



Team formation with complementary skills

Mürüvvet Büyükboyacı¹ | Andrea Robbett²¹Department of Economics, Middle East Technical University, Ankara, Turkey²Department of Economics, Middlebury College, Middlebury University, Middlebury, Vermont**Correspondence**Büyükboyacı Mürüvvet, Department of Economics, Middle East Technical University, Universities District, Dumlupınar Bulvarı No: 1, 06800 Cankaya, Ankara, Turkey.
Email: muruvvet@metu.edu.tr**Funding information**

European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie, Grant/Award Number: 2015ABH67840007

Abstract

One explanation for the prevalence of self-managed work teams is that they enable workers with complementary skills to specialize in the tasks they do best, a benefit that may be enhanced if workers can sort themselves into teams. To assess this explanation, we design a real-effort experiment to study the endogenous formation of teams, and its effect on productivity, when specialization either is or is not feasible. We find a strong positive interaction between endogenous team formation and the ability to specialize, indicating that endogenous team formation is a particularly effective mechanism for promoting team output in production environments that enable the exploitation of skill complementarities.

KEYWORDS

comparative advantage, contest, endogenous team formation, matching, team performance, weakest-link technology

JEL CLASSIFICATION

C91, C92, C71, C72, D72, H41

1 | INTRODUCTION

The use of team-based incentives in the workplace has become increasingly common, with the majority of large firms employing self-managed work teams in recent years (Lazear & Shaw, 2007). One explanation for the increasing prevalence of work teams, despite the inherent free-riding incentives, is that they allow firms to exploit skill complementarities among workers (Lazear, 1998, 2000). Such teams are often formed endogenously, particularly in manufacturing industries, law, academia, and other professional organizations (Bandiera, Barankay, & Rasul, 2013). Self-organization has the potential to further amplify this benefit of teams—provided that agents are able to successfully partner with those individuals whose skills best complement their own. Both laboratory and field studies, along with anecdotal evidence, indicate that enabling workers to voluntarily join teams can have a positive impact on overall productivity (Hamilton, Nickerson, & Owan, 2003; Herbst, Konrad, & Morath, 2015). However, this study has thus far focused on *how much effort* teammates exert, rather than *how they allocate effort* across distinct tasks or whether endogenous team formation can enable agents to form teams that exploit skill complementarities. Despite the prevalence of self-managed work teams, there is little direct evidence that teammates are able to successfully coordinate on an allocation of tasks that maximize team productivity; nor is there evidence that they anticipate such coordination and sort themselves into teams that enable them to take advantage of skill complementarities.¹

To address these questions, we design an experiment to study the endogenous formation of teams and its effect on production, in team contests when specialization either is or is not feasible. By denying some teams the ability to

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.



specialize, we are able to assess whether teammates seek out different partners in different production environments. Additionally, we can isolate the channel through which team formation is likely to be most effective: Facilitating the formation of teams with complementary skills. Specifically, we consider a simple production environment, in which two team members must complete two distinct tasks to produce a unit of output. The introduction of multiple tasks, which can be potentially completed by either team member, incorporates the potential gains from specialization that is a natural feature of most team production environments (Lazear, 1998). Our experimental design allows us to induce and control participants' abilities in each task, by manipulating how much they can accomplish within the time period. Among the population of potential team members, all agents have the same overall capacity to complete tasks, but differ in the distribution of their induced "skills" across the two tasks and which task they are "best" at. Specifically, our population is equally divided among agents who are equally skilled at each task, agents who are slightly better at one task and weaker at the other, and agents who are very good at one task and very poor at the other.

Within this setting, we conduct a 2×2 team contest experiment, in which we independently vary: (a) Whether teammates have some ability to choose their partner or they instead rotate through the different possible match types in the population; (b) whether each worker must complete both tasks to produce or it is possible to specialize. In the endogenously formed team contests, participants first rank each possible teammate on the basis of the agent's observed productivity in each task and agents are sorted into teams on the basis of these rankings. We predict that the attractiveness of teammates with different skill sets will depend on whether the agents anticipate working independently on both tasks or specializing in one task. To assess these predictions, we exogenously vary whether or not specialization is possible by changing how teammates' efforts are aggregated in a perfect-complement's production function. In half the sessions, each teammate must work on both tasks to produce and we expect all productivity types to strictly prefer partners with more balanced skills. In the other sessions, participants have flexibility in how to allocate the division of tasks across teammates. If they anticipate that they will be able to coordinate on an outcome in which team members specialize, then their preferences over teammates should flip, with agents no longer valuing balanced skill sets and instead preferring those who are best at the task at which they themselves are weakest.

We find that endogenous team formation has a positive effect on productivity, primarily through the channel of producing more efficient matches, with little effect on effort or coordination. Our results indicate that team formation promotes productivity by facilitating the formation of teams of complementary skills that enable agents to focus on the task they do best. Teammates who cannot specialize spend most of their time on their weaker task and divide their time similarly regardless of how teams form. When specialization is possible, in contrast, participants spend significantly more time on their stronger task when they are able to form their own teams. Consistent with our predictions, participants prefer to be matched with more balanced types when specialization is not possible, but rank those with more extreme, but complementary, skills higher when the teammates can specialize. As a result, the realized teams in both treatments are characterized by balanced types matched with other balanced types and extreme types matched with other extreme types. Surprisingly, in situations when agents are predicted to be indifferent between two types, participants rank those with complementary skills as more preferred teammates, even when specialization is not possible. Our results highlight the remarkable ability of agents to successfully sort into teams in which they can be most productive and indicate that, as a result, endogenous team formation is particularly effective as a means of exploiting gains from complementarities in worker skills. The ability to seek out and partner with other agents whose skills are complementary is likely to be particularly valuable in situations where workers' "on the ground" knowledge of the relative strengths of their coworkers is superior to the information held by management.

Our paper is most closely related to the strands of the experimental literature concerning behavior in team contests and endogenous group formation in strategic interactions. As discussed above, the primary focus of the team contest literature is on the provision of costly effort and an overarching conclusion is that teams regularly overprovided chosen effort relative to the Nash equilibrium.² With respect to the optimal formation of teams under different production technologies, Brookins, Lightle, and Ryvkin (2015, 2017) investigate how a principal should organize agents who differ in their overall abilities when teammates' efforts are perfect substitutes or perfect complements. These studies differ from the current paper in that they study teammates engaged in an identical task with heterogeneous values for the prize or costs of effort, whereas we consider teammates each engaged in two distinct tasks, with heterogeneous skills at each, while holding constant their overall costs and benefits.

The experiment most similar to the current design is Büyükbayaci and Robbett (2017), which used the same two-task production environment.³ That study considered the effect of team contests by varying whether participants competed in teams or individually, as well as whether team members could specialize and communicate. The results

indicate that there is little difference in output between individual contests and team contests in which teammates must work independently, but that there is a strong positive effect of specialization, especially when communication was possible. That experiment was designed to address different research questions and differs significantly from the current design. Most notably, teams were not formed endogenously, there were two productivity types that were always matched together, and the costs were large enough that participants who could not specialize had an incentive to free-ride. We note, though, that the observation in this paper that the ability to specialize increases output in each type of team replicates the finding of the previous paper, under different conditions and parameters.

Herbst et al. (2015) study endogenous group formation in contests, in which three players can compete independently against each other or two players can form an alliance against the third. The participants, therefore, face the question of *whether* to form a team, but cannot choose with whom to align. They find that players who are stronger competitors when working independently (i.e., those who choose a higher effort either due to intrinsic motivation or induced lower effort costs) are less likely to opt-in to alliances. Despite this, alliances that are formed voluntarily produce significantly more effort than those formed randomly. As less applicable to the current study, a recent review of the alliance formation literature is provided in Herbst et al. (2015).

Other experimental work has considered the effect of endogenous group formation on coordination and cooperation. Riedl, Rohde, and Strobel (2016) find that participants in an endogenously formed network are better able to coordinate on the Pareto-dominant equilibrium in a weakest-link game, whereas Chen (2017) also finds higher levels of coordination when participants can form their own groups. A growing body of experimental work has studied cooperation in public goods games and other social dilemmas when groups are exogenously sorted by the experimenter based on earlier contributions (e.g., Gunthorsdottir, Houser, & McCabe, 2007) or endogenously formed by subjects, via free-mobility (e.g., Ehrhart & Keser, 1999; Robbett, 2016), expulsion (Cinyabuguma, Page, & Putterman, 2005), or voting on requests to enter, exit, or merge groups (Ahn, Isaac, & Salmon, 2008, 2009; Charness & Yang, 2014). Chaudhuri (2011) and Guido, Robbett, and Romaniuc (2018) provide a review of this literature. The public goods experiment with endogenous group formation that is most similar to the current experiment is Page, Putterman, and Unel (2005), whose ranking and matching algorithm we borrow. They find that contributions to the public good are significantly higher under this mechanism, with cooperation levels similar to that under peer punishment. Finally, Baccara and Yariv (2016) provide a theoretical model of group formation, in which groups produce and consume two separate public goods, over which agents have different relative preferences. As the cost of contributing increases relative to the cost of connecting to additional members, they find that groups become less homogeneous and are instead comprised of agents with extreme preferences who are willing to specialize in providing one of the public goods.

