Teamwork in Contests*

Jorge Lemus[†] Guillermo Marshall[‡]

January 21, 2021

Abstract

The increasing use of online contests by firms has prompted managers to understand the implications of different contest designs. Using data from a popular online contest platform, we study the impact of allowing self-organized teams (one feature of competition design) on outcomes. We find that: (1) teamwork has a large and positive causal impact on performance; (2) multiplayer teams make fewer submissions than single-member teams, yet they perform better. We show that information asymmetries play an important role in team formation: (1) the majority of teams are formed by similarly-ranked players; (2) players wait longer before forming teams when performance feedback is noisier.

Keywords: Contests, teamwork, collaboration, contest design

1 Introduction

The organization of innovation has changed dramatically following the rise of online competitions. Over the last decade, firms and government agencies have sponsored thousands of competitions with large monetary prizes on online platforms. A firm's manager who needs to solve a specific problem finds in an online competition access to an otherwise expensive

*Acknowledgements: We thank workshop and conference participants at the Econometric Society Winter Meeting (University of Nottingham), Pontificia Universidad Católica de Chile, and University of British Columbia (Sauder School of Business) for helpful comments and suggestions. Guillermo Marshall is supported in part by funding from the Social Sciences and Humanities Research Council. All errors are our own.

[†]University of Illinois at Urbana-Champaign, Department of Economics; jalemus@illinois.edu

[‡]University of British Columbia, Sauder School of Business; guillermo.marshall@sauder.ubc.ca

production technology: a large number of capable workers, each with a unique set of skills, who are willing to spend effort to solve a problem. The way in which participants organize is crucial, as teamwork has been shown to improve productivity in other settings (Hamilton et al., 2003; Jones, 2009; Waldinger, 2012; Ahmadpoor and Jones, 2019). Different online platforms have different policies regarding teamwork, so before choosing where to sponsor a competition, a manager needs to understand how different teamwork policies will impact competition outcomes and performance.

In this paper, we empirically study the impact of teamwork on team performance in large online competitions and investigate the impact of contest design on team formation. Our empirical setting is Kaggle (`www.kaggle.com`), the largest platform for hosting online data-science competitions (prediction contests), where players create algorithms to predict the outcome of a random variable conditional on a set of covariates.[1] Kaggle competitions usually last several months, offer large monetary prizes, and attract thousands of participants, who are allowed to make multiple submissions over time.

At least four factors make Kaggle an ideal setting to causally estimate the effect of collaboration on team performance. First, each competition attracts thousands of participants, who are allowed to form multiplayer teams. Second, players *must* have made at least one submission prior to forming a multiplayer team. This allows us to observe the performance of each player before and after they form their multiplayer team. Third, players are allowed to make multiple submissions over time, allowing us to keep track of their performance over time. Fourth, players are given access to data to train their algorithms, and each submission is evaluated based on its out-of-sample performance using a predetermined metric (e.g., root mean squared error).[2] That is, Kaggle offers an objective performance measure, and players have access to a real-time public leaderboard providing noisy feedback about the state of the competition.

Our sample includes 149 Kaggle competitions and offers detailed information about the performance of every submission in a competition, the identity of the player making each submission, and team formation. These data allow us to reconstruct the public leaderboard of each competition at every moment of time as well as the organization of players into teams

[1]For instance, the ride-sharing company Lyft is hosting a competition where participants need to predict the movement of traffic agents around an autonomous vehicle.

[2]Kaggle evaluates each submission on two subsets of the data to obtain two scores. The first is the public score, which is posted on a public leaderboard in real time with the objective of informing (with a certain degree of noise) players about the performance of all participants. The second score is the private score, which is concealed until the end of the competition and is used by Kaggle to determine the winner of the competition. Public and private scores are highly correlated.

in a competition over time, which summarize all the public information available to players.

We identify the impact of collaboration on team performance by exploiting team formation during a competition. We compare the performance of players who form a multiplayer team with that of players who never form a multiplayer team both before and after the team is formed. We use exact matching to compare every multiplayer team with a solo team that behaves and has an identical performance leading up to the time when the multiplayer team is formed. The identification argument is that the performance of the members of both teams (i.e., the multiplayer team and the comparison solo team) would have followed the same trend had the multiplayer team not formed.

We find that collaboration increases a team's performance scores by an average of 0.04 to 0.1 standard deviations, which is roughly equivalent to the median score difference between the winner of a competition and the player ranked in the 40th position. This finding is consistent with evidence in other settings that teamwork and collaboration improve performance (e.g Hamilton et al., 2003; Ahmadpoor and Jones, 2019). When estimating dynamic effects, we find that, prior to forming the multiplayer team, members of the team to be formed perform no differently than the comparison group, but their performance significantly increased shortly after starting to collaborate, and these performance gains persisted over time. We also show that all members of a multiplayer team increase their performance after their merger. These findings suggest the existence of research synergies, which could be achieved by exploiting comparative advantages of team members (see, e.g. Büyükboyaci and Robbett, 2019).

Why do collaboration increase scores? Collaboration could increase the rate or the quality of submissions (or both). We find that multiplayer teams do not make more submissions than individual teams. In fact, we present evidence suggesting that collaboration causes teams, on average, to decrease their number of submissions per week. Thus, multiplayer teams do better with fewer submissions, which suggests that the gains from collaboration do not come from the ability to make more submissions, but rather from the ability to make better submissions. This creates a tradeoff between diversity and performance: Multiplayer teams are desirable for a contest sponsor who values obtaining the best possible solution rather than a diverse set of solutions. By teaming up, players may discard certain approaches that they would have otherwise pursued in solo work.[3]

Although players who collaborate improve their performance, multiplayer teams represent

[3] Receiving thousands of potential solutions is valuable for a sponsor who seeks to procure diverse solutions (Terwiesch and Xu, 2008).

less than 8 percent of all teams. The lack of collaboration is consistent with other findings in the literature and can be attributed to a number of factors. Boudreau et al. (2017), for instance, present evidence of the negative impact of matching frictions on collaboration. In our setting, matching frictions could hinder collaboration because players find it difficult to find a potential partner that speaks or writes code in the same language and has a compatible skill set and personality. Second, there may be concerns about moral hazard in teams (see, e.g. Bonatti and Hörner, 2011; Georgiadis, 2015). Third, asymmetric information could prevent productive partnerships from forming. Lin et al. (2013), for instance, study the role of information asymmetry on impeding lending in online peer-to-peer lending platforms. In our setting, information about the type of a potential partner—a player's ability, commitment to work, or preference over approaches for solving a problem—may prevent partnerships from forming. Fourth, teams may not form due to concerns about how to share the prize.[4] Bikard et al. (2015) study the emergence of collaboration as a trade-off between productive efficiency and credit allocation.

Our findings shed light on how some of the factors described above might hinder teamwork in online competitions. First, we find evidence of assortative matching: multiplayer teams are more likely to form among players who are performing similarly at the time of the merger. Forming teams of players who are similar alleviates asymmetric-information concerns (ability) and also balances the "power dynamics" inside the team. Similar assortative matching arises along the dimensions of performance in past competitions and contributions to the community (e.g., code sharing and message posting on public forums). Second, we exploit variation in the amount of noise in the public leaderboard across competitions to assess the role of incomplete information. We find that collaboration occurs earlier in competitions that provide performance feedback that is less noisy. We interpret this finding as indicative of the information content of signals: fewer signals are needed to overcome information asymmetries when signals are more precise.

Our results have implications for contest design. Broadly speaking, the contest should be designed in such a way that it facilitates the formation of self-organized teams.[5] First, a public leaderboard is vital; it allows players to learn about the performance of prospective partners in the current competition. Second, the leaderboard should be as informative as

[4] Players considering forming a team may not agree on how to distribute a prize if they hold divergent beliefs on their marginal contribution to the team or if they have different levels of risk aversion. For a highly-skilled player who believes that she will win the contest with a high probability, the marginal return of collaboration is low. This player will engage in collaboration only if the team allocates a large share of the prize to her.

[5] Blasco et al. (2013) shows that self-organized teams perform better than randomly-formed teams.

possible.[6] Third, information about past performance should be as informative as possible.[7] Fourth, the platform should provide opportunities to signal skills beyond performance in the current competition. In Kaggle, for example, competitors can analyze data and make the code of that analysis public even if they do not participate in a competition. Fifth, the platform should facilitate the enforcement of prize splits among team members.[8]

To complement our empirical analysis, we contacted a number of Kaggle participants with teamwork experience to inquire about how team formation takes place. In particular, we asked: "How concerned are you that your teammate will not be a good match?". Some of their answers, which we reproduce below (*without edits*), are in line with our findings suggesting the importance of screening potential teammates.

