a treatment or control condition, generate higher revenue than those who are never contacted (gray dashed line). Figure 2 also shows that both the control and no-contact groups exhibit a similar decreasing trend over the eight-week time period of our experiment, a pattern similar to the platform's typical attrition rate.[9]

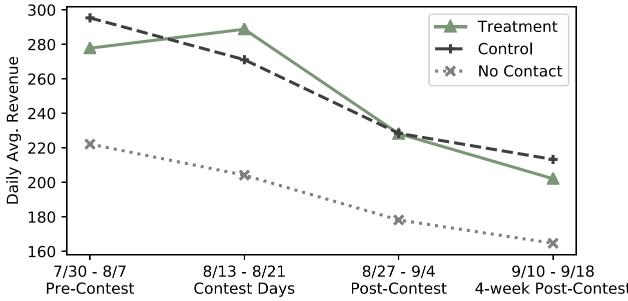
To quantify the average treatment effects on outcome, $Y$, we construct the following difference-in-differences model:

$$\Delta Y_{i,t} = \beta_0 + \beta_1 * \text{Treated} + \epsilon_{i,t}, \qquad (2)$$

where $\Delta Y_{i,t}$ represents the outcome change on the $t$th day in the current period compared with the $t$th day in the precontest period. We report the results of these models in Tables 4 (daily number of hours worked), 5 (daily number of trips completed), and 6 (daily revenue). Each table includes the average treatment effects without (specifications 1–4) and with (specifications 5–8) demographic controls.

Pooling drivers across all treatment and control conditions, we find that treated drivers work an additional 0.77 hours or 46 minutes per day (columns (2) and (6) in Table 4), which translates into 2.4 extra trips per day (columns (2) and (6) in Table 5). This increase in activity corresponds to an increase in treated drivers' daily revenue of 35 CNY (12%) during the contest period compared with that of control drivers (columns (2) and (6) in Table 6). In this and subsequent comparisons, we use as our baseline the average control driver's daily working hours (6.51), number of trips (17.73), or revenue (295.37 CNY) in the five days corresponding to contest days during the two weeks before the contest. Further analyses show that our treatment effect persists during the two-week postcontest period, albeit with half of the effect size (0.38 hours, $p < 0.01$; 1.2 trips, $p < 0.01$; 17.6 CNY, $p < 0.05$, Tables 4–6, column (7), respectively).

**Figure 2.** (Color online) Driver Daily Revenue Before, During, and After the Contest




*Notes.* Contest days refer to August 13, 15, 17, 19, and 21, the dates on which the contests were conducted. We shift the dates by −14, +14, and +28 days to obtain the precontest, postcontest, and four-week postcontest periods. Note that driver revenue is calculated for only the five days in each period accordingly.

In all our analyses, we correct for multiple hypothesis testing (Benjamini and Hochberg 1995) using the Stata code provided by Anderson (2008) and report the false discovery rate adjusted $q$-values in square brackets. We follow the convention of using a 5% (10%) cutoff for $p$-values ($q$-values) to claim statistical significance (Efron 2012). We summarize our first result as follows.

**Result 1** (Average Treatment Effect: Team Contest). During the contest period, drivers who are part of a treatment team work 46 minutes longer, complete 2.4 extra trips, and generate 12% higher revenue per day compared with those in the control condition. These treatment effects remain significant during the two-week postcontest period, albeit with half of the effect size.

By Result 1, we reject the null in favor of Hypothesis 1 that drivers in a team contest exert greater effort than those in the control condition. We find that this effect size is both statistically and economically significant. Furthermore, the average treatment effects during both the contest and the two-week postcontest periods are robust to demographic controls (columns (5)–(8) in Tables 4–6). We also find that adding demographic controls reveals a significant association between work experience (platform age) and revenue both during (17 CNY, $p < 0.01$) and two weeks after (18 CNY, $p < 0.01$) the contest period. Finally, the results in column (6) in Table 6 show that our observed increase in treatment driver revenue is positively associated with the driver's age, hometown proximity to Dongguan, and male gender.[10]

In our experiment, treated drivers' increase in revenue comes primarily from longer hours worked rather than faster driving or location preferences. In an empirical study, Cook et al. (2018) find that their gender earnings gap among Uber drivers can be explained by three factors: driving experience, location preference, and speed. Our study confirms their effect of driving experience as the variable, platform age, has a statistically significant and economically substantial positive correlation with revenue (columns (6) and (7) in Table 6). We do not observe location preferences in our data. Regarding speed, this factor is often out of the control of our drivers because of the high traffic congestion in their driving environment.

**Table 4.** Average Treatment Effects on Daily Working Hours: Difference-in-Differences Linear Regressions

| | Dependent variable: change in daily working hours | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Time period | (1) First day of contest | (2) During contest | (3) Two weeks postcontest | (4) Four weeks postcontest | (5) First day of contest | (6) During contest | (7) Two weeks postcontest | (8) Four weeks postcontest |
| Treated | 0.968*** | 0.772*** | 0.379* | 0.134 | 0.973*** | 0.772*** | 0.379*** | 0.127 |
| | (0.331) | (0.191) | (0.198) | (0.214) | (0.330) | (0.139) | (0.143) | (0.144) |
| | [0.007] | [0.001] | [0.075] | [0.531] | [0.007] | [0.001] | [0.013] | [0.432] |
| Age | | | | | 0.0340** | 0.0138** | 0.00832 | 0.0170** |
| | | | | | (0.0167) | (0.00703) | (0.00722) | (0.00728) |
| Platform age, year | | | | | 0.232 | 0.374*** | 0.397*** | 0.179* |
| | | | | | (0.216) | (0.0910) | (0.0934) | (0.0943) |
| Local | | | | | 0.663** | 0.359*** | 0.152 | 0.686*** |
| | | | | | (0.284) | (0.120) | (0.123) | (0.124) |
| Male | | | | | 1.837** | 0.343 | −0.108 | 0.128 |
| | | | | | (0.843) | (0.356) | (0.365) | (0.368) |
| Constant | −0.393 | −0.521*** | −1.579*** | −1.225*** | −3.754*** | −1.754*** | −2.147*** | −2.270*** |
| | (0.302) | (0.175) | (0.181) | (0.196) | (1.068) | (0.450) | (0.462) | (0.467) |
| Number of drivers | 2,100 | 2,100 | 2,100 | 2,100 | 2,100 | 2,100 | 2,100 | 2,100 |
| Observations (number of drivers × number of days) | 2,100 | 10,500 | 10,500 | 10,500 | 2,100 | 10,500 | 10,500 | 10,500 |

*Notes.* Standard errors in parentheses are clustered at the driver level. False discovery rate adjusted $q$-values are in square brackets to correct for multiple hypothesis testing.
$^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$.