2 | EXPERIMENTAL DESIGN

2.1 | Setting

We consider a simple production environment, in which teammates must choose how to allocate their time across multiple tasks when specialization either is or is not feasible. There exist two tasks, Tasks A and B, and a one-to-one perfect-complement's production function. Let a_i (b_i) denote the units of Task A (Task B) produced by agent i in a given work period. Consistent with our real-effort experiment and natural workplace limitations, the amount of each task that an individual can complete is constrained by the duration of the work period and the individual's skill at the task. Specifically, assume that individual i working only on Task A during the work period can accomplish α_i units of Task A. In other words, α_i is the maximum that i could produce working only on Task A, whereas a_i denotes how many units of Task A i actually completes in the period. If i instead worked only on Task B, it would be possible to accomplish β_i units of Task B. In the model, α_i and β_i represent a worker's natural ability to complete each task. To fully control the types in our experiment, however, we induce these values in the lab.

For each agent, there is a constant marginal opportunity cost of working on Tasks A and B, rather than the best alternative use of their time. Let f denote the total cost of devoting one work period to Task A (or B). Then, each unit of Task A costs individual if/α_i and each unit of Task B costs individual if/β_i . In other words, the marginal cost of the task depends on the per-period opportunity cost and the fraction of the work period required for the individual to accomplish one unit of the task.

Let agents i and j form a team. We consider two types of team contests, which differ in whether a teammate must work on both tasks to produce or it is possible for each teammate to focus on a different task. In the *output-substitutable contest*, each teammate engages separately in a perfect-complement's production environment and their ultimate output is combined. In this case, the team's output is given by $X = \min\{a_i, b_i\} + \min\{a_j, b_j\}$. Each agent i would, therefore, prefer to

allocate his time so as to produce an equal amount of each task ($a_i = b_i$). This requires i to allocate a fraction $\beta_i / (\alpha_i + \beta_i)$ of work time to Task A and a fraction $\alpha_i / (\alpha_i + \beta_i)$ to Task B. An agent who allocates the entire work period according to this division would successfully produce $x_i = a_i = b_i = \alpha_i \beta_i / (\alpha_i + \beta_i)$ units of output. Note that this implies that an agent who is stronger at one task than the other devotes most of his time to the weaker task—that is, the worse the individual is at his weaker task relative to his stronger task, the greater the fraction of time he must devote to the weaker task. In the *input-substitutable contest*, in contrast, the teammates' inputs are combined and thus an individual no longer needs to personally complete one unit of each task to produce a unit of output. Specifically, the team's output is given by $X = \min\{(a_i + a_j), (b_i + b_j)\}$. This implies that teammates with complementary skills will be capable of producing more if they coordinate on a division of tasks, in which each teammate focuses on the task at which she is best.

2.2 | Experimental procedures

The experiment was conducted at the METU-FEAS Behavioral and Experimental Laboratory (BEL) at the Middle East Technical University (METU). Participants were recruited by email using the BEL database, which consists of undergraduate students at METU. All sessions were computerized using z-Tree (Fischbacher, 2007), and exactly 12 participants were admitted to each. Overall, 20 sessions were conducted (for a total of 240 participants) and each lasted 100 min. Throughout the experiment, payoffs were described in terms of “tokens,” with 10 tokens corresponding to 1 Turkish Lira (TL). Participants earned 28.17 TL on average, including a 10-TL participation fee.⁴

Each experimental session consisted of three stages. In the first stage, participants completed two tasks (Tasks A and B, described in detail below) sequentially, for which they were paid a piece rate. This part allowed them to gain familiarity with the tasks and learn their skill in each. In the second and third stages of the experiment, participants engaged in a series of contests in endogenously or exogenously formed teams. In half of the sessions, the exogenously formed team contests came first and, in the other half, the order was reversed. Participants were paid their earnings from the piece rate stage and two randomly selected contests from the second and third parts, to avoid potential wealth effects.⁵

All parts of the experiment used the same real-effort task, which is a modified version of the ball-catching game developed by Gächter, Huang, and Sefton (2016). In the task, computerized balls move from the top of the computer screen downward and the participant must “catch” them by aligning a computerized tray under the balls as they reach the bottom of the screen. The modified game was introduced in Büyükboyacı and Robbett (2017) and involves three main changes from the Gächter et al. (2016) task. First, we create two versions of the ball-catching game, one in which the balls participants catch are red and another in which they are blue. We refer to these as the “red game” and “blue game” in the instructions and refer to them as “Task A” and “Task B” here. Second, rather than having the balls fall randomly, we precisely fix how the balls fall on the screen, to create five different “skill levels,” or values of α_i and β_i . In each version, a fixed number of balls always fall over the course of a 60-s period: Either 80, 70, 50, 30, or 20.⁶ The timing and placement of the balls is such that any participant who is comfortable using a computer is expected to be able to collect every single ball, with the occasional error.⁷ Finally, we eliminate the cost per click that was used by Gächter et al. (2016) and instead use a time cost, consistent with the parameter f described in Section 2.1. The cost of collecting balls for 60 s was $f = 10$ tokens, or 0.167 tokens/s.

In the initial piece rate stage of the experiment, participants completed Task A for 60 s and then Task B for 60 s. Their performance in this stage is later used to provide potential teammates information about their productivity in each task. For the remainder of the experiment, participants had 60 s in each period and chose how to allocate their time across the two tasks by dynamically switching between them as often they liked. A screenshot of the game is shown in Figure 1. The participant could click the “Catch Red” or “Catch Blue” buttons to open that task, which automatically closed the other. Alternatively, they could press the “Neither” button, in which case both tasks disappeared from the screen. This enables us to distinguish and record how much time the participant spends on each task and how much time is spent not working. The participants paid the time cost of 0.167 tokens/s whenever a task was open on their screen, but did not incur costs when the “Neither” state was selected. All participants began each period in the “Neither” state until they actively chose a task.

Using the five different skill levels generated for each of the two tasks (i.e., the possible α and β values), we create five distinct “productivity types”: Balanced, A-Extreme, B-Extreme, A-Moderate, and B-Moderate. The Balanced type had a skill of 50 in each task (i.e., $\alpha_i = \beta_i = 50$), meaning that participants of this type could catch 50 red (or blue) balls in the 60-s period. The A-Extreme type had a skill of 80 at Task A and 20 at Task B ($\alpha_i = 80$ and $\beta_i = 20$), whereas the B-Extreme type had identical skills but was stronger at Task B ($\alpha_i = 20$ and $\beta_i = 80$). For the A-Moderate type, $\alpha_i = 70$ and $\beta_i = 30$, and for the B-Moderate type, $\alpha_i = 30$ and $\beta_i = 70$.⁸

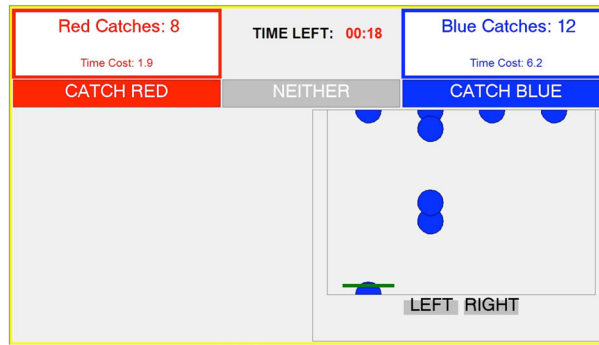


FIGURE 1 Contest screen (translated to English) [Color figure can be viewed at wileyonlinelibrary.com]

In each session of 12 people, four were assigned to be Balanced and two were assigned to be each of the other four types. Participants were told that the balls would fall at different rates for different people, but were not told anything about the existence of specific types or the specific number of balls different people were expected to catch. Instead, participants learned about their own productivity in the two tasks during the piece rate stage, which came first and enabled them to see how many of each colored ball they could catch. Additionally, they always learned the skill set of their teammate before competing together, by finding out that person's piece rate performance in the two tasks.⁹

The four treatment conditions are shown in Table 1. Each session was either input-substitutable or output-substitutable, so the participants experienced only one type of production technology. All participants participated in 15 endogenously formed team contests 15 exogenously formed team contests, with the order balanced across sessions. Participants were always matched with the same teammate for three consecutive contest periods. Participants in the exogenous contests were always assigned a new partner at the end of the three periods, whereas participants in the endogenous contests were matched with a partner according to each person's preference rankings and could be matched with an earlier teammate.