> "I would team up with a person only if I am very sure that I will learn something from that person. I would check that person LinkedIn profile and would also have conversations with that person over call before teaming up. LinkedIn and their previous kaggle work can serve as good indicator. Also during the call, I ask them what have they done so far in the competition. I decide based on the answers which they give to this question"

> "previous experience at kaggle, posts in the current competition, and the current results. Also it is very important if I already participated in another competition with the person. So I know the capabilities of the person, and how hard he/she can work."

> "I just want to team up with someone smart who I'll enjoy collaborating with. If they've done well in other competitions, that's good enough. If they are doing well in the same competition, it could be do to noise."

Related Literature. Jones (2009) documents that teamwork has increased over time as a response to the "burden of knowledge." Teamwork and specialization allow inventors to

[6] The contest designer needs to also consider that players might overfit if the leaderboard is perfectly informative.

[7] Kaggle allocates "medals" based on performance. However, some have questioned the real value of a medal, especially if each member of a multiplayer team gets one regardless of their contribution.

[8] In some competitions, it is up to the winning team to reallocate the prize money among its members. In others, the platform allocates the prize in even shares between the team members unless the team requests an alternative prize distribution. See, e.g., some competitions hosted in the platform DrivenData.org.

cope with an expanding knowledge frontier.[9] Bloom et al. (2020) document that ideas are harder to find, meaning that research effort is rising but research productivity is declining. With respect to the impact of teamwork on outcomes, Ahmadpoor and Jones (2019) find that teamwork has a greater impact than solo work. Azoulay et al. (2010) and Jaravel et al. (2018) uncover long-lasting effects of collaboration and teamwork from exploiting the premature death of high-skilled collaborators and showing that this event has a negative effect on the future performance of the remaining collaborators.

Büyükboyacı and Robbett (2017) and Büyükboyaci and Robbett (2019) study the allocation of effort across two tasks by workers of heterogeneous skills and find evidence of strong productivity gains by exploiting comparative advantages. Girotra et al. (2010) find that it is counterproductive for team members to work together all the time—it is better for them to work independently and then share their ideas. Our results show that the organization within the team allows teams to send fewer submissions that perform better, but we do not observe task-allocation within a team. Regarding team size, LiCalzi and Surucu (2012) model how knowledge diversity in a team affects performance. Wu et al. (2019) use data on academic papers, patents, and software products to show that smaller teams produce more disruptive research, whereas larger teams expand on the existing knowledge. In our data, 80 percent of multiplayer teams are composed of two or three members.

Some articles have also studied teamwork in Kaggle competitions. These articles are mostly descriptive and, in contrast to our work, they do not provide causal estimates of the impact of teamwork on contest outcomes. Dissanayake et al. (2019) document that teams are formed by members who share similar characteristics, although teams with diverse members perform better. Similarly, Dissanayake et al. (2015) find that non-diverse teams perform worse unless most of their members are high-skilled. Wang et al. (2019) discuss repeated participation in Kaggle competitions.

Another literature has examined the level of strategic sophistication of groups versus individuals. Charness and Sutter (2012) review the literature exploring the differences between group and individual decision-making. A striking finding is that group decisions are less likely to be influenced by biases, cognitive limitations, and social considerations. Most of this literature is experimental, including static and dynamic settings (see, e.g., Cooper and Kagel, 2005; Sutter et al., 2013; Müller and Tan, 2013; Feri et al., 2010). These findings could explain why teams perform better than individuals in our setting. In contrast to this literature, our findings stem from competitions with large monetary prizes and thousands of

[9] In Figure A.2 we show that the share of multiplayer teams in our sample that reach the top 100 in a competition has increased over time.

participants rather than experiments.

Finally, there is a theoretical literature on collusion in contests. Alexeev and Leitzel (1991) provide conditions under which cooperation in contests is sustainable. Similarly, Huck et al. (2002) study mergers and collusion in contests. They show that collusion without synergies is profitable for coalitions that approach the size of the grand coalition when collusion is observable and that it is always profitable when collusion is unobservable.

2 Background and Data

2.1 Kaggle Competitions

Kaggle is a platform that hosts online prediction competitions. A prediction competition is one where participants have to make predictions about a random variable (e.g., YouTube sponsored a competition where players had to create an algorithm that could predict video tags for YouTube videos uploaded by users), and the player with the most accurate predictions wins the competition. We focus on *featured* contests, which are hosted by a company (e.g., Expedia, Google), pay a monetary prize, and usually attract significant participation (more on this below). Contests can last several months, and participants can submit multiple submissions over time (though there is a limit on the number of submissions that can be submitted in a given day).

Participants of Kaggle competitions have access to two datasets. The first dataset is the training dataset, which is to be used by the participants to train their algorithms and includes both an outcome variable and covariates. The second dataset is the *test* dataset, which only includes covariates. When making a submission, the player must submit outcome variable predictions for each observation in the test dataset. Kaggle partitions the test dataset in two subsets and evaluates the out-of-sample performance of each submission on these two subsets.[10] The out-of-sample performance score on the first dataset is called the *public* score, which is instantly posted on a public leaderboard containing the public scores of all submissions up until that moment of time.[11] The second score is the *private* score, which is only made public at the end of the competition and is used by Kaggle to determine the

[10]These subsets are fixed and players do not know whether a given observation in the test dataset belongs to the first or second dataset.

[11]The evaluation criterion for the out-of-sample performance of a submission varies across competitions. Examples of evaluation criteria include the root mean squared error or R^2.

competition winner. The public and private scores are highly correlated (the correlation in our sample is 0.99), making the public scores an informative signal about performance.

Players are free to form multiplayer teams (i.e., mergers) subject to some restrictions. First, each member of the new team must have made at least one submission prior to the merger. In our sample, members of two-player teams submitted an average of 16 submissions prior to the time of the merger, rendering this a non-binding constraint in most cases. Second, the cumulative number of submissions by all team members prior to the merger cannot exceed a threshold (i.e., maximum allowed submissions per day times the number of days the competition has been running).

At least three things make Kaggle an ideal setting for the purposes of our study. First, Kaggle provides rich data on a number of large competitions. Second, because these competitions are prediction contests, there is a well-defined and objective measure of performance for each submission based on the submission's out-of-sample performance. Lastly, we observe team formation. These combined factors provide us with a setting to measure the impact of collaboration on team performance in a competitive environment.

2.2 Data and Descriptive Evidence

We use publicly available information on 149 featured competitions hosted by Kaggle.[12] An observation in our dataset is a submission in a contest. For each submission, we observe the timestamp of the submission, an identifier for the player who made the submission, a team identifier, and the public and private scores of the submission. We complement these data with data on team formation, which allows us to observe the exact date when a player joins a team. While each player has a unique identifier, the player's team identifier changes as the player joins a new team. These data allow us to keep track of the performance of a player (or team) during the contest as well as reconstruct both the public and private leaderboard at every instant of time.

Table 1 reports some competition-level summary statistics. The table shows that these competitions offer a monetary prize of \$48,434 (USD) on average, with some competitions offering as much as \$1,200,000. Featured competitions also attract a significant number of participants and submissions. On average, 1,495 teams submit at least one submission, and the competitions receive an average of 24,787 submissions. With respect to the public and private scores of the submissions, we standardize these variables at the competition level

[12] https://www.kaggle.com/kaggle/meta-kaggle

Table 1: Competition-level summary statistics

	Count (1)	Mean (2)	St. Dev. (3)	Min (4)	Max (5)
Total number of submissions	149	24,787.48	32,416.34	139.00	159,810
Total number of teams	149	1,495.04	1,772.96	29	10,450
Total number of players	149	1,587.93	1,881.41	30	11,111
Average team size	149	1.17	0.13	1.01	1.74
Reward quantity (USD)	149	48,434.21	128,676.46	0	1,200,000

Notes: An observation is a competition.

(i.e., the public and private scores have mean 0 and standard deviation 1, respectively) to facilitate comparison across competitions.[13]

Table 2 presents the distribution of team size across competitions. Panel A includes the full sample of teams, whereas Panel B restricts attention to the teams that finish the contest within the top 50 positions of the ranking. Panel A shows that 92 percent of teams have a single member and 97 percent of teams have no more than 2 members. Panel B shows that collaboration is more frequent among top teams, with only 71 percent of these teams having a single member. Figure A.1 in the Online Appendix shows that mergers take place throughout the competition, with the timing of mergers being roughly uniformly distributed over time.