Next, we explore whether there is any treatment effect on efficiency, that is, revenue per hour worked. The results in Table EC.5 show that there is no significant treatment effect on efficiency ($-0.0703$, $p > 0.10$, column (4)). Consistent with Cook et al. (2018), we find that male drivers are more efficient than their female counterparts (7.096, $p < 0.01$, column (6)) although the latter comprises only 2% of our sample.

Finally, we acknowledge that our observed treatment effect could be driven by a combination of factors, including monetary prizes, competition, and team membership. Within our experimental design, our control drivers receive no opportunity to increase pay, whereas our treatment drivers are given the opportunity to earn more via the contest.[11] Thus, it is possible that the treatment drivers show an increase

**Table 5.** Average Treatment Effects on Daily Number of Trips: Difference-in-Differences Linear Regressions

| | Dependent variable: change in daily trips | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Time Period | (1) First day of contest | (2) During contest | (3) Two weeks postcontest | (4) Four weeks postcontest | (5) First day of contest | (6) During contest | (7) Two weeks postcontest | (8) Four weeks postcontest |
| Treated | 3.114*** | 2.392*** | 1.219** | 0.462 | 3.131*** | 2.393*** | 1.223*** | 0.445 |
| | (0.958) | (0.522) | (0.544) | (0.543) | (0.955) | (0.387) | (0.400) | (0.379) |
| | [0.003] | [0.001] | [0.034] | [0.395] | [0.003] | [0.001] | [0.004] | [0.276] |
| Age | | | | | 0.0791 | 0.0405** | 0.0386* | 0.0393** |
| | | | | | (0.0482) | (0.0195) | (0.0202) | (0.0191) |
| Platform age, year | | | | | 0.534 | 0.952*** | 0.785*** | 0.218 |
| | | | | | (0.624) | (0.253) | (0.262) | (0.248) |
| Local | | | | | 1.799** | 1.057*** | 0.120 | 1.582*** |
| | | | | | (0.821) | (0.332) | (0.344) | (0.326) |
| Male | | | | | 5.919** | 1.355 | −0.191 | 0.227 |
| | | | | | (2.438) | (0.987) | (1.022) | (0.968) |
| Constant | −3.137*** | −2.032*** | −4.408*** | −5.082*** | −12.63*** | −5.869*** | −6.289*** | −7.253*** |
| | (0.874) | (0.477) | (0.497) | (0.496) | (3.089) | (1.250) | (1.294) | (1.226) |
| Number of drivers | 2,100 | 2,100 | 2,100 | 2,100 | 2,100 | 2,100 | 2,100 | 2,100 |
| Observations (number of drivers × number of days) | 2,100 | 10,500 | 10,500 | 10,500 | 2,100 | 10,500 | 10,500 | 10,500 |

*Notes.* Standard errors in parentheses are clustered at the driver level. False discovery rate adjusted $q$-values are in square brackets to correct for multiple hypothesis testing.
$^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$.

**Table 6.** Average Treatment Effects on Daily Revenue: Difference-in-Differences Linear Regressions

| | Dependent variable: change in daily revenue | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Time period | (1) First day of contest | (2) During contest | (3) Two weeks postcontest | (4) Four weeks postcontest | (5) First day of contest | (6) During contest | (7) Two weeks postcontest | (8) Four weeks postcontest |
| Treated | 38.69** | 35.24*** | 17.36* | 6.369 | 38.99** | 35.39*** | 17.62** | 6.308 |
| | (16.61) | (9.246) | (9.534) | (9.512) | (16.58) | (6.984) | (7.111) | (6.752) |
| | [0.032] | [0.001] | [0.092] | [0.504] | [0.032] | [0.001] | [0.032] | [0.401] |
| Age | | | | | 1.642** | 0.741** | 0.801** | 0.829** |
| | | | | | (0.836) | (0.352) | (0.359) | (0.341) |
| Platform age, year | | | | | 7.620 | 17.16*** | 18.19*** | 6.455 |
| | | | | | (10.83) | (4.561) | (4.644) | (4.410) |
| Local | | | | | 26.98* | 14.23** | 3.998 | 24.99*** |
| | | | | | (14.24) | (5.999) | (6.108) | (5.800) |
| Male | | | | | 91.16** | 35.42** | 27.18 | 34.82** |
| | | | | | (42.30) | (17.82) | (18.15) | (17.23) |
| Constant | −43.98*** | −24.24*** | −66.96*** | −82.06*** | −204.4*** | −103.4*** | −138.6*** | −157.0*** |
| | (15.16) | (8.440) | (8.703) | (8.683) | (53.59) | (22.58) | (22.99) | (21.83) |
| Number of drivers | 2,100 | 2,100 | 2,100 | 2,100 | 2,100 | 2,100 | 2,100 | 2,100 |
| Observations (number of drivers × number of days) | 2,100 | 10,500 | 10,500 | 10,500 | 2,100 | 10,500 | 10,500 | 10,500 |

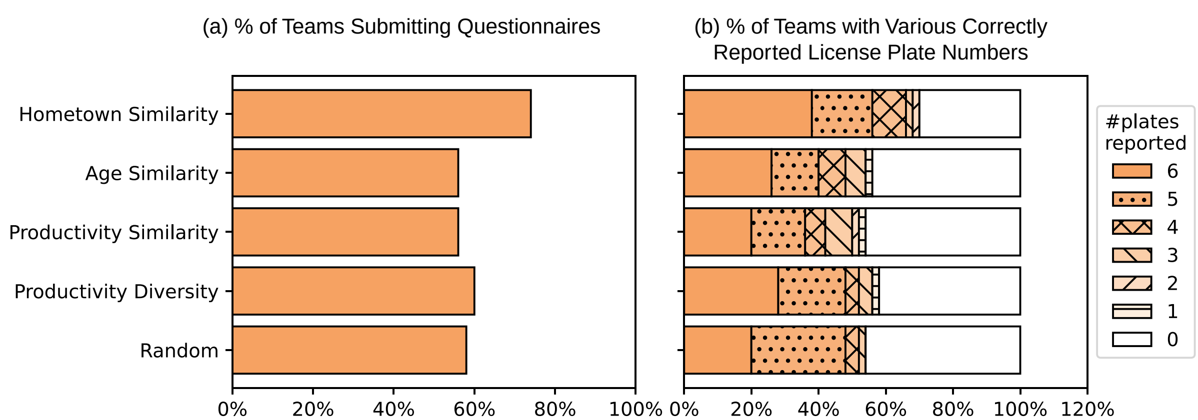
*Notes.* Standard errors in parentheses are clustered at the driver level. False discovery rate adjusted $q$-values are in square brackets to correct for multiple hypothesis testing.
*$p < 0.1$; **$p < 0.05$; ***$p < 0.01$.

in revenue because they have higher earnings expectations.