The likelihood of winning the contest was determined by the team's points and its competitor's points.¹⁰ The winner of the contest was determined probabilistically according to a piecewise linear contest success function (Büyükboyacı & Robbett, 2017; Gächter et al., 2016; Gill and Prowse, 2012). Specifically, team m producing X_m units of output competing against team k producing X_k units of output wins the contest with probability $(50 + X_m - X_k)/100$ (bounded by 0 and 1).¹¹ The winning team shares a prize of 200 tokens, whereas the losing team earns nothing. Regardless of whether they won the contest, the incurred time cost was subtracted from the participant's earnings. At the end of each period, participants saw how many of each colored ball they collected, their points (output) during the contest, and whether they won the contest or not. They did not receive feedback on the points scored by their opponents.

In the exogenously formed team contests, each person was matched with each of the other possible types for three periods. For instance, an A-Extreme type would experience five distinct matches (of three periods each) with each of the possible other types: Balanced, A-Moderate, B-Moderate, and B-Extreme, as well as with the other A-Extreme.¹² The order with which each type was matched with each other type was systematically rotated in each of the sessions and was perfectly balanced across treatment cells. The team always competed against another team of the same skill profile.¹³

In the endogenously formed team periods, participants were asked to rank their 11 potential teammates before the formation of each match. Preference rankings are, therefore, elicited before the first, fourth, seventh, tenth, and thirteenth contest periods in the 15-period endogenous matching stage. The matching mechanism is borrowed from Page et al.'s (2005) study of public good provision in endogenously formed groups. In each ranking period, the participants were shown the Tasks A and B piece rate (first stage) performance of the other 11 participants, in random order. Note that this information is equivalent to the assigned α and β values as long as participants catch every ball in the piece rate stage without any errors. No other information was provided. The participants were asked to enter a

TABLE 1 Treatment conditions

	Output-substitutable	Input-substitutable
Endogenous then exogenous	5 sessions (60 participants)	5 sessions (60 participants)
Exogenous then endogenous	5 sessions (60 participants)	5 sessions (60 participants)

Note. Each cell contains 900 distinct team periods (or 450 separate contests) and 300 ranking observations.

ranking of each person into a box beside that person's productivity, with lower numbers indicating more preferred teammates. The ranking could be any number from 1 to 100, ties were allowed, and only the order (not the magnitudes) of the ranking numbers mattered.¹⁴ After participants entered their rankings, the program first rescaled each person's rankings to be from 1 (most preferred) to 11 (least preferred). If the participant gave multiple people the same number, the ranking for each of these people was coded as the average of each of the spots these people would hold in the 1–11 ranking.¹⁵ This forces all participants' rescaled rankings to sum to the same number ($66 = 1 + 2 + \dots + 11$). Next, the computer matches participants into teams by searching across all possible pairs for the team of two individuals whose mutual ranks' sum is lowest and assigning them to a team. Then, it searches across all possible remaining pairs, finds the two unmatched participants whose rankings of each other sum to the lowest number, and matches them into a team. The process continues until all six teams have been formed.

In the endogenous treatment, the first two teams to be formed (whose members ranked each other best) always compete against each other, the next two teams compete against each other, and the last two teams selected to compete. If the participants rank each other as predicted, this implies that perfectly symmetric teams will always compete against each other in both the input-substitutable and output-substitutable treatments.¹⁶

In both the endogenously and exogenously formed team contests, participants were able to communicate with their teammates through a chat box before the contest. We included the chat period, which participants could potentially use to coordinate efforts with their teammate, because the ability to communicate is a natural feature of most workplaces and eliminating all such channels would be artificial. During the communication stage, participants were shown their own and their teammates' piece rate performance in the two tasks.¹⁷ The teammates were able to chat before each of the three periods that they were matched. To reduce boredom and irrelevant conversations, teammates were given a full minute to chat in the first period that they were matched, 30 s in the second period they were matched, and 15 s in the third period they were matched.

2.3 | Predictions

We begin by considering the equilibrium output level for each team composition in the output-substitutable and input-substitutable contests; based on the team's predicted output, we then make predictions regarding preferences over teammates, the formation of teams in the endogenous treatment, and the expected output of endogenously formed teams.

The predicted output for each team in the output-substitutable contests is straightforward. In this case, the marginal cost to player i of producing output is equal to the cost of completing one unit of Task A plus the cost of completing one unit of Task B: $f/\alpha_i + f/\beta_i$. The marginal benefit of producing an additional unit of output is equal to the marginal increase in the likelihood of winning, 0.01, times the agent's share of the prize, $V/2$. Given the parameters of the experiment, the marginal cost is between 0.40 and 0.625 for all productivity types, which is less than the marginal benefit of 1. In equilibrium, the team members would, therefore, spend the entire period working and allocate their time optimally such that $a_i = b_i = \alpha_i\beta_i/(\alpha_i\beta_i)$. The Balanced types would thus produce 25 units, the Moderate types 21 units, and the Extreme types 16 units. The equilibrium team output for each team composition is presented in Figure 2. The dark gray triangle lists the output for the team when each member works independently at producing an equal number of each task, as is predicted in the output-substitutable treatment.

The predicted output for each team in the input-substitutable contests is less clear-cut, because the optimal behavior of each agent depends on the production decisions of his teammate. If i 's teammate j chooses $a_j = b_j$, then i would also wish to choose $a_i = b_i$ and allocate his time identically to the output-substitutable case. Therefore, the output-substitutable equilibrium, in which each of the teammates work independently to complete an equal number of each task, is also an equilibrium in the input-substitutable treatment. Alternatively, if i 's teammate chooses $a_j < b_j$, then i would wish to produce such that $a_i = (b_j - a_j) + b_i$. In other words, he would wish to complete Task A to balance his teammate's surplus production of Task B and then would use the remaining time to complete an equal number of both tasks. Because the marginal cost of completing both tasks is lower than the marginal benefit, in equilibrium teams always spend the entire time period producing. However, for most team compositions, there exist a range of team outputs that could be supported in equilibrium.

The white triangle in Figure 2 lists the range of equilibrium outputs for the team in the input-substitutable contest. Because each team member working independently to complete an equal number of each task is an equilibrium outcome in the input-substitutable contest, these ranges always include the "independent" prediction listed in the dark gray triangle. Finally, each white triangle in Figure 2 also lists the maximum output that the team could produce if each teammate *only specialized* by working on a single task. In cases where the teammates have different skills at the task they are specializing in, this is not an

	Balanced ($\alpha=50, \beta=50$)	A-Moderate ($\alpha=70, \beta=30$)	B-Moderate ($\alpha=30, \beta=70$)	A-Extreme ($\alpha=80, \beta=20$)	B-Extreme ($\alpha=20, \beta=80$)
Balanced ($\alpha=50, \beta=50$)	Independent: 50 [50] Only specialize: 50	Independent: 46 [40, 56] Only specialize: 50	Independent: 46 [40, 56] Only specialize: 50	Independent: 41 [35, 56] Only specialize: 50	Independent: 41 [35, 56] Only specialize: 50
A-Moderate ($\alpha=70, \beta=30$)	Independent: 46 [40, 56] Only specialize: 50	Independent: 42 [42] Only specialize: 30	Independent: 42 [30,70] Only specialize: 70	Independent: 37 [35, 40] Only specialize: 30	Independent: 37 [27, 72] Only specialize: 70
B-Moderate ($\alpha=30, \beta=70$)	Independent: 46 [40, 56] Only specialize: 50	Independent: 42 [30,70] Only specialize: 70	Independent: 42 [42] Only specialize: 30	Independent: 37 [27, 72] Only specialize: 70	Independent: 37 [35, 40] Only specialize: 30
A-Extreme ($\alpha=80, \beta=20$)	Independent: 41 [35, 56] Only specialize: 50	Independent: 37 [35, 40] Only specialize: 30	Independent: 37 [27, 72] Only specialize: 70	Independent: 32 [32] Only specialize: 20	Independent: 32 [20, 80] Only specialize: 80
B-Extreme ($\alpha=20, \beta=80$)	Independent: 41 [35, 56] Only specialize: 50	Independent: 37 [27, 72] Only specialize: 70	Independent: 37 [35, 40] Only specialize: 30	Independent: 32 [20, 80] Only specialize: 80	Independent: 32 [32] Only specialize: 20

FIGURE 2 Equilibrium output by team composition.