Figure 1 offers a first approximation to measuring the impact of collaboration on team performance. The figure displays the share of multiplayer teams by final ranking. The figure shows that, among teams that finished a competition in the top 30 positions of the final ranking, higher ranked teams were likelier to collaborate.[14] In conjunction with Table 2, the figure shows that top teams were far more likely to collaborate than the average team (recall that only 8 percent of teams in the sample are multiplayer teams). In the rest of the paper, we study whether this positive relationship between collaboration and performance is, in fact, a causal relationship.

[13] Depending on the submission evaluation metric, players are competing to achieve low scores (e.g., root mean squared error) or high scores (e.g., R^2). We transform scores so that higher scores can always be interpreted as better scores.

[14] Figure A.2 in the Online Appendix also shows that collaboration among top teams has increased over time.

Table 2: Distribution of team size across competitions

Number of members	Freq. (1)	Percent (2)	Cumulative (3)
Panel A: All teams			
1	205,193	92.11	92.11
2	10,302	4.62	96.74
3	3,873	1.74	98.48
4	1,799	0.81	99.28
5 or more	1,594	0.72	100.00
Total	222,761	100.00	
Panel B: Top 50 teams			
1	5,260	71.35	71.35
2	984	13.35	84.70
3	503	6.82	91.52
4	273	3.70	95.23
5 or more	352	4.77	100.00
Total	7,372	100.00	

Notes: An observation is a competition–team combination. The top 50 teams are the teams who finished within the first fifty positions of the private leaderboard in each competition.

3 Empirical Strategy

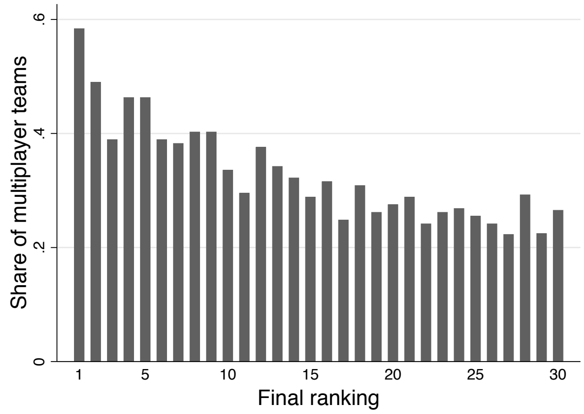
What is the impact of collaboration on player and team performance? To answer this question, we exploit team formation during competitions and compare the performance of the members of a team before and after they merge and start collaborating with the performance of players who never collaborate.

Our main estimating equation is

$$y_{i,j,c,t} = 1\{\text{post-merger}\}_{i,j,c,t}\beta + h(\mathbf{x}_{i,j,c,t}, \delta) + \mu_{j,c} + \lambda_{c,t} + \varepsilon_{i,j,c,t}, \tag{1}$$

where $y_{i,j,c,t}$ is the score of submission i of team (or player) j in competition c at time t, $1\{\text{post-merger}\}_{i,j,c,t}$ is an indicator that takes the value one after the members of team j complete their merger, $\mathbf{x}_{i,j,c,t}$ is a vector of time-varying team-level state variables (e.g., the team's distance to the maximum score on the public leaderboard), $h(\cdot, \delta)$ is a flexible (parametric) function of these state variables, $\mu_{j,c}$ and $\lambda_{c,t}$ are team–competition and competition–time fixed effects, respectively, and $\varepsilon_{i,j,c,t}$ is an error term clustered at the team level. We

Figure 1: How do teams who welcome a new member fair?



Notes: An observation is a team that finished a competition in the top 30 positions of the final ranking.

also estimate a version of equation (1) that allows for time-varying effects,

$$y_{i,j,c,t} = \sum_{\tau=-6}^{6} 1\{\tau \text{ weeks before/after time of merger}\}_{i,j,c,t}\beta_\tau + h(\mathbf{x}_{i,j,c,t}, \delta) + \mu_{j,c} + \lambda_{c,t} + \varepsilon_{i,j,c,t}, \tag{2}$$

where $\beta_{-\tau}$ and β_τ capture, respectively, the performance effects of collaboration τ weeks before and τ weeks after the players actually merge.[15] We note that, in our analysis, *all* the submissions of all members of team j have the same team identifier, even those that are submitted before the players merge. The coefficient of interest, β, therefore measures the impact of collaboration on the overall performance of all team members.

The main identification assumption is that treatment assignment is unconfounded. That is, the probability that a team was exposed to the treatment (i.e., a merger) may depend on team-level state variables ($\mathbf{x}_{i,j,c,t}$) and the team's ability to produce high scores (captured in the team-level fixed effects), but it does not depend on the potential outcomes (Imbens and Rubin, 2015). In our framework, this can also be interpreted as treatment being exogenous conditional on the team-level state variables and the team's ability to produce high scores, thus implying that the treatment is uncorrelated with performance-related unobservables in the error term. Under this assumption, β can be identified by comparing the observed scores of treated and non-treated teams that have similar state variables.

We note that the unconfoundedness assumption accommodates the case in which teams choose to merge based on observable state variables (e.g., their position in the leaderboard)

[15] We normalize the coefficient β_{-1} to be zero. β_0 captures the effect of collaboration at the week of the merger.

or performance-unrelated unobservables (e.g., the size of their social network). This assumption, however, does not accommodate the case in which performance-related unobservables in the error term affect the decision to merge. For example, a violation of the unconfoundedness assumption would occur if all participants had perfect foresight about their collaboration gains and team formation only occurred among those with sufficiently large collaboration gains. Although treatment unconfoundedness is not testable, we can assess the plausibility of this assumption in two ways. First, we use the estimates of equation (2), which allow us to evaluate whether the performance of treated and non-treated teams, conditional on state variables, exhibit similar trends running up to the time of the merger. Second, we present descriptive evidence suggesting that the gains of collaboration are uncertain from the perspective of a participant, which implies that post-merger performance-related unobservables are unlikely to drive team formation.[16]

We estimate equations (1) and (2) in two ways. The first approach makes use of the full sample of teams of up to two members. We exclude larger teams to insulate our estimates of the impacts of collaboration from instances of multiple treatments during the competition (i.e., teams that invite multiple players during the competition and thus experience the benefits of collaboration in multiple different occasions). The second approach makes use of matching on state variables to further restrict the sample. Specifically, we match every two-member team with a non-treated team that has the same state variables at the time of the merger (e.g., the same number of cumulative submissions and distance to the maximum score on the leaderboard). Although all of our specifications control for these state variables, the matched subsample ensures that we are comparing teams that are observationally equivalent except for being exposed to a merger.

Although players are required to submit at least one submission prior to merging with another player, they are not required to make more submissions after merging. We observe 8,466 mergers between two players for which submissions were recorded after the time of the merger. Our matching procedure is able to match 7,474 of these teams with a non-treated team with the same characteristics at the time of the merger (i.e., the same number of cumulative submissions and the same distance to the maximum score on the leaderboard). Table A.1 in the Online Appendix presents a balance analysis for the treated and control teams in the matched subsample.

[16]Another concern is that we may not observe collaboration instances that are informal. That is, players who share information or code but never formally merge. We note that to the extent that collaboration increases performance, not observing these informal arrangements would lead us to underestimate the impact of collaboration on team performance.

4 The Impact of Collaboration on Performance

4.1 Scores

We begin our discussion on the impacts of collaboration by measuring its effect on team performance. Figure 2 presents our estimates for Equation 2, which allows us to measure the performance effects of collaboration starting from 6 weeks prior to the actual merger until 6 weeks after. We conduct the analysis for both the public and private scores, and we make use of two samples. In Panel A, we make use of the full sample of teams of one or two members, which implies that teams with a single member serve as a control for treated teams (i.e., teams with two members). In Panel B, we further restrict the sample so that every treated team (i.e., teams with a merger) is matched with another team with the same covariates at the time of the merger (i.e., the same number of cumulative submissions and the same distance to the maximum score on the leaderboard). All specifications include team fixed effects, competition–day fixed effects, and a second-degree polynomial of a number of team (or player) level state variables.[17] Although the decision to form a team may respond to state variables, which we are flexibly controlling for, our identification assumption is that the decision to form a team does not respond to performance-related unobservables (i.e., treatment assignment is unconfounded).