In what follows, we provide some evidence that our treatment effect is not entirely driven by an incentive to compete for monetary prizes. First, treated drivers continue to work longer hours, complete more trips, and generate higher revenue in the two weeks postcontest, absent any monetary prize or formal competition during this period (column (7) in Tables 4–6). Furthermore, our postcontest survey (Section EC.7) indicates that more than 88% of the drivers either like or very much like the team contest (question 2), citing team belonging (66%), making friends (70%), a sense of honor from winning (61%), and monetary incentives (68%) as the top benefits. When asked whether they prefer a temporary or a long-lasting team (question 12), 79% of the drivers choose "a long-lasting team, so team members can keep in touch after the contest." The long-lasting bonds among team members are further corroborated in a postexperiment interview with 14 drivers conducted by the platform staff and the first author, in which drivers mention finding friends from their hometown as one of the top benefits of the contest. They also mention that they continue to socialize with teammates.

As the experimental economics literature provides extensive evidence on the positive effects of communication on team performance, we next examine treatment

**Figure 3.** (Color online) Team Responsiveness in Different Treatments




*Notes.* Team responsiveness is coded based on the precontest survey. Panel (a) codes the responsiveness binarily with a team deemed responsive if the captain submits the survey on team member characteristics. Panel (b) codes responsiveness based on the number of correctly reported license plate numbers.

effects on team responsiveness. Recall that a team is responsive if the captain submits the precontest survey, a measure of likely communication among team members. Figure 3 reports the proportion of responsive teams using both a binary variable to indicate whether a captain submits the survey (left panel) and the number of correctly reported license plate numbers (right panel). From the left panel, we see that teams comprising members from the same hometown exhibit the highest proportion of responsiveness. This result is consistent with prior research that shows location similarity is a strong predictor of whether a member of an online community joins a team (Ai et al. 2016).

Table 7 reports the corresponding regression analysis results for the team formation algorithms and team responsiveness. The omitted group is randomly formed teams. Specifications (1) and (2) use a probit regression to examine the treatment effect along the extensive margin with the likelihood of submitting the survey as the dependent variable. By contrast, specifications (3) and (4) use an ordinary least squares (OLS) regression to examine the treatment effect along the intensive margin with the number of license plates reported correctly as the dependent variable. The results again show that teams based on hometown similarity show the highest level of responsiveness. Quantitatively, these teams are 17 (18) percentage points marginally more likely to be responsive than randomly formed (age- or productivity-similar) teams ($p < 0.10$ in all cases). In comparison, none of the other team-formation algorithms performs better than the randomly formed teams ($p > 0.10$). Along the intensive margin, we find no significant differences among any of the team formations across teams whose captains have submitted the survey. One possible reason for this finding may be that captains submit their surveys only if they had sufficient information.[12] Indeed, more than 75% of the captains who submit the survey correctly report at least five out of six license plate numbers. Whereas the team formation algorithms are not revealed to captains or drivers, we expect that people from the same hometown can infer this similarity from information cues such as a similar accent.[13] We now formally state our second result.

**Result 2** (Team Responsiveness). Hometown-similar teams are 17 (18) percentage points marginally more likely to be responsive than randomly formed (age- or productivity-similar) teams.

**Table 7.** Treatment Effects on Team Responsiveness Omitting the Random Group

| | Extensive margin | | Intensive margin | |
|---|---|---|---|---|
| | Probit, $Y = P$(Responsive) | | OLS, $Y$ = # Correct Plate Numbers | |
| | (1) | (2) | (3) | (4) |
| Hometown similarity | 0.167* | 0.165* | 0.111 | 0.146 |
| | (0.0973) | (0.0967) | (0.375) | (0.376) |
| | [0.353] | [0.353] | | |
| Age similarity | −0.0193 | −0.0172 | 0.102 | 0.0980 |
| | (0.0953) | (0.0954) | (0.400) | (0.406) |
| | [0.917] | [0.917] | | |
| Productivity Similarity | −0.0193 | −0.0101 | −0.326 | −0.256 |
| | (0.0953) | (0.0958) | (0.400) | (0.403) |
| | [0.917] | [0.917] | | |
| Productivity diversity | 0.0195 | 0.0187 | 0.105 | 0.154 |
| | (0.0957) | (0.0951) | (0.393) | (0.399) |
| | [0.917] | [0.917] | | |
| Average precontest revenue, 100 CNY | | 0.0866 | | 0.176 |
| | | (0.0591) | | (0.237) |
| Average age | | 0.00452 | | 0.0134 |
| | | (0.00766) | | (0.0305) |
| Average platform age, year | | −0.0677 | | −0.502 |
| | | (0.140) | | (0.554) |
| Proportion of local driver | | −0.00936 | | −0.501 |
| | | (0.135) | | (0.505) |
| Proportion of male driver | | 0.0205 | | −3.080 |
| | | (0.522) | | (2.008) |
| Constant | | | 4.862*** | 7.455*** |
| | | | (0.280) | (2.332) |
| Observations (number of teams) | 250 | 250 | 152 | 152 |

*Notes.* Standard errors are in parentheses. False discovery rate adjusted $q$-values are in square brackets to correct for multiple hypothesis testing. Coefficients in columns (1) and (2) report the average marginal effects of the probit estimates.
*$p < 0.1$; **$p < 0.05$; ***$p < 0.01$.

By Result 2, we provide some support in favor of Hypothesis 2 that drivers in teams based on strong and salient identities, such as hometown similarity, perform better than those based on weaker identities, such as randomly formed teams.

We next conduct a two-stage least squares instrumental variable regression analysis to establish any causal effect of communication on team performance. Because team responsiveness is not randomly assigned, we instead use random placement into hometown-similar teams as our instrument. The results in Table EC.8 (specifications (1)–(5)) and corresponding OLS results (specifications (6)–(9)) show that responsive team drivers earn 53.5 CNY more than those in a nonresponsive team ($p < 0.01$, column (7)), whereas the local average treatment effect is not statistically significant ($p > 0.10$, columns (2)–(5)).