Note. This figure presents the equilibrium output levels for each productivity type (row) matched with each other possible productivity type (column). The dark triangle in each box displays the unique equilibrium output in the output-substitutable treatment, where the two teammates can only produce by working independently on both tasks. In the input-substitutable treatment, working independently remains an equilibrium outcome, but there generally exist a wider range of equilibrium output levels, presented in the white triangle for each team, which includes the output-substitutable equilibrium output. As not necessarily an equilibrium outcome, at the bottom of each box, we also list the maximum output the team could achieve if each teammate only ever works on one task [Color figure can be viewed at wileyonlinelibrary.com]

equilibrium outcome, because the stronger teammate would prefer to complete a single task only up to the units the other teammate could complete, and then work on both tasks for the remaining time. However, we include it as a behavioral prediction, because this may be an obvious outcome for teammates to coordinate on.

The predicted equilibrium output for each team, presented in Figure 2, leads to several hypotheses. First, we compare the predicted output under the two production environments. When the teammates are the same type, output under the two types of contests is identical. Otherwise, there are a range of predicted outputs in the input-substitutable contest that includes both lower and higher outputs than the output-substitutable prediction. Assuming that teams are able to coordinate on the equilibrium associated with the greatest output, we predict the output will be higher in the input-substitutable case. However, we emphasize that achieving this higher output requires more careful coordination among the teammates (relative to simply working independently) and carries the risk of coordination failure.

Prediction H1 (Team output in output-substitutable and input-substitutable contests). *Teams will produce more output in the input-substitutable contests compared with the output-substitutable contests.*¹⁸

Looking across each row (or down each column) in Figure 2, we can assess each type's preference over teammates. In the output-substitutable contests, all types always achieve the highest equilibrium output when matched with the Balanced type and lowest equilibrium output when matched with an Extreme type. In the input-substitutable contests, the attractiveness of different types depends on whether the participants anticipate that the team will specialize or work independently. If the participants in the input-substitutable contest anticipate that they will be able to coordinate on the highest output equilibrium, then Balanced types prefer any of the more extreme types to other Balanced types, whereas Moderates and Extreme types prefer the types in the following order: The Extreme type with complementary skills, the



Moderate type with complementary skills, the Balanced type, the Moderate type with similar skills, and the Extreme type with similar skills.¹⁹ This leads to our second prediction.

Prediction H2 (Rankings in output-substitutable and input-substitutable contests). *All participants will favor more balanced types in the output-substitutable contests. In the input-substitutable contests, participants favor types with extreme, but complementary, skills to the more balanced types.*

From Figure 2, we also observe that comparing equilibrium output across types will sometimes generate indifferences in the agents' preference rankings. For instance, all agents in output-substitutable contests should expect the same team output if they are matched with an A-Extreme or with a B-Extreme. When all else is equal, we predict that participants will favor types who are more similar to themselves (e.g., A-Moderates and A-Extreme types would prefer A-Extreme to B-Extreme, and so on) consistent with the homophily principle, which describes the tendency of individuals to associate with others of similar characteristics (McPherson, Smith-Lovin, & Cook, 2001).

Prediction H3 (Rankings when expected output is the same). *When participants are expected to be indifferent between two types on the basis of equilibrium output, they will rank those whose skill profiles are similar to their own more favorably.*

In contrast, one area in which heterophily has been documented is in collaborative networks (Johnson et al., 2009; Xie et al., 2016). It is possible that past experience favoring those with complementary skills in team production environments leads participants to naturally prefer teammates who are dissimilar, all else equal, in this setting.

Given Prediction H2, we anticipate that participants will form teams in which similar types are matched together. In the output-substitutable treatment, the sum of mutual rankings will be lowest among the Balanced types and they will thus form two teams of Balanced types. Among the remaining teammates, Moderates will mutually rank each other lowest and will form two teams of Moderates (and, if Prediction H3 holds, A-Moderate will be matched with A-Moderate and B-Moderate with B-Moderate). Finally, the Extremes will be matched with each other (and, if Prediction H3 holds, A-Extreme will be matched with A-Extreme and B-Extreme with B-Extreme). In the input-substitutable treatment, the preference rankings differ but the resulting teams are also characterized by Extremes matched with Extremes, Moderates with Moderates, and Balanced with Balanced. Extreme-A and Extreme-B types will mutually rank each other lowest and thus form the first two teams. The next two teams formed will be Moderate-A with Moderate-B, and the remaining two teams will be composed of Balanced types.

Prediction H4 (Team formation in output-substitutable and input-substitutable contests). *When teams can form endogenously in both input-substitutable and output-substitutable treatments, we expect that teams will form in which Balanced types are matched with Balanced, Moderate types with Moderate types, and Extreme types with Extreme types. In the input-substitutable treatment, we further expect Moderate-A to be matched with Moderate-B and Extreme-A to be matched with Extreme-B.*

Finally, we consider the predicted output in the endogenously formed teams compared with the exogenously formed teams. In the output-substitutable contests, participants work independently and there is no productivity benefit from forming specific teams. Theoretically, output should be identical in the two treatments: Regardless of who they are matched with, Balanced types produce 25, Moderates produce 21, and Extremes produce 16, for an average team output of 41.33.²⁰ In the input-substitutable treatment, we expect that teams will form with similar, but complementary types, and teammates specialize. In this case, the expected team output is 66.67 (corresponding to the average maximum equilibrium output in the team compositions predicted in Prediction H4).²¹ In the exogenously formed team contests, each type is rotated through meeting with the other types and the average maximum equilibrium output over each of these meetings is 54.8. Therefore, we expect that output will be greater when participants can endogenously form teams *only* in the input-substitutable treatment.

Prediction H5 (Team output in exogenously formed and endogenously formed teams). *In output-substitutable contests, the average output will be the same regardless of whether the teams are formed endogenously or*

exogenously. In the input-substitutable contests, the average output will be greater when teams can be formed endogenously.

3 | RESULTS

In this section, we examine the overall productivity of teams across the four treatments. We then assess how teammates allocate their time in each treatment, consider participants' ranking over teammates in the endogenously formed contests and the resulting team formation, and analyze the content of the communication stage.

We first consider performance in the initial piece rate stage, when participants completed one task at a time and were paid per catch. Because participants choose teammates on the basis of how many balls were caught in this stage, it is important to first check that the productivity of each type in the piece rate stage is similar to the prediction. Table A1 in the Appendix reports the predicted and actual performance for each type on each task during the piece rate stage and indicates that the average catches were very close to the predicted level (within 0.4) for each type. We, therefore, conclude that our assigned productivities were successful.

3.1 | Productivity of teams

We first consider the overall productivity of teams that are formed exogenously and endogenously when teammates either cannot specialize (output-substitutable) or can focus on one task (input-substitutable). Table 2 presents the average output level for all treatments along with nonparametric tests for treatment differences, taking each session of 12 interacting participants as the unit of observation. First, we observe a positive effect of endogenous team formation (first row). Overall, participants who could choose their teammates produced five more units of output than those who could not (44.7 vs. 39.7; $p = 0.019$). This positive effect of team formation is entirely driven by the input-substitutable treatment, as can be seen from the next two rows of the table. When teammates' output is substitutable, endogenous team formation has no effect on the average team output ($p = 0.80$). When teammates can specialize, teams that formed endogenously produce more than 11 additional units of output (55.4 vs. 44.2; $p = 0.005$). Therefore, we confirm Prediction H5 that endogenous team formation has a positive effect on output in teams that can specialize (and no effect when teams cannot specialize). Looking down the columns in Table 2, we see that teams produce more in the input-substitutable treatment (overall, 49.8 vs. 34.6; $p < 0.0001$), confirming Prediction H1. Furthermore, the positive effect of enabling teammates to specialize is significantly greater when they can form teams than when teams are assigned exogenously (difference in differences: 12.45; $p < 0.001$).