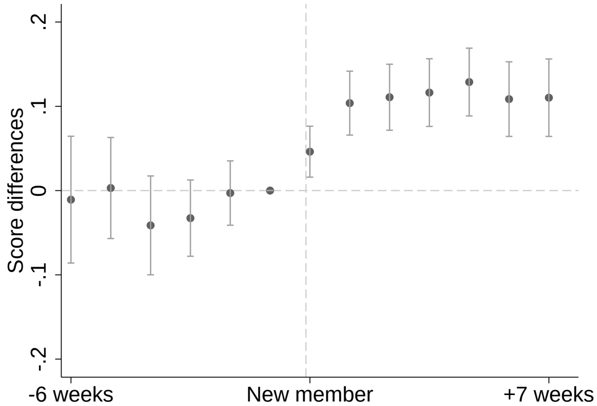
Figure 2 (Panel A) shows that, prior to the actual merger, treated and non-treated teams present no statistical differences in their performance, whether measured using the public or private scores, which provides support for our assumption of unconfounded treatment assignment. After the actual merger, treated teams perform significantly better than non-treated teams, with effects that manifest immediately and last for at least 6 weeks after the merger. In the first week after the merger, the effect is about 0.05 or 0.06 standard deviations, and then it climbs to about 0.11 or 0.12 standard deviations and remains at that level thereafter.[18] Panel B repeats the exercise using the matched subsample. The figures look very similar to those in Panel A, with the exception that the estimated effects are smaller in magnitude than those in Panel A. The smaller magnitudes likely reflect that the control and treatment groups in Panel B are less different in the state variables that predict good performance.

[17] These state variables include the total number of submissions by all teams up until the submission time, total number of submissions by the team making the submission up until the submission time, total number of submissions by the member of the team making the submission up until the submission time, the submitting team's distance to the maximum score on the public leaderboard at the submission time, and the fraction of contest time that had elapsed at the submission time.

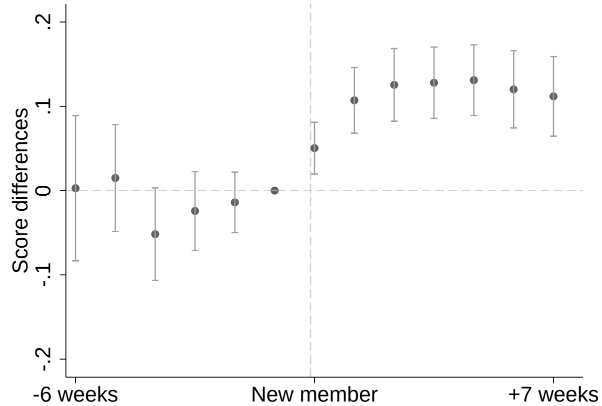
[18] Recall that both public and private scores are standardized (i.e., have mean 0 and standard deviation 1).

Figure 2: The impact of collaboration on scores: Team-level estimates

Panel A: Baseline estimates

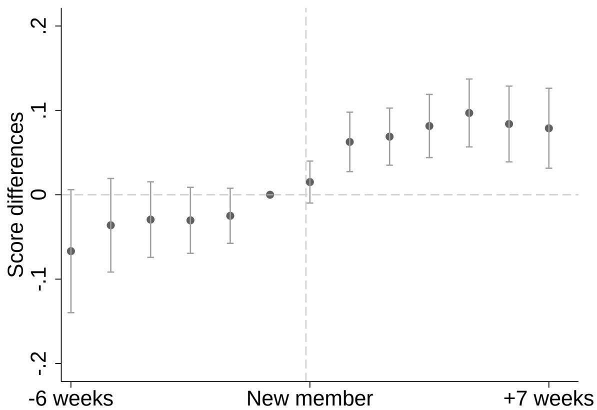


1) Public score

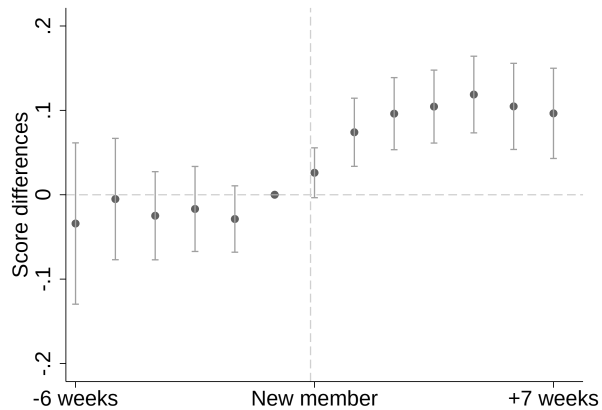


2) Private score

Panel B: Matching estimates



1) Public score



2) Private score

Notes: Standard errors are clustered at the team-level, and 95-percent confidence intervals are depicted in the figures. An observation is a submission made by a team in a competition. All specifications include team fixed effects, competition–day fixed effects, and a second-degree polynomial of variables: total number of submissions by all teams up until the submission time, total number of submissions by the team making the submission up until the submission time, total number of submissions by the member of the team making the submission up until the submission time, the submitting team's distance to the maximum score on the public leaderboard at the submission time, and the fraction of contest time that had elapsed at the submission time. The sample is restricted to include submissions by treated teams that took place six weeks before or after the week in which the team changed its team size, and it also restricts attention to teams with one or two members. Panel B further restricts the sample to ensure balance in observables (measured at the time of treatment).

Table 3 presents estimates for Equation 1, which constrains the treatment effect to be constant in the post-merger period. Panel A shows that collaboration causes an increase in public and

Table 3: The impact of collaboration on scores: Team-level estimates

	Public score (1)	Private score (2)
Panel A: Baseline estimates		
New member	0.078^{***}	0.085^{***}
	(0.014)	(0.016)
Observations	3,248,210	3,179,632
R^2	0.439	0.448
Panel B: Matching estimates		
New member	0.041^{***}	0.050^{***}
	(0.012)	(0.015)
Observations	342,716	338,431
R^2	0.335	0.361

Notes: Standard errors clustered at the team-level in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. An observation is a submission made by a team in a competition. All specifications include team fixed effects, competition–day fixed effects, and a second-degree polynomial of variables: total number of submissions by all teams up until the submission time, total number of submissions by the team making the submission up until the submission time, total number of submissions by the member of the team making the submission up until the submission time, the submitting team's distance to the maximum score on the public leaderboard at the submission time, and the fraction of contest time that had elapsed at the submission time. The sample is restricted to include submissions by treated teams that took place six weeks before or after the week in which the team changed its team size, and it also restricts attention to teams with one or two members. Panel B further restricts the sample to ensure balance in observables (measured at the time of treatment).

private scores of 0.078 and 0.085 standard deviations, respectively. When restricting the sample to the matched subsample, these estimates drop to 0.041 and 0.05, respectively. How large are these magnitudes? The median score difference between the winner of the contest and the player who finishes in the 40th position is about 0.05, which suggests that these collaboration effects are economically significant.

Table 4 replicates Table 3 using indicators for whether a submission has a score that exceeds percentile x of the competition-level score distribution as the dependent variable.[19] The table shows that teamwork has a positive impact on a team's probability of achieving extreme scores, e.g., the probability of achieving a private score that exceeds the 95th and 99th percentile of the distribution increases by 6.6 and 2.3 percentage points on average as a consequence of teamwork. These findings suggest that the performance gains of collaboration are payoff-relevant from the perspective of a team as collaboration impacts their ability to score in the upper tail of the score distribution. Moreover, they suggest that collaboration is

[19]The competition-level score distribution is the final distribution of scores of every competition (i.e., all submissions are used to compute this distribution).

Table 4: The impact of collaboration on extreme scores: Team-level estimates

	$1\{\text{score} > p75\}$	$1\{\text{score} > p90\}$	$1\{\text{score} > p95\}$	$1\{\text{score} > p99\}$
	(1)	(2)	(3)	(4)
Panel A: Public score				
New member	0.074***	0.084***	0.070***	0.023***
	(0.008)	(0.009)	(0.009)	(0.005)
Observations	3,248,210	3,248,210	3,248,210	3,248,210
R^2	0.564	0.570	0.633	0.836
Panel B: Private score				
New member	0.067***	0.077***	0.066***	0.023***
	(0.008)	(0.009)	(0.009)	(0.005)
Observations	3,179,632	3,179,632	3,179,632	3,179,632
R^2	0.541	0.454	0.433	0.453

Notes: Standard errors clustered at the team-level in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. An observation is a submission made by a team in a competition. $1\{\text{score} > pX\}$ is an indicator that takes the value one if the submission's score exceeded percentile X of the competition-level score distribution. All specifications include team fixed effects, competition–day fixed effects, and a second-degree polynomial of variables: total number of submissions by all teams up until the submission time, total number of submissions by the team making the submission up until the submission time, total number of submissions by the member of the team making the submission up until the submission time, the submitting team's distance to the maximum score on the public leaderboard at the submission time, and the fraction of contest time that had elapsed at the submission time. The sample is restricted to include submissions by treated teams that took place six weeks before or after the week in which the team changed its team size, and it also restricts attention to teams with one or two members.

likely to benefit the contest sponsor in the form of a thicker upper tail of scores.