In our experiment, we are also interested in whether different team formations impact revenue. Table 8 presents our results using team formation algorithm as the independent variable in specifications (1)–(3). As a robustness check, we use alternative measures of team diversity as the independent variables in specifications (4)–(6). More specifically, we measure driver diversity based on the standard deviation in driver age, productivity, and platform age within a team; we measure hometown diversity using the average distance (in kilometers) between the hometowns of any two drivers within the same team. Our results in Table 8 show that, irrespective of our independent variables, team formation has no significant effect on driver revenue during the contest. Interestingly though, we find that teams based on age similarity exhibit significantly higher revenue during the two-week period after the contest, earning 33 CNY more, on average, compared with drivers in randomly formed teams (column (2), $p < 0.05$). This observation is confirmed by the negative correlation between the standard deviations of age and team productivity (column (5), $p < 0.05$). Furthermore, we see that productivity-similar teams exhibit marginally higher revenue during the two-week period after the contest, earning 21 CNY more, on average, compared with drivers in randomly formed teams (column (2), $p < 0.10$).[14] We summarize our analysis as follows.

**Result 3** (Team Composition). Whereas team formation has no significant effect on driver revenue during the contest, age- (productivity-) similar teams are

**Table 8.** Similarity and Diversity on Driver Revenue: Difference-in-Differences Linear Regressions on Treated Drivers

| | Dependent variable: change in daily revenue, CNY | | | | | |
|---|---|---|---|---|---|---|
| | By treatment group | | | By diversity metrics | | |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Time period | Contest | Two weeks postcontest | Four weeks postcontest | Contest | Two weeks postcontest | Four weeks postcontest |
| Age similarity | 0.933 | 33.19** | 9.806 | | | |
| | (15.10) | (12.93) | (10.76) | | | |
| | [0.951] | [0.124] | [0.445] | | | |
| Hometown similarity | 5.838 | 20.70 | 17.12 | | | |
| | (16.21) | (13.28) | (12.91) | | | |
| | [0.785] | [0.439] | [0.439] | | | |
| Productivity similarity | −14.65 | 21.47* | 13.85 | | | |
| | (14.95) | (12.36) | (12.18) | | | |
| | [0.445] | [0.439] | [0.439] | | | |
| Productivity diversity | −17.50 | 17.50 | 11.33 | | | |
| | (15.36) | (13.21) | (12.66) | | | |
| | [0.439] | [0.439] | [0.445] | | | |
| Age standard deviation | | | | −0.417 | −3.357** | −0.123 |
| | | | | (1.581) | (1.386) | (1.214) |
| Average hometown distance | | | | 0.0297 | −0.00706 | −0.0196 |
| | | | | (0.0237) | (0.0232) | (0.0197) |
| Productivity standard deviation | | | | 0.0953 | −0.0347 | −0.00401 |
| | | | | (0.111) | (0.0951) | (0.0956) |
| Platform age standard deviation | | | | −0.0646 | −0.0370 | −0.0852 |
| | | | | (0.0984) | (0.0840) | (0.0859) |
| Constant | 16.07 | −68.17*** | −86.12*** | 4.701 | −15.89 | −48.15** |
| | (12.04) | (9.269) | (8.033) | (29.33) | (22.64) | (21.99) |
| Number of drivers | 1,750 | 1,750 | 1,750 | 1,750 | 1,750 | 1,750 |
| Observations | 8,750 | 8,750 | 8,750 | 8,750 | 8,750 | 8,750 |

*Notes.* For specifications (1)–(3), the omitted category is the random treatment. Standard errors in parentheses are clustered at the team level. False discovery rate adjusted $q$-values are in square brackets to correct for multiple hypothesis testing.
*$p < 0.1$; **$p < 0.05$; ***$p < 0.01$.

more (marginally more) productive than randomly formed teams two weeks postcontest.

Result 3 yields partial support for Hypothesis 2 that drivers in age- (productivity-) similar teams generate higher (marginally higher) revenue than those in randomly formed teams, albeit only during the two weeks postcontest. We conjecture that the null effect of team formation during the contest might be because drivers may be unable to bond sufficiently during the five-day contest period. This conjecture is corroborated by our postcontest interviews. Drivers tell us that they continue to interact with their teammates even after the end of the contest period. For example, one driver states, "We did not dissolve our team after this campaign. We still chat every day in our group chat [in WeChat]. Indeed, the contest has ended, but our friendship has just started." Another driver mentions "an in-person get-together to celebrate [the end of] the contest."

Next, we examine the effects of prize structure on driver revenue. Table EC.10 presents the results of pairwise comparisons between the three prize structures under each team formation algorithm. The outcome variable is the difference in driver revenue, which is the same as the change in daily revenue (CNY) in Table 8. Based on the multiplicity-adjusted $p$-values (List et al. 2019), we see that variations in the prize structure have a significant effect on changes in revenue only for teams formed based on hometown similarity. Specifically, the group prize treatment leads to a greater increase in daily revenue than either the hybrid (68.4 CNY, multiplicity-adjusted $p = 0.0003$) or individual (41.8 CNY, multiplicity-adjusted $p = 0.062$) prize treatment, thus providing some support for Hypothesis 3. However, these results are not robust to either excluding the 11% of team pairs from different team formation algorithms (Table EC.11) or using a regression analysis interacting prize structure and team formation (Table EC.12). Therefore, we conclude that our prize structure has no effect on aggregate team performance in our experiment.

Finally, we are interested in the productivity of those who volunteer to be captains in our study. Table EC.13 presents a probit specification. The dependent variable is whether a driver volunteers to be a team captain, whereas the independent variables include a driver's precontest revenue, work experience (platform age), and demographics. The results show that those who generate higher revenue in the two weeks prior to the announcement of the contest (0.0253, $p < 0.01$) as well as those who work for the platform for a longer period of time (0.0297, $p < 0.05$) are significantly more likely to volunteer to lead a team.