From Table 2, we also note that the average output in each case is significantly below the maximum equilibrium team output discussed at the end of Section 2.3. The average team output in the output-substitutable teams if participants always allocate their time optimally is 41.3. Instead, teams produce, on average, less than 35 units, indicating that they are not most efficiently allocating their time across the two tasks. As there exist a range of predicted output levels in the input-substitutable contests, we note that the teams are producing less than the maximum equilibrium output for the teams they are assigned in the exogenous treatment (44.15 instead of 54.8) and for the output-maximizing team compositions in the endogenous treatment (55.39 instead of 66.67). Figure 3 further breaks

TABLE 2 Average team output by treatment

	All	Exogenous	Endogenous	Endogenous = Exogenous
All	42.19	39.69	44.70	$Z = 2.351^{**}$
Output-substitutable	34.61	35.22	34.01	$Z = 0.255$
Input-substitutable	49.77	44.15	55.39	$Z = 2.803^{***}$
Output-Subs. = Input-Subs.	$Z = 5.775^{***}$	$Z = 3.628^{***}$	$Z = 3.78^{***}$	

Note. The table reports the average team output in each treatment condition with the session as the unit of observation, meaning that each session of 12 participants is treated as one observation in the endogenous treatment and one observation in the exogenous treatment. The final column reports two-tailed Wilcoxon signed-rank tests comparing output when teams are formed endogenously and exogenously in each session ($n = 20$ overall and $n = 10$ when separately considering the output-substitutable and input-substitutable treatments). The final row reports two-tailed Mann-Whitney U significance tests for the difference in output-substitutable and input-substitutable sessions ($n = 10$ in each comparison group when separately considering the Endogenous and Exogenous treatments). *10%, **5%, ***1%.

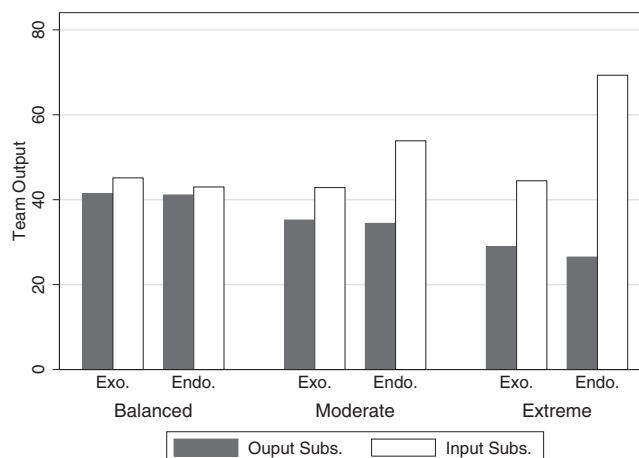


FIGURE 3 Team output for Balanced (left), Moderate (center), and Extreme (right) types by treatment

down the team output in each treatment by the different productivity types: Balanced (left), Moderate (center), and Extreme (right). For the balanced types, who do not gain from specializing themselves, there is little difference across treatments. Balanced types achieve only slightly higher output in the input-substitutable treatment, and the difference is significant only when teams are formed exogenously ($p = 0.0284$ with the session as the unit of observation). This is consistent with the prediction that they will be matched with fellow Balanced types when team formation is endogenous and thus not experience productivity benefits from specialization. Moving to the Moderate and Extreme types, we observe, as predicted, that these types achieve less output than the Balanced types in the output-substitutable contest, but more output in the input-substitutable contests with endogenous team formation.²² Furthermore, there is a strong benefit of endogenous team formation for Moderate and Extreme types in the input-substitutable contest, when they can choose teammates with complementary skills.²³

The regression estimates reported in Table 3 confirm the results of the nonparametric tests. The first five columns present multilevel regression models of the determinants of team output, with random effects for individual and session. The model in Col. (1) regresses team output on indicators for the input-substitutable and endogenous treatment and confirms the positive effect of each on output. The second column additionally includes an interaction for the Input-Substitutable and Endogenous indicators (with exogenous output-substitutable contests as the omitted condition) and confirms that there is little effect of endogenous team formation on its own, but a strong positive interaction with input-substitutable contests. The next models include indicators for Moderate and Extreme types (with Balanced omitted) and demonstrate that there is little effect of type on output overall [Col. (3)], but a strong treatment effect [Col. (4)]: Moderates and Extremes produce significantly less in the output-substitutable contests and significantly more in the input-substitutable contests. Further, Col. (5) indicates that Moderates and Extremes produce more in exogenous input-substitutable contests, but far more still in endogenous input-substitutable contests. The last column replicates the model of Col. (2), except that each observation is the session-wide average team output in the endogenous or exogenous stage (in which case the only regressors are the treatment indicators and interaction) and reports essentially identical estimates to Col. (2). Finally, we include a period as a regressor in each model presented in Cols. (1)–(5) and find that output increases by approximately 0.40 units/period.

Figure 4 graphs the average team output over the course of the 15 contest periods for each treatment and demonstrates that the rank order of the treatments is consistent with Table 2 across all periods. Although endogenous team formation does not affect team output level under the output-substitutable production technology, there is a stable positive effect in input-substitutable contests. We also observe a weak increase in output over time in all four treatments.²⁴

3.2 | Allocation of time across tasks

To determine the source of these treatment differences, we turn to the question of how teammates allocate their time across the two tasks in each of the treatment conditions. First, we note that teammates spend very little time on *neither* task and the amount of time spent not working is similar across treatments.²⁵ We, therefore, conclude that the

TABLE 3 Determinants of team output

	(1)	(2)	(3)	(4)	(5)	(6)
Input-Subs.	15.16*** (1.401)	8.929*** (1.520)	8.929*** (1.520)	-3.452* (1.817)	3.667** (1.639)	8.929*** (1.562)
Endogenous	5.016*** (1.956)	-1.213 (2.029)	-1.213 (2.029)	-1.213 (2.029)	-0.362 (1.997)	-1.213 (1.941)
Endogenous × Input-Subs.		12.46*** (2.670)	12.46*** (2.670)	12.46*** (2.670)	-1.780 (2.750)	12.46*** (2.812)
Moderate			-1.095 (1.458)	-6.478*** (0.675)	-6.267*** (0.934)	
Extreme			-0.373 (3.163)	-13.56*** (1.041)	-12.50*** (1.348)	
Moderate × Input-Subs.				10.77*** (1.550)	3.977** (1.604)	
Extreme × Input-Subs.				26.38*** (1.848)	11.81*** (1.691)	
Moderate × Endogenous					-0.423 (1.352)	
Extreme × Endogenous					-2.132 (1.665)	
Moderate × Endogenous × Input-Subs.					13.58*** (3.848)	
Extreme × Endogenous × Input-Subs.					29.13*** (3.817)	
Period	0.394*** (0.137)	0.394*** (0.137)	0.394*** (0.137)	0.394*** (0.137)	0.394*** (0.137)	
Constant	28.96*** (1.648)	32.07*** (1.454)	32.56*** (2.099)	38.75*** (1.524)	38.33*** (1.538)	35.22*** (0.904)
Observations	7,200	7,200	7,200	7,200	7,200	40
Clusters	20	20	20	20	20	-

Note. Cols. (1)–(5) include random effects for individuals and session. Standard errors clustered at a session level. In Col. (6), each observation is the average session output in the endogenous or exogenous phase. *10%, **5%, ***1%.

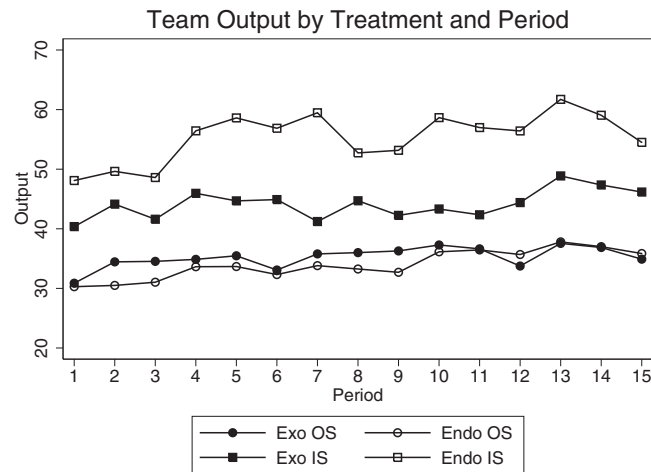


FIGURE 4 Output over time by treatment . IS: input substitutable; OS: output substitutable

treatment differences are not driven by differences in overall time spent working and instead focus on how the participants allocate their work time across the two tasks.

Figure 5 presents the percentage of work time (i.e., time spent on either task) the teammates devote to each task in each of the treatments, along with the predicted allocation of time if teammates work independently (dashed line). Regardless of how much time they spend working in the output-substitutable contest, Balanced types should always devote 50% of that work time to each task and Moderate (Extreme) types should spend 30% (20%) on their stronger task, to complete an equal amount of each task. The left panel shows the amount of time that the Balanced types devote to Tasks A and B. As predicted, we observe that the Balanced types devote close to half of their working time to each task and that there is no difference in time allocation across the four treatments.