A remaining identification concern is whether players who form multiplayer teams *only* do so because they know that they will benefit from it. In other words, the concern is that the results in Figure 2 and Table 3 are driven by selection along the dimension of performance gains of collaboration. A number of facts render this threat unlikely. First, most players participate in a single competition (70 percent of players), and the overwhelming majority of those who ever form a multiplayer team do so only once (84.3 percent of players). These facts suggest that players must face uncertainty about whether teamwork will be productive for them. In Table A.2 in the Online Appendix, we replicate Table 3, restricting the sample of treated teams to those who are forming a multiplayer team for the first time (i.e., those for whom the benefits of collaboration are more uncertain), and find that our estimates become larger in absolute value, suggesting that this form of selection is not explaining our results. Second, Table A.3 in the Online Appendix shows that the impact of collaboration on performance does not vary by whether the multiplayer team was formed early or late in the competition. If players know that they would benefit from collaboration, one would expect that they form a multiplayer team as early as possible to maximize the benefits of teamwork.[20] In particular, those who expect the greatest benefits of collaboration should form multiplayer teams earlier. We do not see this happening.

4.2 What drives these performance gains?

One possibility is that the original member of the team recruits a "star" and this new member is responsible for the increased performance. A second possibility, which does not preclude the first, is that collaboration creates performance-enhancing synergies, making both players stronger. We can already rule out that the first is the sole explanation of our findings because of our research design—as previously explained, *all* the submissions of all members of team j have the same team identifier, even those that were submitted before the players merged. That is, our estimates measure the impact of collaboration on the overall performance of all team members, so a positive estimated effect of collaboration on scores would necessarily suggest the existence of synergies.

A second question is whether both players improve their performance after they start collaborating. We explore this in Table 5, where we exploit the fact that we observe which player (i.e., the original member or the new member) makes each submission. The table resembles Table 3, but the analysis is now conducted at the player level (i.e., fixed effects are at the player rather than at the team level), and it is run separately for the original and new member of each team. The table shows that collaboration causes an increase in the scores of both the original and new members, though the effects are larger for the new members. These findings suggest that both players become stronger when collaborating and that the performance gains are not simply driven by one of the two.

4.3 Heterogeneity

We next explore whether the impact of collaboration on performance is heterogeneous across different types of teams and contests. In Table A.4 in the Online Appendix, we replicate Table 3 using a subsample of "competitive teams" and find that our estimates do not significantly change as a result (i.e., the point estimates change by less than 10 percent of the

[20] As previously mentioned, Figure A.1 in the Online Appendix shows that mergers occur throughout the competition and are not concentrated at the beginning.

Table 5: Decomposing the effects of collaboration on scores: Player-level estimates

	Public score	Private score
	(1)	(2)
Panel A: Submissions by original members		
New member	0.063***	0.065***
	(0.018)	(0.021)
Observations	3,139,108	3,074,864
R^2	0.445	0.454
Panel B: Submissions by team switchers		
New team	0.092***	0.102***
	(0.016)	(0.016)
Observations	3,530,852	3,455,601
R^2	0.440	0.450

Notes: Standard errors clustered at the player-level in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. An observation is a submission made by a team in a competition. All specifications include player fixed effects, competition–day fixed effects, and a second-degree polynomial of variables: total number of submissions by all teams up until the submission time, total number of submissions by the team making the submission up until the submission time, total number of submissions by the member of the team making the submission up until the submission time, the submitting team's distance to the maximum score on the public leaderboard at the submission time, and the fraction of contest time that had elapsed at the submission time. The sample is restricted to include submissions by treated teams that took place six weeks before or after the week in which the team changed its team size. Panel A is restricted to include submissions made by teams of up to two members and submissions made by the member that started the team (i.e., the "original member"). Panel B restricts the sample of treated teams to include only submissions by players who switched teams during the competition.

standard error).[21] These findings also suggest that our results are not driven by stronger teams. Table A.5 in the Online Appendix explores whether the gains of collaboration are different in contests which are expected to be harder (e.g., contests where players must analyze image data, contests with larger rewards, or contests with larger datasets).[22] We find that the performance gains of collaboration are no different (in statistical terms) in these competitions.

4.4 Number of submissions

We next study the impact of collaboration on the number of submissions by a team. To this end, we compute the number of submissions made by each team in every week of the

[21]A team is classified as competitive if it achieved a score within the top quartile of the competition-level score distribution by the end of the competition.

[22]Lemus and Marshall (2020) present evidence showing that the reward quantity is associated with difficulty.

Table 6: The impact of collaboration on the number of submissions: Team-level estimates

	Number of submissions per week (1)	Number of submissions per week (while active) (2)
Panel A: Baseline estimates		
New member	-1.488***	0.162
	(0.213)	(0.405)
Observations	1,307,553	424,819
R^2	0.657	0.688
Panel B: Matching estimates		
New member	-0.119	0.279
	(0.181)	(0.201)
Observations	75,962	43,690
R^2	0.537	0.573

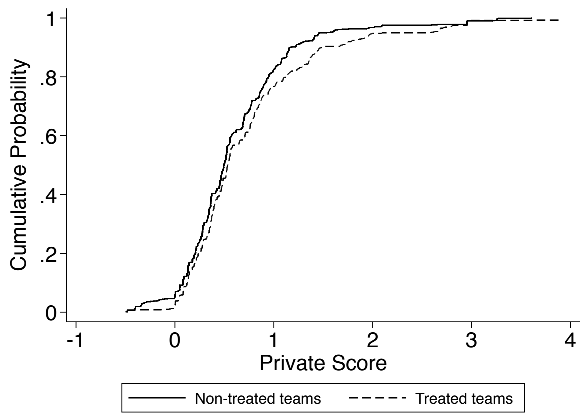
Notes: Standard errors clustered at the team-level in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. An observation is a competition–team–week combination. All specifications include team fixed effects and competition–week fixed effects. The sample is restricted to include submissions that took place in the first twelve weeks of the competition and teams of up to two members. Panel B further restricts the sample to ensure balance in observables (measured at the time of treatment).

competition (dependent variable), and we estimate a version of Equation 1. In the analysis, an observation is a team–week–competition combination. We estimate specifications that differ in how we treat the weeks in which a team makes zero submissions. In the first specification, all team–week combinations are included, whereas in the second we only include the team-week combinations that lie between the first and last week with a positive submission count for that team (i.e., the weeks when the team was active).[23] Both specifications should lead to similar estimates unless treated teams choose to exit the competition sooner or later than non-treated teams.

Table 6 presents the estimates of our analysis. Panel A shows that collaboration causes the number of submissions per week by a team to decrease by 1.5 submissions when including all team–week combinations (Column 1) or to not decrease at all when considering only the active periods of teams (Column 2). These estimates suggest that treated teams may be quitting sooner than non-treated teams but make the same number of submissions as non-treated teams while active. We find similar qualitative effects when looking at the matched subsample (Panel B), but the effects are not statistically different than zero.

[23]For example, if a teams makes 0, 1, 0, 1, and 0 submissions in the five weeks of a competition, respectively, we only include weeks 2, 3 and 4 in the estimation sample. This choice is based on the assumption that the team did not enter until week 2 and was already inactive in week 5.

Figure 3: Distribution of best scores in the competition, by treatment status



Notes: The figure plots the CDF of the 10 best scores in each competition coming from treated and non-treated teams.