In the 141 teams with two or more drivers who express interest in being a captain, only one of these drivers is randomly appointed as the team captain. Kolmogorov–Smirnov tests find no significant difference in prior revenue, age, platform age, or gender between the volunteers who are appointed as captains and those who are not ($p > 0.10$). Table 9 presents nine OLS specifications investigating the effects of being randomly selected as a team captain on a driver's daily working hours (specifications (1)–(3)), number of trips (specifications (4)–(6)), and revenue (specifications (7)–(9)). We find that, among our 298 volunteers, those who are randomly chosen to be a captain work harder than those who are not chosen, earning 34 CNY more per day, on average, during contest days although this result is only marginally significant ($p < 0.10$) likely because of the small number of observations. The results in Table EC.14 show that this effect is driven by the lower number of hours worked by volunteers who are not chosen to be captains. Compared with nonvolunteers, volunteers who are not chosen as captains work 32 fewer minutes per day during the contest ($-0.537$, $p < 0.10$, column (4)), whereas those who are chosen do not differ from nonvolunteers in their working hours (0.134, $p > 0.10$, column (4)). Therefore, this effect is driven by the lottery losers being discouraged. We summarize this analysis as follows.

**Table 9.** Effect of Being Randomly Chosen as a Captain: Difference-in-Differences Linear Regressions

| | Dependent variable | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Change in daily working hours | | | Change in daily number of trips | | | Change in daily revenue, CNY | | |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
| Time period | Contest | Two weeks postcontest | Four weeks postcontest | Contest | Two weeks postcontest | Four weeks postcontest | Contest | Two weeks postcontest | Four weeks postcontest |
| Assigned captain | 0.668* | 0.390 | −0.00657 | 1.663 | 0.595 | −0.321 | 34.18* | 23.65 | −5.278 |
| | (0.376) | (0.400) | (0.424) | (1.043) | (1.097) | (1.079) | (17.73) | (19.29) | (18.30) |
| Constant | −0.380 | −1.342*** | −0.844*** | −1.511** | −3.647*** | −4.167*** | −17.91 | −57.15*** | −65.08*** |
| | (0.258) | (0.275) | (0.291) | (0.717) | (0.754) | (0.742) | (12.20) | (13.27) | (12.59) |
| Number of volunteers | 298 | 298 | 298 | 298 | 298 | 298 | 298 | 298 | 298 |
| Observations | 1,490 | 1,490 | 1,490 | 1,490 | 1,490 | 1,490 | 1,490 | 1,490 | 1,490 |

*Notes.* Standard errors in parentheses are clustered at the driver level.
$^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$.

**Result 4** (Team Captain). Among drivers who volunteer to be team captains, captains work 40 minutes marginally longer and earn 34 CNY marginally more revenue per day during the contest than those not chosen to be captains, who work fewer hours than captains or nonvolunteers.

We next explore whether our team treatment effect on revenue differs for captains and noncaptains by rerunning our analyses excluding team captains from the analysis. The results in Tables EC.15–EC.17 show a similar treatment effect, indicating that the contest benefits less productive drivers as much as it benefits more productive ones.

## 4.2. Contest Dynamics

Drivers in our contest receive feedback on their team and individual performance each day during the contest. Previous research on repeated contests shows that effort and performance in later rounds depends on success or failure in earlier rounds (Descamps et al. 2022). To assess if our drivers' effort is influenced by feedback on their contest performance, we perform a panel analysis that examines the extent to which an individual or team's likelihood of winning on day $t$ is influenced by whether the individual or team won on day $t-1$, where $t \in \{2, 3, 4, 5\}$.

Table 10 presents the results from seven probit specifications. The dependent variable is whether an individual driver wins the battle on day $t$, whereas the independent variables include the indicator variables *Individual Won on Day* $t-1$, *Team Won on Day* $t-1$, the interaction of the two, an individual driver's precontest revenue, and the driver's opponent's precontest revenue. Specifications (1)–(4) include all treated drivers, whereas specifications (5)–(7) repeat the analysis for the respective prize structures. The results show that a driver who wins the individual (team) battle on the previous day is 9.4 pp (10.6 pp) more likely to win again (column (1)). Controlling for the precontest revenue of a driver and opponent as well as the interaction term, the momentum effects decrease slightly to 6.65 pp (7.19 pp) if the individual (team) wins the previous day (column 4). Interestingly, when we examine the momentum effect by prize structure, we find that a driver who is rewarded solely by the outcome of the individual battle increases the driver's likelihood of winning by 9.68 pp (14.9 pp) if the driver (team) wins the previous day. This latter result underscores the positive effect of team identity on individual performance.

To uncover the mechanisms driving our findings, we repeat our dynamic analysis using the number of trips as the dependent variable. The results in Table 11 show that a driver who wins an individual battle on day $t-1$ completes 2.39 extra trips on day $t$ (column (4)). This effect size is robust under different prize structures (columns (5)–(7)).[15]

Finally, Table 12 presents the effects of a team's previous win on the likelihood that the team wins again, controlling for the precontest revenue of a driver's team members as well as that of the opponent team's members. Specifications (1)–(3) include all teams, whereas specifications (4)–(6) include only the 222 matched teams within the same team formation algorithm group. The results show that a team that wins on day $t-1$ is 11.8 pp more likely to win on day $t$ (column (4)). This effect size is reduced to 9.45 pp and is marginally significant when we control for own and opponent team's precontest revenue (column (6)). We summarize the results as follows.

**Table 10.** Contest Dynamics on the Likelihood of Individual Winning: Probit

| | Dependent variable: Individual wins on day $t$ (probit) | | | | | | |
|---|---|---|---|---|---|---|---|
| | All drivers | | | | By prize structure | | |
| | (1) | (2) | (3) | (4) | (5) Individual | (6) Hybrid | (7) Group |
| Individual won on day $t-1$ | 0.0944*** | 0.0917*** | 0.0876*** | 0.0665*** | 0.0968** | 0.0215 | 0.0920** |
| | (0.0223) | (0.0216) | (0.0213) | (0.0256) | (0.0475) | (0.0470) | (0.0374) |
| Team won on day $t-1$ | 0.106*** | 0.101*** | 0.0912*** | 0.0719*** | 0.149*** | 0.0194 | 0.0498 |
| | (0.0158) | (0.0153) | (0.0150) | (0.0192) | (0.0333) | (0.0311) | (0.0329) |
| Individual and team won on day $t-1$ | | | | 0.0409 | 0.0185 | 0.0511 | 0.0496 |
| | | | | (0.0256) | (0.0425) | (0.0499) | (0.0398) |
| Precontest revenue, 100 CNY | | 0.0620*** | 0.0880*** | 0.0879*** | 0.0564*** | 0.1199*** | 0.0885*** |
| | | (0.00655) | (0.00881) | (0.00881) | (0.0152) | (0.0144) | (0.0142) |
| Opponent's precontest revenue, 100 CNY | | | −0.0440*** | −0.0440*** | −0.0180 | −0.0793*** | −0.0369** |
| | | | (0.00933) | (0.00934) | (0.0150) | (0.0169) | (0.0147) |
| Number of drivers | 1,750 | 1,750 | 1,750 | 1,750 | 588 | 574 | 588 |
| Observations (number of drivers × number of days) | 7,000 | 7,000 | 7,000 | 7,000 | 2,352 | 2,296 | 2,352 |

*Notes.* The dependent variable uses outcomes from the second to the fifth contest days. Coefficients are average marginal effects using the delta method. Standard errors in parentheses are clustered at the team level.
*$p < 0.1$; **$p < 0.05$; ***$p < 0.01$.