For the Moderate and Extreme types, Figure 5 displays the proportion of time the teammates spent on their stronger and weaker tasks. As predicted, teammates in the output-substitutable contests spend most of their time on their weaker task.²⁶ Compared with the theoretical predictions of 30% or 20%, however, both types still spend too much time on their stronger task.²⁷ As predicted, the Extremes devote significantly less time to their stronger task than the Moderates do in the output-substitutable contests and there are no differences in the amount of time spent on the stronger task between the exogenous and endogenous treatments for either type.²⁸ Participation in an input-substitutable contest has a strong positive effect on the amount of time allocated to the stronger task.²⁹ Furthermore, the effect is strongest when teammates can form their own teams than when they are formed exogenously.³⁰

Thus, the treatment differences in output (observed in Section 3.1) are mirrored in the amount of time the participants devote to the task they are best at. As there is no difference in time allocation across the treatments for Balanced types, there are substantial differences in how that time is allocated across the two tasks by participants who are better at one task than the other. Moderate and Extreme types spend most of their time on their weaker task in

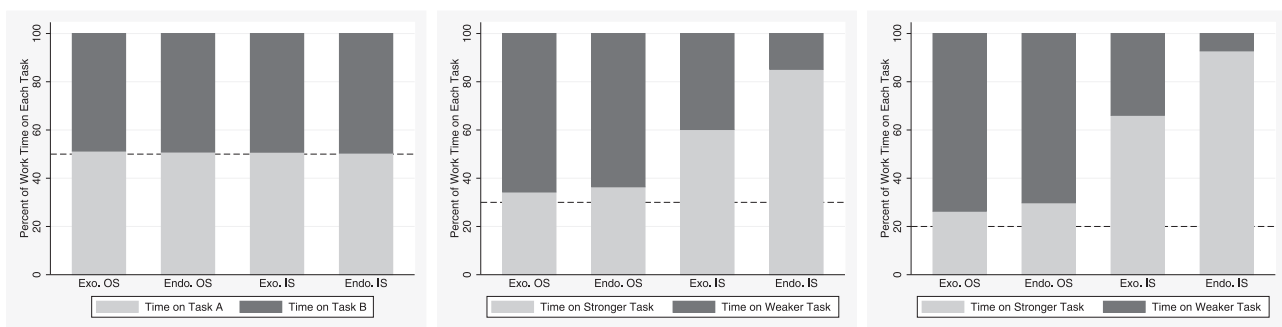


FIGURE 5 Time allocation across treatments for Balanced (left), Moderate (center), and Extreme (right) types. The dashed line shows the equilibrium time allocation if teammates work independently. IS: input substitutable; OS: output substitutable

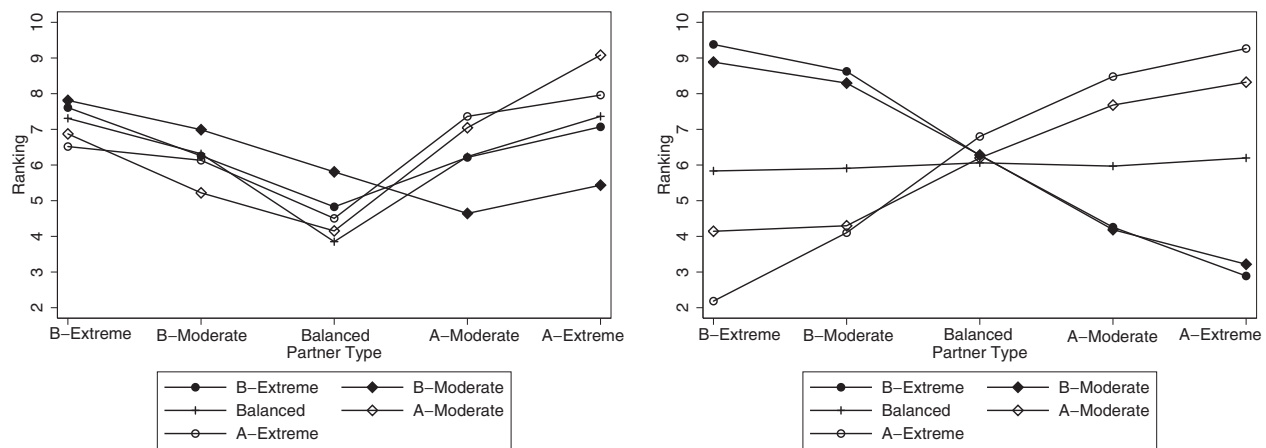


FIGURE 6 Rankings of potential partners in output-substitutable (left) and input-substitutable (right) team contests

output-substitutable contests. In input-substitutable contests with exogenous team formation, teammates spend more time on the task they are best at, and this effect is even stronger when teammates can form their teams endogenously. To explore this finding further, we next examine how participants sort into teams.

3.3 | Rankings over teammates and realized matches

We now turn to the participants' rankings of their potential teammates and the realized team compositions, which directly follow from these rankings. To assess the participants' rankings, we focus on the *rescaled* rankings that force the sum of each participant's rankings to equal 66, rather than the raw ordinal ranking that they enter for each potential teammate. The left panel of Figure 6 presents the average rankings that each type of participant assigns to each type of potential partner in the output-substitutable contests. Consistent with the prediction, we find that all types rank Balanced types significantly better than Moderates and Moderates significantly better than Extremes.³¹ The potential partner types on the horizontal axis are ordered by skill at Task A (with Balanced types in the center). Recall that lower rankings indicate more preferred partners. For four of the five types, we observe a clear "V-shaped" pattern, with the Balanced types receiving the lowest rankings and more extreme types receiving strictly higher rankings. The exception is the B-Moderate type, who ranks A-Moderate lowest.³² Other than the B-Moderates, all other types rank the Balanced types significantly better than either the Moderates or the Extremes.³³ We, therefore, confirm the first part of H2: In the output-substitutable contests, more balanced types are ranked significantly better than more extreme types.

We also hypothesized (Prediction H3) that, in the case of indifference based on expected output, participants would favor those with more similar skills. We see little indication of this—in fact, B-Moderates, A-Moderates, and A-Extremes all rank those with *opposite skills* as significantly better than those with similar skills.³⁴ The difference is substantial (on average, nearly two ranking slots higher) and there is no indication that this preferential treatment of complementary types diminishes over the course of the experiment. Therefore, we reject Prediction H3 and instead find evidence that participants prefer those with different skills from their own *even* when there is no productivity advantage to matching with a teammate with complementary strengths.

If participants in the input-substitutable contests assume that teammates will work independently, then they should rank their potential teammates identically to the output-substitutable case. However, teammates in the input-substitutable contests can produce higher output if they form teams with complementary types and specialize. Recall that Moderate and Extreme types who anticipate achieving the highest equilibrium output feasible for their team would rank their potential teammates in the order listed on the horizontal axis (reading from left to right for those stronger at A and right to left for those stronger at B): Extreme with complementary skills, Moderate with complementary skills, Balanced, Moderate with the same skills, and Extreme with the same skills. As not generally an equilibrium outcome, participants may instead anticipate that each type would only specialize in the task they were best at (and then stop working when they hit the maximum their teammate could produce). In this case, Moderates would become indifferent between A-Extremes and A-Moderates and between B-Extremes and B-Moderates, whereas Extremes' preferences

	Balanced	Moderate	Extreme		Balanced	Moderate	Extreme
Balanced	Frequency: 21.33% Output: 44.73	Frequency: 12% Output: 35.39	Frequency: 12% Output: 32.39	Balanced	Frequency: 22.67% Output: 41.78	Frequency: 16% Output: 48.71	Frequency: 5.33% Output: 42.06
	Frequency: 20% Output: 45.12	Frequency: 13.3% Output: 39.43	Frequency: 13.3% Output: 34.1		Frequency: 20% Output: 43.9	Frequency: 13.3% Output: 46.33	Frequency: 13.3% Output: 45.83
Similar Moderate		Frequency: 1.67% Output: 39.8	Frequency: 8.00% Output: 29.33	Similar Moderate		Frequency: 1.67% Output: 27.4	Frequency: 2.00% Output: 9.33
		Frequency: 6.67% Output: 37.6	Frequency: 13.3% Output: 30.85		Frequency: 6.67% Output: 28.3	Frequency: 13.3% Output: 29	
Complementary Moderate		Frequency: 14.3% Output: 38.77	Frequency: 14.67% Output: 28.32	Complementary Moderate		Frequency: 16.3% Output: 64.76	Frequency: 12.67% Output: 63.39
		Frequency: 6.67% Output: 36.4	Frequency: 13.33% Output: 29.3		Frequency: 6.67% Output: 58.5	Frequency: 13.33% Output: 56.48	
Similar Extreme			Frequency: 3.33% Output: 22.2	Similar Extreme			Frequency: 0.33% Output: 4
			Frequency: 6.67% Output: 25.7		Frequency: 6.67% Output: 19.55		
Complementary Extreme			Frequency: 12.67% Output: 25.24	Complementary Extreme			Frequency: 23% Output: 78.71
			Frequency: 6.67% Output: 26.65		Frequency: 6.67% Output: 65.55		

FIGURE 7 Resulting teams and output with endogenous team formation treatment (white rectangles) and exogenous team formation (gray rectangles) in output-substitutable contests (left) and input-substitutable contests (right)

would remain strict. Finally, Balanced types are expected to be indifferent between Moderates and Extremes, both of which are preferred to Balanced. If, instead, the Balanced types assume that each person would only specialize, then they would be fully indifferent between all types.