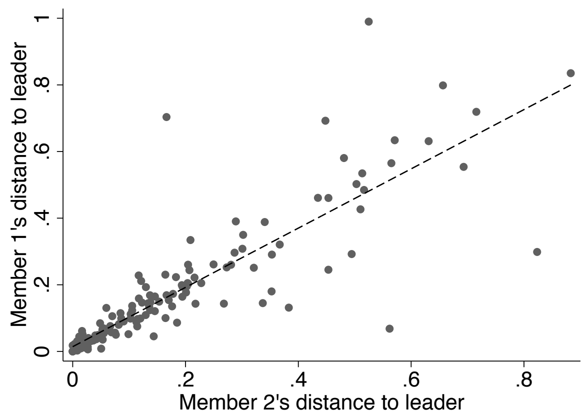
What do these results imply for contest design? The competition sponsor cares about the best submissions in the competition. Allowing players to collaborate and form teams appears to create a tradeoff: teams increase their performance but may send fewer submissions per week. To examine what these countervailing forces do to the upper tail of the score distribution, Figure 3 plots the cumulative distribution function of the 10 best scores in each competition coming from treated and non-treated teams, respectively.

The figure shows that the distribution of top scores of treated teams first order stochastically dominates that of the non-treated teams and has a longer right tail.[24] If we are willing to assume that the distribution of non-treated teams is a good counterfactual for the distribution of treated teams had they not merged, the figure suggests that there is no tradeoff.[25] Even with fewer submissions, treated teams achieve higher scores than non-treated teams, and, since the contest sponsor only cares about the very best submissions, collaboration should be encouraged.

[24]The figure is in line with Table A.4 in the Online Appendix, which replicates our analysis of the impact of collaboration on scores on a subsample of "competitive teams". That analysis shows that collaboration causes an increase in performance even among top teams.

[25]Figure 2 is supportive of this assumption in that treated and non-treated teams do not significantly differ in their performance prior to the time of the merger.

Figure 4: Team member heterogeneity at the time of the merger



Notes: The figure plots the distance of each team member to the leader at the moment of forming a team. These figures restrict attention to teams in which both members had submitted at least 5 submissions prior to the merger.

5 Asymmetric Information: The Need for Screen

Before forming multiplayer teams, players need to assess the value of collaboration. As we mentioned in the introduction, there are many factors that can deter players from collaborating. One of special importance is asymmetric information. Players may be concerned that a potential partner has a low ability. Naturally, players will try to screen potential partners before collaborating. The quotes in the introduction from Kaggle participants suggest that this is, in fact, the case.

There are different ways of screening potential partners. First, players can look at their performance in the current competition. Figure 4 plots each team member's distance to the maximum score on the leaderboard immediately before their merger. The figure shows that mergers largely occur among players who are performing similarly at the time of the merger and usually take place among players who are close to the leader of the competition. This suggests that players choose partners carefully to avoid unproductive mergers and that they want to choose 'someone similar,' possibly to avoid conflicting power dynamics.

While the leaderboard is informative about the performance of potential teammates, the public score is only a noisy signal of the private score (i.e., the payoff-relevant performance measure). We exploit variation in the precision of the public score as a signal of the private score across contests to measure the impact of asymmetric information on team formation. Table 7 shows that both the number of submissions and the time of the merger decrease with

Table 7: The impact of performance feedback noise on team formation outcomes: Player-level estimates

	Number of submissions prior to merger (in logs)	Time of merger (in logs)
	(1)	(2)
Feedback precision (in St. Dev.)	-0.037**	-0.062***
	(0.016)	(0.018)
Observations	4,410	4,410
R^2	0.201	0.043

Notes: Robust standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. An observation is a competition–player combination, and the sample is restricted to players who formed a team. Feedback precision is a measure of feedback precision, which is a contest design feature readily available in the data. We standardize the feedback precision variable to simplify the interpretation of our results (before standardizing, the variable takes value between 0 and 100, with a mean of 31.6 and a standard deviation of 23.4). When the precision is 100 percent, the public and private scores take the same value; when it is 0, the public score is uninformative about the private score (see Section 2). All specifications include contest-level controls (i.e., total reward quantity, number of prizes, maximum daily submissions, contest length, dataset size, image data indicator) and player-level controls (i.e., public score of first submission in the competitions, number of participations in past competitions). Column 1 further controls for the time of the merger. Time of merger is normalized to take values between 0 and 1 (i.e., it can be interpreted as the fraction of the contest time elapsed at the time of the merger).

the precision of the information in the leaderboard. We interpret this finding as indicating the higher information content of more precise signals: When information is more precise, fewer signals are needed to form a more precise posterior belief about the type of a potential teammate, which leads to earlier collaboration.

Second, players could learn about potential teammates by looking at their performance in past competitions. Figure A.3 in the Online Appendix shows that players who have performed well in the past are more likely to form teams with other players who have similarly performed well in the past. However, for first-time Kaggle participants, this information is missing.

Third, players could learn about potential teammates by looking at whether they have shared code with other Kaggle users or have participated on the Kaggle discussion board.[26] Kaggle users can rate both shared code and forum messages, which provides a signal about the expertise or ability of a potential teammate. Figure A.3 in the Online Appendix shows evidence of assortative matching along these dimensions. We find that players who have

[26] Code sharing might be altruistically motivated by the joy of enhancing learning in the community, or strategically motivated to signal skills to other player. Tausczik and Wang (2017) investigate community-level sharing of code in 25 Kaggle contests. They find that 10 percent of users, those doing relatively well but not at the top of the competition, were the most likely to share code. They also find that sharing code improved individual, but not collective performance. Hutter et al. (2011) and Bullinger et al. (2010) also discuss the trade-off between cooperation and competition in online contests.

posted more messages in forums or who have shared code that has been well-received in the community are more likely to work together in a multiplayer team.

These findings suggest that players face asymmetric information when forming teams and respond by acquiring information about potential teammates (e.g., Figure 4) as well as providing additional signals to potential teammates when asymmetric information is more severe (Table 7). Our findings have clear implications for contest design. Collaboration has a positive impact on team performance (even at the upper tail of the distribution of scores), but few players choose to form multiplayer teams, which is due at least in part to asymmetric information. Contest platforms should invest in alleviating these information frictions by making the public score a more precise signal of the private score, facilitating access to information about the past performance of players, implementing peer-review evaluations within multiplayer teams, and creating more instances in which players can communicate with others (e.g., discussion boards).

6 Discussion

We ask whether collaboration has an impact on team performance in contests. Our findings suggest that teamwork has an economically significant effect on team performance that equals the median score difference between the winner of a competition and the player ranked in the 40th position. The effect of collaboration is significant for both high- and low-ranked teams and improves a team's chances of reaching the top positions in a competition.

Although we do not observe the reason why these performance gains exist, we present several pieces of evidence pointing towards synergies. First, all team members improve their scores after a merger. Second, we do not find evidence that collaboration leads to a higher rate of submissions—if anything, we find some evidence showing that collaboration decreases the weekly number of submissions. These together suggest that performance gains do not come from doing more, but rather from doing the same better.

We then examine which players are teaming up and whether players are screening for potential teammates. We find that players who are closer to the top of the competition are more likely to team up. Moreover, we find that players are teaming up with players who are similarly ranked. This suggests that players use the current standing in the competition to screen for potential partners.[27] Players also use information about past performance to

[27] The quotes in the introduction also point in this direction.

screen for potential partners, and players with good past performance tend to team up with other players with good past performance. We also exploit variation in the informativeness of the public leaderboard to study whether more precise feedback facilitates collaboration. Consistent with economic theory, we find that, when signals are more precise, players observe fewer signals before teaming up.

Our results have implications for managers seeking to sponsor competitions in order to solve a particular problem as well as for the design of online platforms. First, our finding that self-organized teams perform better than single-member teams suggests that platforms should encourage the formation of self-organized teams. Second, our findings of assortative matching in team formation suggest that platforms should provide opportunities for players to signal their types. For instance, players can signal their types through their performance in the competition. Thus, reducing information asymmetry through an informative leaderboard may facilitate teamwork, although this needs to be traded off against the risk of overfitting. Also, creating instances for players to signal their type outside competitions is valuable. On Kaggle, players can upload their own datasets, share code to analyze a dataset, and post messages on forums that are rated by other users. Other online contest platforms do not allow for these opportunities, which could be detrimental for collaboration and overall performance, so managers should prefer platforms that permit and encourage teamwork.

7 References

Ahmadpoor, Mohammad and Benjamin F Jones (2019) "Decoding team and individual impact in science and invention," *Proceedings of the National Academy of Sciences*, 116 (28), 13885–13890.

Alexeev, Michael and Jim Leitzel (1991) "Collusion and rent-seeking," *Public Choice*, 69 (3), 241–252.

Azoulay, Pierre, Joshua S Graff Zivin, and Jialan Wang (2010) "Superstar extinction," *The Quarterly Journal of Economics*, 125 (2), 549–589.

Bikard, Michaël, Fiona Murray, and Joshua S Gans (2015) "Exploring trade-offs in the organization of scientific work: Collaboration and scientific reward," *Management science*, 61 (7), 1473–1495.