**Table 11.** Contest Dynamics on the Number of Trips: OLS

| | Dependent variable: Daily number of trips | | | | | | |
|---|---|---|---|---|---|---|---|
| | All drivers | | | | By prize structure | | |
| | (1) | (2) | (3) | (4) | (5) Individual | (6) Hybrid | (7) Group |
| Individual won on day $t-1$ | 2.290*** | 2.563*** | 2.698*** | 2.391*** | 2.189*** | 2.454*** | 2.577*** |
| | (0.297) | (0.294) | (0.296) | (0.392) | (0.734) | (0.727) | (0.588) |
| Team won on day $t-1$ | 0.779** | 0.596** | 0.729** | 0.442 | 1.097 | −0.255 | 0.631 |
| | (0.328) | (0.303) | (0.306) | (0.406) | (0.740) | (0.650) | (0.725) |
| Individual and team won on day $t-1$ | | | | 0.625 | 0.330 | 0.952 | 0.420 |
| | | | | (0.500) | (0.828) | (0.952) | (0.824) |
| Precontest revenue, 100 CNY | | 4.892*** | 4.446*** | 4.442*** | 3.705*** | 4.831*** | 4.805*** |
| | | (0.192) | (0.247) | (0.246) | (0.396) | (0.402) | (0.444) |
| Opponent's precontest revenue, 100 CNY | | | 0.755*** | 0.757*** | 0.932** | 0.412 | 0.912** |
| | | | (0.251) | (0.251) | (0.395) | (0.443) | (0.449) |
| Constant | 14.75*** | 1.855*** | 0.915 | 1.003* | 2.422** | 0.694 | −0.152 |
| | (0.354) | (0.514) | (0.578) | (0.587) | (1.094) | (0.920) | (0.994) |
| Number of drivers | 1,750 | 1,750 | 1,750 | 1,750 | 588 | 574 | 588 |
| Observations (number of drivers × number of days) | 7,000 | 7,000 | 7,000 | 7,000 | 2,352 | 2,296 | 2,352 |

*Notes.* The dependent variable uses outcomes from the second to the fifth contest days. Standard errors in parentheses are clustered at the team level.
*$p < 0.1$; **$p < 0.05$; ***$p < 0.01$.

**Result 5** (Momentum Effects). A driver who wins a pairwise battle on day $t-1$ completes 2.29 more trips and is 9.4 pp more likely to win again on day $t$. In comparison, a driver whose team wins on day $t-1$ is 10.6 pp more likely to win on day $t$. Under the individual prize structure, the individual (team) momentum effect is 9.7 pp (14.9 pp).

By Result 5, we confirm the individual momentum effects that are widely reported in the experimental and empirical contest literature (Descamps et al. 2022). To our best knowledge, ours is the first study to document a team momentum effect.

Finally, to investigate any treatment effects on retention, we examine the fourth week after the experiment, which is also the last seven days in our observation period, and code a driver who completed any ride as active. The results in Table EC.18 show no significant treatment effect on retention (0.00719, $p > 0.10$), likely because of the short intervention duration and the relatively small number of observations. Indeed, in a subsequent three-city team contest experiment with a longer duration and larger sample size, we find that drivers in a team leaderboard treatment continue to work longer hours three months after the intervention ends (Ye et al. 2022).

## 5. Discussion

Our study uses a natural field experiment at a ride-sharing platform in China to understand how team formation and other factors impact team responsiveness, driver working hours, and driver revenue. Applying social identity theory to the ride-sharing context, we use different team-formation algorithms to place drivers in teams and compare the revenue of our treatment and

**Table 12.** Contest Dynamics on the Likelihood of Team Winning: Probit

| | Dependent variable: Team wins on day $t$ (probit) | | | | | |
|---|---|---|---|---|---|---|
| | All teams | | | Team pairs in the same algorithm | | |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Team won on day $t-1$ | 0.108* | 0.0968* | 0.0731 | 0.118** | 0.111** | 0.0945* |
| | (0.0587) | (0.0546) | (0.0462) | (0.0584) | (0.0556) | (0.0488) |
| Team members' average precontest revenue, 100 CNY | | 0.158*** | 0.258*** | | 0.156*** | 0.275*** |
| | | (0.0399) | (0.0375) | | (0.0446) | (0.0428) |
| Opponent team members' average precontest revenue, 100 CNY | | | −0.257*** | | | −0.275*** |
| | | | (0.0378) | | | (0.0431) |
| Number of teams | 250 | 250 | 250 | 222 | 222 | 222 |
| Observations (Number of teams × Number of days) | 1,000 | 1,000 | 1,000 | 888 | 888 | 888 |

*Notes.* The dependent variable uses outcomes from the second to the fifth contest days. Coefficients are average marginal effects using the delta method. Standard errors in parentheses are clustered at the team level.
*$p < 0.1$; **$p < 0.05$; ***$p < 0.01$.

control drivers during and after a contest. Our results show that, compared with those in the control condition, treated drivers work 0.8 longer hours, complete 2.4 more trips, and earn 12% higher revenue per day during the contest. This treatment effects persist two weeks after the contest, albeit with half of the effect size. Furthermore, we find that drivers in teams composed of drivers with similar age continue to work longer hours and generate higher revenue during the two weeks after the contest, absent any cash prize or formal competition.

A question for platforms interested in implementing a team contest structure is whether doing so yields a positive return on investment (ROI). Based on the average treatment effects for all treated drivers during the contest and the two-week postcontest period, we find a total revenue increase of 612,150 CNY. The cash reward for winners as well as the payment for team captains sums to 163,850 CNY, yielding a total ROI of 3.74 and a platform ROI (19% of revenue) of 0.71 for every dollar invested.[16] Despite this platform financial loss, the company was motivated by the driver engagement benefits in our study to consider a version of our team contest as part of its organizational structure.