The average rankings that each type of participant assigns to each type of potential partner in the input-substitutable contests are depicted in the right panel of Figure 6. Rather than the “V-shaped” pattern of the output-substitutable contests, we see, as predicted, that Moderates and Extremes assign monotonically decreasing rankings to participants as we move from those with extreme but similar skills to those with extreme and complementary skills. Across all Moderates and Extremes, we find that participants assign more preferred rankings to Extreme types with complementary skills than Moderate types with complementary skills ($p < 0.0001$, $Z = 5.616$ at individual level and $p = 0.0008$, $Z = 3.3354$ at session level). Likewise, Moderates and Extremes both assign more preferred rankings to Moderate types with complementary skills than Balanced types ($p < 0.0001$, $Z = 8.036$ at individual level and $p < 0.0001$, $Z = 4.224$ at session level), more preferred rankings to Balanced types than Moderates with similar skills ($p < 0.0001$, $Z = 7.738$ at individual level and $p < 0.0001$, $Z = 4.135$ at session level), and more preferred rankings to Moderate types with similar skills than Extreme types with similar skills ($p = 0.0002$, $Z = 3.711$ at individual level and $p = 0.0161$, $Z = 2.41$ at session level). We, therefore, confirm the second part of Prediction H2. As can be seen from Figure 6, these patterns are similar across all four Moderate and Extreme types (with the possible exception of the A-Moderates, who rank B-Extreme and B-Moderates similarly), but there is some indication that the lines are less steep for Moderates than for Extremes. However, the significance tests suggest that the rankings are strict for Moderate types as well: Moderates rank complementary Extremes better than complementary Moderates (at individual level, $p = 0.0241$, $Z = 2.256$ and at session level, $p = 0.13$, $Z = 1.513$) and similar Extremes worse than similar Moderates (at individual level, $p = 0.012$, $Z = 2.511$ and at session level, $p = 0.0638$, $Z = 1.853$). Therefore, Moderate and Extreme participants appear to rank their potential partners according to the output-dominant equilibrium. In contrast, the Balanced type rankings are the same over all types and we see little indication that Balanced types rank other Balanced types less favorably than the Moderates or Extremes ($p > 0.78$, $Z < 0.28$ at both session and individual levels).

Finally, we consider the resulting matches in the endogenous treatment, noting that the formation of teams directly follows from the ranking behavior already discussed. The realized relative frequency of each team composition in the endogenous treatment is given in the white rectangles in Figure 7, with the left panel reporting the outcomes in the output-substitutable contests. As described by Prediction H4, in both types of contests, we expect Balanced types to form teams with other Balanced types, Moderates to form teams with Moderates, and Extremes to form teams with Extremes. The modal outcomes for each type are in line with the predicted team compositions, but other matches are frequently observed as well. At the individual level, Extreme types are matched with other Extreme types 48% of the time, Moderate types are matched with other Moderates 48% of the time, and Balanced types are matched with other Balanced 64% of the time.³⁵ In output-substitutable contests, there are no productivity benefits to forming teams with complementary types. We, therefore, expect Moderates and Extremes to be equally inclined to form teams of

TABLE 4 Effect of team formation on output

	(1) Output-Subs.	(2) Output-Subs.	(3) Input-Subs.	(4) Input-Subs.	(5) Input-Subs.	(6) Input-Subs.
Endogenous	-1.164 (0.819)	-1.111 (0.769)	2.135 (1.370)	2.175* (1.188)	8.574*** (2.421)	8.714*** (1.974)
Similar Mod.	-7.208*** (1.229)	-6.848*** (2.027)	-13.80*** (2.700)	-13.51*** (2.926)		
Comp. Mod.	-6.930*** (1.090)	-7.154*** (1.426)	19.61*** (2.107)	19.52*** (1.992)	1.076	1.019
Similar Ext.	-20.83*** (2.149)	-20.73*** (1.875)	-22.71*** (2.576)	-22.80*** (3.170)		
Comp. Ext.	-18.77*** (1.466)	-18.69*** (1.464)	32.27*** (1.882)	32.12*** (1.853)	13.27***	13.08***
Similar Mod. & Ext.	-14.81*** (1.359)	-14.64*** (1.417)	-15.01*** (2.291)	-15.06*** (2.329)		
Comp. Mod. & Ext.	-15.97*** (1.444)	-15.72*** (1.302)	17.44*** (2.339)	17.37*** (1.907)		
Balanced & Mod.	-7.072*** (1.500)	-6.456*** (1.353)	5.039*** (1.694)	4.611** (1.843)		
Balanced & Ext.	-11.85*** (1.392)	-11.57*** (1.346)	2.742 (2.239)	2.649 (2.144)		
Match Num.	1.635*** (0.275)	1.643*** (0.266)	1.492*** (0.361)	1.491*** (0.384)	0.541	0.645
Constant	40.58*** (1.278)	40.35*** (1.359)	37.08*** (1.830)	37.17*** (1.896)	54.12***	53.86***
Observations	600	600	600	600	236	236

Note. Each observation is the average output for a type over the third periods they were matched. *10%, **5%, ***1%. Match Num. specifies when this three-period interaction occurred out of the five separate team matchings each person experiences [(1)-(5)]. Even numbered columns include session random effects. Final two columns include only complementary matches.

complementary or similar teammates, unless participants have an intrinsic preference for those with similar skills (Prediction H3) or complementary skills (as indicated in the ranking results described above). If teams of two Extremes (or two Moderates) were to form randomly, we would expect that two-thirds of the resulting teams would be composed of complementary types and one-third of similar types.³⁶ Instead, among all teams of two Extremes (Moderates), 79.2% (89.6%) are composed of complementary types.

In input-substitutable contests, we expect the teams of Moderates and Extremes to be comprised of members with complementary skills. Again, the modal match for each type aligns with the predictions, with somewhat more variation than expected. Extreme types are matched with the complementary Extreme type 69% of the time, whereas Moderate types are matched with the complementary Moderate type 49% of the time, and Balanced types are matched with other Balanced types 69% of the time. Finally, we can also quantify the extent to which participants successfully form the most productive teams in the input-substitutable sessions. If we normalize the maximum possible output in these sessions to 100, then the maximum possible output given the exogenously formed teams that were assigned is 82.2, indicating an output loss of 17.8% resulting from inefficient partnerships. The maximum possible output given the endogenously formed teams that were realized is 95.7, indicating that roughly three-quarters of the output associated with inefficient teams, is eliminated when participants can form their own matches.

3.4 | Output by team composition

The results thus far indicate that endogenous team formation has a positive impact on output as a result of participants forming teams with complementary skills such that each teammate can specialize in what they do best, with no such impact observed when teams are unable to specialize. To more directly assess whether team formation *in itself* has a positive impact on output, we consider team output *conditional* on team composition. Figure 7 also reports the average output produced by each team composition in the output-substitutable (left panel) and input-substitutable contests (right panel), with the white squares displaying the outcomes for endogenously formed teams and the gray squares displaying the outcomes for exogenously formed teams. In output-substitutable contests, there appears to be little difference in output between the endogenously and exogenously formed teams of each type. These results are supported in the regression estimates reported in Table 4. The average output for a team over their three-period match is regressed on an indicator for the endogenous treatment and variable coding when the match occurred ("Match Num.") in the series of five distinct three-period matches, as well as dummies for each of the possible team compositions (with two balanced types as the omitted condition).³⁷ The first two columns use only the output-substitutable contests, with the second column also including session random effects.³⁸ Controlling for the composition of the team, there is a negative but insignificant effect of endogenous team formation on output. (The team composition coefficients indicate that all team compositions achieve significantly less output compared with the two balanced types, with the relative magnitudes essentially as expected.)