Blasco, Andrea, Kevin J Boudreau, Karim R Lakhani, Michael Menietti, and Christoph Riedl (2013) "Do Crowds have the Wisdom to Self-Organize?".

Bloom, Nicholas, Charles I Jones, John Van Reenen, and Michael Webb (2020) "Are ideas getting harder to find?" *American Economic Review*, 110 (4), 1104–44.

Bonatti, Alessandro and Johannes Hörner (2011) "Collaborating," *American Economic Review*, 101 (2), 632–63.

Boudreau, Kevin J, Tom Brady, Ina Ganguli, Patrick Gaule, Eva Guinan, Anthony Hollenberg, and Karim R Lakhani (2017) "A field experiment on search costs and the formation of scientific collaborations," *Review of Economics and Statistics*, 99 (4), 565–576.

Bullinger, Angelika C, Anne-Katrin Neyer, Matthias Rass, and Kathrin M Moeslein (2010) "Community-based innovation contests: Where competition meets cooperation," *Creativity and innovation management*, 19 (3), 290–303.

Büyükboyacı, Mürüvvet and Andrea Robbett (2017) "Collaboration and free-riding in team contests," *Labour Economics*, 49, 162–178.

Büyükboyaci, Mürüvvet and Andrea Robbett (2019) "Team formation with complementary skills," *Journal of Economics & Management Strategy*, 28 (4), 713–733.

Charness, Gary and Matthias Sutter (2012) "Groups make better self-interested decisions," *Journal of Economic Perspectives*, 26 (3), 157–76.

Cooper, David J and John H Kagel (2005) "Are two heads better than one? Team versus individual play in signaling games," *American Economic Review*, 95 (3), 477–509.

Dissanayake, Indika, Sridhar Nerur, and Jie Zhang (2019) "Team Formation and Performance in Online Crowdsourcing Competitions: The Role of Homophily and Diversity in Solver Characteristics."

Dissanayake, Indika, Jie Zhang, and Bin Gu (2015) "Task division for team success in crowdsourcing contests: Resource allocation and alignment effects," *Journal of Management Information Systems*, 32 (2), 8–39.

Feri, Francesco, Bernd Irlenbusch, and Matthias Sutter (2010) "Efficiency gains from team-based coordination—large-scale experimental evidence," *American Economic Review*, 100 (4), 1892–1912.

Georgiadis, George (2015) "Projects and team dynamics," *The Review of Economic Studies*, 82 (1), 187–218.

Girotra, Karan, Christian Terwiesch, and Karl T Ulrich (2010) "Idea generation and the quality of the best idea," *Management science*, 56 (4), 591–605.

Hamilton, Barton H, Jack A Nickerson, and Hideo Owan (2003) "Team incentives and worker heterogeneity: An empirical analysis of the impact of teams on productivity and participation," *Journal of political Economy*, 111 (3), 465–497.

Huck, Steffen, Kai A Konrad, and Wieland Müller (2002) "Merger and collusion in contests," *Journal of Institutional and Theoretical Economics (JITE)*, 563–575.

Hutter, Katja, Julia Hautz, Johann Füller, Julia Mueller, and Kurt Matzler (2011) "Communitition: The tension between competition and collaboration in community-based design contests," *Creativity and innovation management*, 20 (1), 3–21.

Imbens, Guido W and Donald B Rubin (2015) *Causal inference in statistics, social, and biomedical sciences*: Cambridge University Press.

Jaravel, Xavier, Neviana Petkova, and Alex Bell (2018) "Team-specific capital and innovation," *American Economic Review*, 108 (4-5), 1034–73.

Jones, Benjamin F (2009) "The burden of knowledge and the "death of the renaissance man": Is innovation getting harder?" *The Review of Economic Studies*, 76 (1), 283–317.

Lemus, Jorge and Guillermo Marshall (2020) "Dynamic Tournament Design: Evidence from Prediction Contests."

LiCalzi, Marco and Oktay Surucu (2012) "The power of diversity over large solution spaces," *Management Science*, 58 (7), 1408–1421.

Lin, Mingfeng, Nagpurnanand R Prabhala, and Siva Viswanathan (2013) "Judging borrowers by the company they keep: Friendship networks and information asymmetry in online peer-to-peer lending," *Management Science*, 59 (1), 17–35.

Müller, Wieland and Fangfang Tan (2013) "Who acts more like a game theorist? Group and individual play in a sequential market game and the effect of the time horizon," *Games and Economic Behavior*, 82, 658–674.

Sutter, Matthias, Simon Czermak, and Francesco Feri (2013) "Strategic sophistication of individuals and teams. Experimental evidence," *European economic review*, 64, 395–410.

Tausczik, Yla and Ping Wang (2017) "To Share, or Not to Share? Community-Level Collaboration in Open Innovation Contests," *Proceedings of the ACM on Human-Computer Interaction*, 1 (CSCW), 1–23.

Terwiesch, Christian and Yi Xu (2008) "Innovation contests, open innovation, and multiagent problem solving," *Management science*, 54 (9), 1529–1543.

Waldinger, Fabian (2012) "Peer effects in science: Evidence from the dismissal of scientists in Nazi Germany," *The Review of Economic Studies*, 79 (2), 838–861.

Wang, Xuan, Hanieh Javadi Khasraghi, and Helmut Schneider (2019) "Towards an Understanding of Participants' Sustained Participation in Crowdsourcing Contests," *Information Systems Management*, 1–14.

Wu, Lingfei, Dashun Wang, and James A Evans (2019) "Large teams develop and small teams disrupt science and technology," *Nature*, 566 (7744), 378–382.

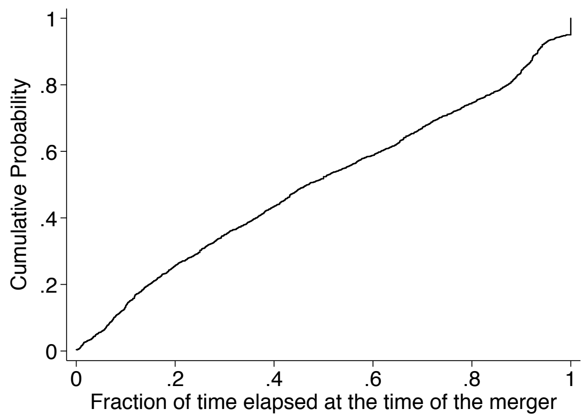
Online Appendix

Teamwork in Contests

by Jorge Lemus and Guillermo Marshall

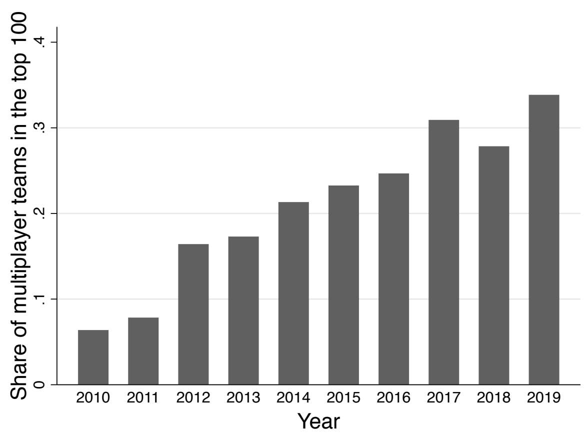
Supplemental Material – Intended for Online Publication