Indeed, encouraged by the results of our experiment, the platform shipped two of our team-formation algorithms (hometown and age similarity) into production within the platform. To address the financial loss issue in our experiment, the platform adapted our contest format from daily prizes into a final prize for the entire duration of a contest. In 2018 alone, the platform conducted 1,548 team contests across 180 cities in China, involving more than two million drivers. These contests, typically one week long, helped the platform meet the high tourist demand during national holidays and increased both driver income and retention (Ye et al. 2020). In a subsequent three-city team contest experiment, we find that team and individual ranking information alone can increase driver revenue by 2%–6% and drivers in the team leaderboard condition continue to be more engaged with the platform three months after the contest ends (Ye et al. 2022). These experiments demonstrate the efficacy of teams and helped convince the platform to make driver teams a permanent feature of the platform structure.

Whereas our experiment examines the effect of team formation on one platform, our results indicate that team identity shows great promise as a design tool that can be leveraged to increase worker productivity and engagement in the gig economy. Future research could use our study as a foundation for exploring the full potential of social identity theory, examining the impact of longer contests and more persistent teams.

## Acknowledgments

The authors thank George Akerlof, Roland Bénabou, Ceren Budak, Alain Cohn, Donna Harris, Rachel Kranton, Erin Krupka, Steve Leider, Jingfeng Lu, Dilip Mookherjee, Daniel Romero, and Tanya Rosenblat as well as seminar and conference participants at the Monash University Prato Center, University of Michigan, the Economics Research Network on Identity, Norms and Narratives (2019, London, UK), Monash Conference on Social Influence, the 2019 Economic Science Association meetings (Los Angeles, CA), and the 2020 North America Economic Science Association meetings (online) for helpful comments. The authors thank Miao Liang, Mei Mei, Tao Song, Guobin Wu, and Lulu Zhang for their help in implementing the experiment. Three anonymous referees and an associate editor provided insightful comments which significantly improved the paper. Anonymized and deidentified data are available on a secure server located in the platform headquarters in Beijing to any researcher for purposes of reproducing or extending the analysis. The platform contact email is gaia@didichuxing.com.

## Endnotes

[1] Statistics are from the company website, retrieved July 29, 2022.

[2] Using this particular data set for our power calculation is based on the availability of processed data during the experiment-design phase.

[3] Contest rules are announced after the team-formation process and before the start of the contest. See Section EC.1 for recruiting materials. Around the time of our experiment, the exchange rate was US$1 ≈ 6.7 CNY.

[4] Similarity is based on precontest productivity using the number of trips in a two-week window and ensuring additional treatment-specific criteria. For example, if the dropout driver is assigned to a team based on hometown similarity, we require that the new driver is from the same province. If the dropout driver is assigned to a team based on age similarity, we require that the substitute driver is in the same age group. We present the results for the treated drivers in the main text and repeat our analyses with the originally randomized drivers (including the 15 dropouts) in Section EC.6.

[5] WeChat is the dominant social media and communication app in China, which allows for both group texts and calls.

[6] We thank an anonymous reviewer for suggesting this approach.

[7] The figure looks almost identical if we replace daily revenue with the number of hours worked or the number of completed trips.

[8] In the subsequent analysis, we focus on revenue comparisons with the understanding that alternative comparisons using driver income or platform profit reach similar conclusions.

[9] This high attrition rate also appears to be a problem with Uber drivers (Scheiber 2017).

[10] To check the robustness of Result 1, we repeat the analysis with the original drivers, using the 15 dropouts rather than their replacements, and find similar results (see Tables EC.2–EC.4).

[11] This difference is modeled in Equation (1).

[12] To check the robustness of our results, we repeat the analysis with productivity similarity as the omitted group, using treated drivers (Table EC.6) or the original randomly assigned drivers, including the 15 dropouts, and find similar results (Table EC.7).

[13] During our contest period, each province imposes a roaming charge for mobile phones from other provinces. Because most people change their phone numbers when they move to a new province, drivers cannot infer others' hometowns from their phone numbers.

[14] To check robustness, we repeat the analysis using the original randomly assigned drivers, including the 15 dropouts, and find similar results (Table EC.9, Section EC.6.).

[15] We repeat the analysis using revenue as the outcome variable in Table EC.19 and find similar results.

[16] The average treatment effect for treated drivers is 35.24 CNY per day during the contest period (five days in total) and 17.36 CNY during the two-week post contest period (10 days in total). Therefore, the total revenue increase during both periods is $(35.24 \times 5 + 17.36 \times 10) \times 1{,}750 = 612{,}150$. The expected cash reward is 15 CNY per driver per day, plus 100 CNY per team captain and 50 CNY for the 152 captains who submit the precontest survey. Therefore, the total cost is $15 \times 1{,}750 \times 5 + 100 \times 250 + 50 \times 152 = 1{,}63{,}850$. This yields an ROI = 612,150/163,85 = 3.74.

## References

Ai W, Chen R, Chen Y, Mei Q, Phillips W (2016) Recommending teams promotes prosocial lending in online microfinance. *Proc. Natl. Acad. Sci. USA* 113(52):14944–14948.

Akerlof GA, Kranton RE (2000) Economics and identity. *Quart. J. Econom.* 115(3):715–753.

Akerlof GA, Kranton RE (2010) *Identity Economics: How Our Identities Shape Our Work, Wages, and Well-Being* (Princeton University Press, Princeton, NJ).

Anderson ML (2008) Multiple inference and gender differences in the effects of early intervention: A reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects. *J. Amer. Statist. Assoc.* 103(484):1481–1495.

Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. Roy. Statist. Soc. B* 57(1):289–300.

Brewer MB (1999) The psychology of prejudice: Ingroup love and outgroup hate? *J. Soc. Issues* 55(3):429–444.

Brewer MB, Silver MD (2000) Group distinctiveness, social identification, and collective mobilization. Stryker S, Owens TJ, White RW, eds. *Self, Identity and Social Movements*, chapter 7 (University of Minnesota Press, Minneapolis, MN), 153–171.

Cason TN, Sheremeta RM, Zhang J (2012) Communication and efficiency in competitive coordination games. *Games Econom. Behav.* 76(1):26–43.

Charness G, Chen Y (2020) Social identity, group behavior and teams. *Annual Rev. Econom.* 12:691–713.

Charness G, Rigotti L, Rustichini A (2007) Individual behavior and group membership. *Amer. Econom. Rev.* 97(4):1340–1352.

Chen R, Chen Y (2011) The potential of social identity for equilibrium selection. *Amer. Econom. Rev.* 101(6):2562–2589.