Figure A.1: Timing of team mergers: Cumulative probability function



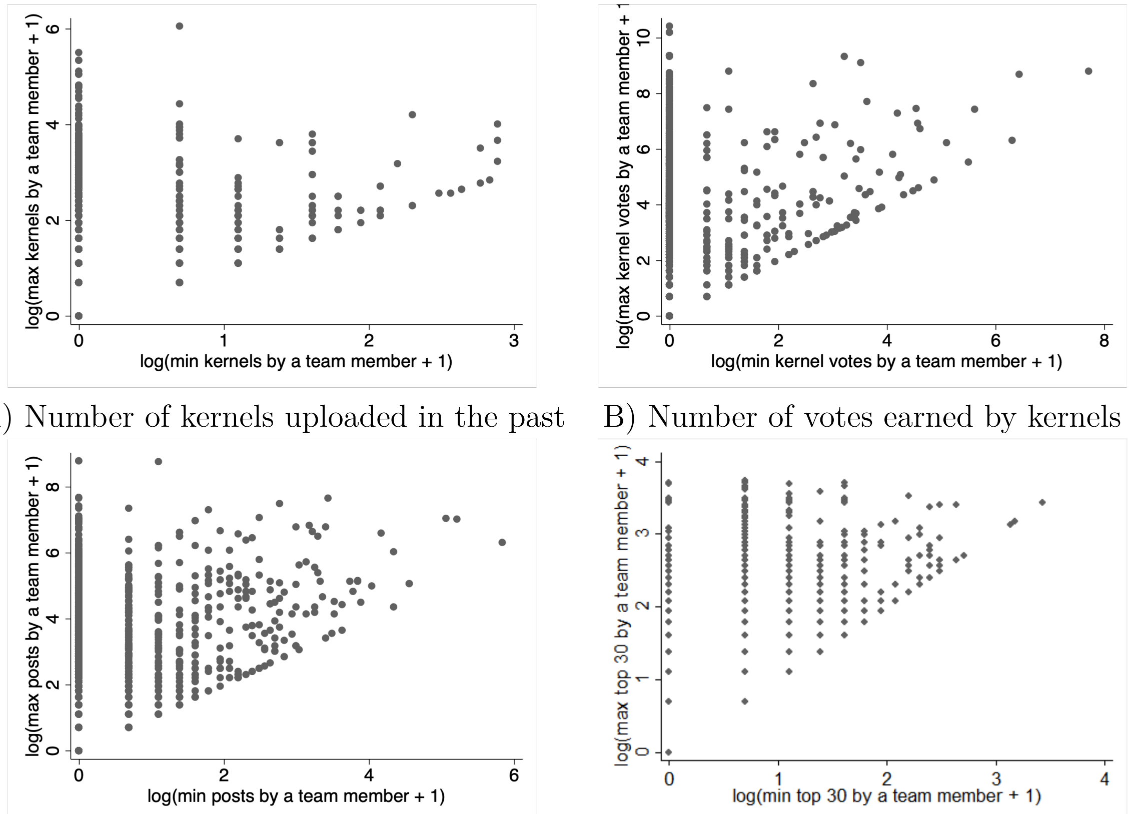
Notes: An observation is a team that welcomed a new member during the competition.

Figure A.2: How have collaboration rates changed over time?



Notes: An observation is a competition.

Figure A.3: Are players matching up with players who have similar observed outcomes?



A) Number of kernels uploaded in the past

B) Number of votes earned by kernels

C) Number of past messages on the discussion board

D) Number of past participations in the top 30

Notes: An observation is a multiplayer team. Panel A plots the number of kernels (also known as notebooks) posted by the team members in the past. Kernels are code that players can post so that any user can make use of it. Panel B plots the number of votes earned by the kernels posted by the different team members. Panel C plots the number of discussion board messages posted by the team members in the past. Panel D plots the number of past participations where the team members finished in the top 30 positions.

Table A.1: Balance table: pre-merger covariates across treated and non-treated (matched) teams

	Number of submissions up to time of merger (1)	Distance to max score on public leaderboard at time of merger (2)	Team size at time of merger (3)
Non-treated teams	15.955	1.446	1
Treated teams	15.955	1.447	1
p-value	1.000	0.978	1.000

Notes: Treated teams are teams who welcomed a new member during the competition, non-treated teams are teams who did not change their team size during the competition. The last row of the table reports the p-value of a differences-in-mean test.

Table A.2: The impact of collaboration on scores: Team-level estimates, first-time collaborators subsample

	Public score	Private score
	(1)	(2)
New member	0.114^{***}	0.121^{***}
	(0.019)	(0.023)
Observations	3,121,785	3,054,446
R^2	0.444	0.453

Notes: Standard errors clustered at the team-level in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. An observation is a submission made by a team in a competition. All specifications include team fixed effects, competition–day fixed effects, and a second-degree polynomial of variables: total number of submissions by all teams up until the submission time, total number of submissions by the team making the submission up until the submission time, total number of submissions by the member of the team making the submission up until the submission time, the submitting team's distance to the maximum score on the public leaderboard at the submission time, and the fraction of contest time that had elapsed at the submission time. The sample is restricted to include untreated teams and treated teams whose members are participating in a multiplayer team for the first time (i.e., in all previous competitions, if any, they participated in a single-member team). Further, the sample is restricted to include submissions by treated teams that took place six weeks before or after the week in which the team changed its team size, and it also restricts attention to teams with one or two members.

Table A.3: The impact of collaboration on scores: Team-level estimates, heterogeneity analysis with respect to time of merger

	Public score	Private score
	(1)	(2)
New member (early merger)	0.080^{***}	0.081^{**}
	(0.024)	(0.035)
New member (late merger)	0.078^{***}	0.086^{***}
	(0.017)	(0.018)
Observations	3,248,210	3,179,632
R^2	0.439	0.448

Notes: Standard errors clustered at the team-level in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. An observation is a submission made by a team in a competition. All specifications include team fixed effects, competition–day fixed effects, and a second-degree polynomial of variables: total number of submissions by all teams up until the submission time, total number of submissions by the team making the submission up until the submission time, total number of submissions by the member of the team making the submission up until the submission time, the submitting team's distance to the maximum score on the public leaderboard at the submission time, and the fraction of contest time that had elapsed at the submission time. The sample is restricted to include submissions by treated teams that took place six weeks before or after the week in which the team changed its team size, and it also restricts attention to teams with one or two members. An early merger (late merger) is defined as a merger that took place when less (more) than 50 percent of the contest time had elapsed.

Table A.4: The impact of collaboration on scores: Team-level estimates, competitive teams subsample

	Public score	Private score
	(1)	(2)
New member	0.075***	0.084***
	(0.016)	(0.018)
Observations	2,008,287	2,008,287
R^2	0.302	0.308

Notes: Standard errors clustered at the team-level in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. An observation is a submission made by a team in a competition. The sample is restricted to competitive teams, which are defined as teams that obtained scores above the 75th percentile of the score distribution in their competition. All specifications include team fixed effects, competition–day fixed effects, and a second-degree polynomial of variables: total number of submissions by all teams up until the submission time, total number of submissions by the team making the submission up until the submission time, total number of submissions by the member of the team making the submission up until the submission time, the submitting team's distance to the maximum score on the public leaderboard at the submission time, and the fraction of contest time that had elapsed at the submission time. The sample is restricted to include submissions by treated teams that took place six weeks before or after the week in which the team changed its team size, and it also restricts attention to teams with one or two members.

Table A.5: The impact of collaboration on scores: Team-level estimates, heterogeneity analysis with respect to contest characteristics

	(1)	(2)	(3)	(4)	(5)
Panel A. Dependent variable: Public score					
New member	0.078***	0.086***	0.073***	0.038	0.080***
	(0.014)	(0.018)	(0.021)	(0.044)	(0.015)
New member * Image data		-0.023			
		(0.026)			
New member * Large reward			0.012		
			(0.026)		
New member * Post 2015				0.044	
				(0.047)	
New member * Large dataset					-0.011
					(0.032)
Observations	3,248,210	3,248,210	3,248,210	3,248,210	3,248,210
R^2	0.439	0.439	0.439	0.439	0.439
Panel B. Dependent variable: Private score					
New member	0.085***	0.088***	0.077***	0.041	0.083***
	(0.016)	(0.019)	(0.021)	(0.045)	(0.018)
New member * Image data		-0.009			
		(0.036)			
New member * Large reward			0.019		
			(0.031)		
New member * Post 2015				0.047	
				(0.048)	
New member * Large dataset					0.012
					(0.038)
Observations	3,179,632	3,179,632	3,179,632	3,179,632	3,179,632
R^2	0.448	0.448	0.448	0.448	0.448

Notes: Standard errors clustered at the team-level in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. An observation is a submission made by a team in a competition. All specifications include team fixed effects, competition–day fixed effects, and a second-degree polynomial of variables: total number of submissions by all teams up until the submission time, total number of submissions by the team making the submission up until the submission time, total number of submissions by the member of the team making the submission up until the submission time, the submitting team's distance to the maximum score on the public leaderboard at the submission time, and the fraction of contest time that had elapsed at the submission time. The sample is restricted to include submissions by treated teams that took place six weeks before or after the week in which the team changed its team size, and it also restricts attention to teams with one or two members. Image data is an indicator for whether the contest requires use of video or image data; large reward is an indicator for contests with above average reward quantity; post 2015 is an indicator for contests taking place after 2015 (when the platform incorporated new features that facilitated communication among players, e.g., notebooks); and large dataset is an indicator for whether the dataset has an above average size (in GBs).