Chen Y, Li SX (2009) Group identity and social preferences. *Amer. Econom. Rev.* 99(1):431–457.

Chen R, Chen Y, Riyanto YE (2020) Best practices in replication: A case study of common information in coordination games. *Experiment. Econom.* 24:2–30.

Chen Y, Li SX, Liu TX, Shih M (2014) Which hat to wear? Impact of natural identities on coordination and cooperation. *Games Econom. Behav.* 84:58–86.

Chen MK, Chevalier JA, Rossi PE, Oehlsen E (2019) The value of flexible work: Evidence from Uber drivers. *J. Political Econom.* 127(6):2735–2794.

Cohen P, Hahn R, Hall J, Levitt S, Metcalfe R (2016) Using big data to estimate consumer surplus: The case of Uber. NBER Working Paper No. 22627, National Bureau of Economic Research, Cambridge, MA.

Cook C, Diamond R, Hall J, List JA, Oyer P (2018) The gender earnings gap in the gig economy: Evidence from over a million rideshare drivers. NBER Working Paper No. 24732, National Bureau of Economic Research, Cambridge, MA.

Descamps A, Ke C, Page L (2022) How success breeds success. *Quant. Econom.* 13(1):355–385.

Eckel CC, Grossman PJ (2005) Managing diversity by creating team identity. *J. Econom. Behav. Organ.* 58(3):371–392.

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

Erev I, Bornstein G, Galili R (1993) Constructive intergroup competition as a solution to the free rider problem: A field experiment. *J. Experiment. Soc. Psych.* 29(6):463–478.

Fu Q, Lu J, Pan Y (2015) Team contests with multiple pairwise battles. *Amer. Econom. Rev.* 105(7):2120–2140.

Gee LK, Schreck MJ, Singh A (2020) From laboratory to field: Social distance and charitable giving in teams. *Econom. Lett.* 192: 109128.

Girvan M, Newman MEJ (2002) Community structure in social and biological networks. *Proc. Natl. Acad. Sci. USA* 99(12):7821–7826.

Goette L, Huffman D, Meier S, Sutter M (2012) Competition between organizational groups: Its impact on altruistic and antisocial motivations. *Management Sci.* 58(5):948–960.

Hall JV, Krueger AB (2018) An analysis of the labor market for Uber's driver-partners in the United States. *ILR Rev.* 71(3):705–732.

Harmel R, Yeh YY (2015) China's age cohorts: Differences in political attitudes and behavior. *Soc. Sci. Quart.* 96(1):214–234.

Heller N (2017) Is the gig economy working? *The New Yorker* (May 8). Accessed December 7, 2022, https://www.newyorker.com/magazine/2017/05/15/is-the-gig-economy-working.

Jackson E, Looney A, Ramnath S (2017) The rise of alternative work arrangements: Evidence and implications for tax filing and benefit coverage. Office of Tax Analysis Working Paper 114. Accessed November 1, 2022, https://www.census.gov/content/dam/Census/about/about-the-bureau/adrm/FESAC/meetings/Ramnath%20Background%20Document.pdf.

Kessler JB, Milkman KL (2018) Identity in charitable giving. *Management Sci.* 64(2):845–859.

Kim PH, Aldrich HE (2002) Teams that work together, stay together: Resiliency of entrepreneurial teams. *Babson Kauffman Entrepreneurship Res. Conf.*

Konrad KA (2009) *Strategy and Dynamics in Contests* (Oxford University Press, New York).

Krishnan HA, Miller A, Judge WQ (1997) Diversification and top management team complementarity: Is performance improved by merging similar or dissimilar teams? *Strategic Management J.* 18(5):361–374.

Lazear EP, Rosen S (1981) Rank-order tournaments as optimum labor contracts. *J. Political Econom.* 89(5):841–864.

List JA, Sadoff S, Wagner M (2011) So you want to run an experiment, now what? Some simple rules of thumb for optimal experimental design. *Experiment. Econom.* 14:439–457.

List JA, Shaikh AM, Xu Y (2019) Multiple hypothesis testing in experimental economics. *Experiment. Econom.* 22:773–793.

Liu TX, Wan Z, Yang C (2019) The efficiency of a dynamic decentralized two-sided matching market. Preprint, submitted February 19, http://dx.doi.org/10.2139/ssrn.3339394.

McPherson M, Smith-Lovin L, Cook J (2001) Birds of a feather: Homophily in social networks. *Annual Rev. Sociol.* 27:415–444.

Page SE (2007) *The Difference: How the Power of Diversity Creates Better Groups, Firms, Schools, and Societies* (Princeton University Press, Princeton, NJ).

Ravenelle AJ (2019) *Hustle and Gig: Struggling and Surviving in the Sharing Economy*, 1st ed. (University of California Press, Oakland, CA).

Ruef M, Aldrich HE, Carter NM (2003) The structure of founding teams: Homophily, strong ties, and isolation among us entrepreneurs. *Amer. Sociol. Rev.* 68(2):195–222.

Saad L, Wigert B (2021) Remote work persisting and trending permanent. Accessed December 7, 2022, https://news.gallup.com/poll/355907/remote-work-persisting-trending-permanent.aspx.

Scheiber N (2017) How Uber uses psychological tricks to push its drivers' buttons. *The New York Times*. Accessed December 7, 2022, https://www.nytimes.com/interactive/2017/04/02/technology/uber-drivers-psychological-tricks.html.

Sheremeta RM (2018) Behavior in group contests: A review of experimental research. *J. Econom. Surveys* 32(3):683–704.

Tajfel H, Turner J (1979) An integrative theory of intergroup conflict. Worchel S, Austin W, eds. *The Social Psychology of Intergroup Relations* (Brooks/Cole, Monterey, CA), 94–109.

Tajfel H, Billig M, Bundy R, Flament CL (1971) Social categorization and inter-group behavior. *Eur. J. Soc. Psych.* 1(2):149–177.

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

Ye T, Ai W, Chen Y, Mei Q, Ye J, Zhang L (2022) Virtual teams in a gig economy. *Proc. Natl. Acad. Sci. USA* Forthcoming.

Ye T, Ai W, Zhang L, Luo N, Zhang L, Ye J, Mei Q (2020) Predicting individual treatment effects of large-scale team competitions in a ride-sharing economy. *Proc. 26th ACM SIGKDD Internat. Conf. Knowledge Discovery Data Mining*, 2368–2377.

Zhang C, Xie Y (2013) Place of origin and labour market outcomes among migrant workers in urban China. *Urban Stud.* 50(14): 3011–3026.