Turning to the input-substitutable contests, the right panel of Figure 7 again reveals that some team compositions achieve higher output when teams are formed endogenously, whereas others experience lower output. However, there are some notably large differences in this case. Conditional on matching with a complementary type (in teams of two Extremes, two Moderates, or one Moderate and one Extremes), participants who formed the team endogenously produce, on average, higher output. As far less commonly observed, Extreme types that match endogenously with other Extremes or Moderates with similar skills achieve far less output than those in the same matches exogenously. These observations are confirmed in the regression results in Table 4. Models (3) and (4) include all team matches in the input-substitutable contests and indicate a positive but weak effect of endogenous team formation. The final two models include only matches of complementary types (two complementary moderates, two complementary extremes, or, the omitted condition, a complementary moderate and extreme) and indicate a strong positive effect of endogenous team formation: Complementary teams that formed endogenously produce more than eight additional units of output. We, therefore, find no overall indication that team formation *in itself* promotes productivity. This is consistent with our previous interpretation of the results that the positive interaction between team formation and the ability to specialize is due to teammates forming more productive matches. However, we also do find evidence that agents who deliberately seek out complementary teammates are far better at coordinating with those teammates on the highest possible equilibrium (relative to those who were randomly assigned the same partnership).

More generally, we can also consider the output in each session as a fraction of the maximum possible given the teams that were actually formed. In the output-substitutable contests, output is 85.2% of the maximum for exogenously formed and slightly (but not significantly) lower for endogenously formed teams at 82.2% ($p > 0.79$, session-level

Wilcoxon signed-rank test). In the input-substitutable treatments, output is 80.6% in exogenously assigned teams and 86.7% in endogenously formed teams ($p = 0.09$, session-level Wilcoxon signed-rank test). This marginally higher productivity, even conditional on teams formed, is consistent with the finding above that highly productive teams are most productive when they specifically seek each other out.

3.5 | Communication content

Finally, we also conducted a content analysis of the communication between team members in each treatment. The content of the chats was quantified by independent coders, following the procedures of Cason et al. (2012, 2017). The details of the procedure and chat content can be found in the online appendix. Overall, we observe that the most common discussion topic in the output-substitutable contests concerned the strategy of catching an equal number of each ball, which was discussed in approximately 80% of teams in the output-substitutable contests, but fewer than a quarter of teams in the input-substitutable contests. In contrast, three-quarters of input-substitutable teams discussed a specialization plan and many also discussed a strategy in which players “subsidize” the team by completing both tasks after reaching the target. Finally, more than half of teams in both output-substitutable treatments and the exogenous input-substitutable treatment talked about setting a target output level. This topic was somewhat less common (32%) in the endogenous input-substitutable contests, in which teams are more likely to be formed with perfectly complementary types.

4 | CONCLUSION

Both laboratory and observational studies have found that enabling workers to voluntarily join work teams can have a positive impact on overall productivity (Hamilton et al., 2003; Herbst et al., 2015). However, this study, as well as the experimental team contest literature more broadly, largely focuses on the provision of costly effort and not on how workers allocate their effort across different tasks. In contrast, the ability of workers to match with teammates who have complementary skills and specialize in the task they do best is considered a primary benefit to the formation of work teams (Lazear & Shaw, 2007). We, therefore, design a controlled laboratory experiment that allows us to isolate the impact of team formation itself, as well as the channel through which it is likely to be most effective: Facilitating the formation of teams with complementary skills. To do so, we design a simple real-effort team contest experiment, in which work teams must complete two distinct tasks to produce. We then independently vary agents' ability to form their own teams and their ability to specialize by focusing on one task. This design additionally enables us to assess how agents seek out teammates differently under production environments, in which each team member must work independently or in which collaboration is possible.

Overall, we find that there is a strong positive interaction between the ability to specialize and the ability of participants to match with particular teammates. When specialization is not possible, endogenous team formation has no effect on output. The ability to specialize, in itself, has a positive effect on output, which is significantly magnified when agents have a say in who joins their team. We further find that this impact on productivity is driven, not by how much time teammates spend working, but how they direct their time. When specialization is not possible, teammates each spend most of their time on the tasks they do poorly; in contrast, teammates who can specialize switch to an allocation of time in which they focus on their stronger task and endogenous team formation has a significant positive effect on their ability to specialize in the task they do best. As predicted, we find that agents who cannot specialize seek out teammates with average skills at both tasks. Surprisingly, however, when evaluating potential teammates with imbalanced skill sets, they exhibit a preference for those with complementary skills, even when they are predicted to be indifferent on the basis of output. Although preference for similarity in traits, such as education and age, has been well-documented, there is also evidence that individuals seek out complementary teammates in collaborative production environments ranging from online video games to street gangs (Johnson et al., 2009; Xie et al., 2016). It is possible that participants' past experiences cause them to use a heuristic in which they naturally seek out complementary teammates in a collaborative environment, even when there is no productivity advantage. Future research may consider whether such behavior persists even when there is a cost to doing so. When specialization is possible, teammates' preferences switch such that they seek out those with more extreme (but complementary) skill sets, which is consistent with participants anticipating that the team will be able to coordinate on an equilibrium in which each team member focuses on one task.

Our results, therefore, suggest that team formation is likely to be most effective at promoting production in environments where team members have heterogeneous skills and collaboration is possible. We find that the

participants in our experiment are remarkably adept at forming the teams in which they can be most productive. In situations where individuals can specialize, our participants demonstrate a strong ability to sort themselves into teams with complementary skills and coordinate on an allocation of work time in which each teammate specializes in the task they do best. This finding suggests that team members are capable of effectively organizing their own work groups, without the authority of a manager or principal, who may have inferior information about the relative skills of their employees. At the same time, we note that beyond the context of our simple production environment, workers may not have information about the skill profile of their potential teammates until after they have worked together. Additionally, the process of forming teams in the workplace is likely to be informal and decentralized and, compared with the formal matching mechanism used in our experiment, may lead to less efficient matches. As our team contest environment is theoretically equivalent to a setting in which team members are paid based on output, it is possible that teammates working outside of the more competitive contest setting may focus more on fairness considerations, such as relative earnings or production, rather than the sole objective of maximizing output, causing them to favor similarly productive partners. Finally, teammates may have specific preferences over tasks, which are independent of their relative skills, as well as preferences for working with specific people they like, regardless of their strengths. All of these considerations may complicate both the team matching process and the team's ability to agree on how each person will allocate their time across tasks and indicate important topics for future research.

ACKNOWLEDGMENTS

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement (2015ABH67840007). We thank Jeff Carpenter and Peter Matthews for useful discussions and comments.

END NOTES

- ¹ Bandiera et al. (2013) study team formation among soft fruit pickers, where there is little room for specialization, and find that workers sort into teams on the basis of their social connections when incentives are relatively weak and form teams with others of similar ability when incentives are stronger. Hamilton et al. (2003) study production in a garment factory where workers can voluntarily opt-in to team production and find that more productive workers were among the earliest to join teams and that there was a positive effect of team formation even beyond this, indicating that teams facilitate the use of collaborative skills. A large theoretical and experimental literature is devoted to studying individual behavior in team contests and other team production environments. This study primarily focuses on the provision of costly effort on a single task, with teammates' efforts either serving as perfect complements or perfect substitutes, and not on how workers allocate their time across distinct tasks (Sheremeta, 2018, provides a comprehensive and current review of the literature).
- ² See Sheremeta (2018) for a review. Experiments that allow teams to communicate, as in our design, include Cason, Sheremeta, and Zhang (2012, 2017), Leibbrandt and Sääksvuori (2012), and Sutter and Strassmair (2009).
- ³ The only other experiment we are aware of in which team members choose how to divide two separate tasks is Cooper and Sutter (2013). In that experiment, teams of two people each play takeover games (Samuelson and Bazerman, 1985) against outside opponents. In each team, one player takes the (strategically challenging) role of a buyer, who must bid on a company, and the other takes the (strategically straightforward) role of a seller, who must decide whether to accept a bid. They vary whether the roles are assigned randomly by the computer or endogenously by mutual agreement and find that those teams that can choose how to assign the roles are more likely to have the more able player fill the role of the buyer, but that these teams do not perform better, perhaps because the choice of task assignment is a distraction from considering the optimal bidding behavior. There is also a distinct literature on multitasking, in which individual agents work on multiple tasks, finding that agents who are forced to multitask or choose their own schedule generally perform worse than those who work on distinct tasks sequentially (Buser & Peter, 2012).
- ⁴ At the time of the experiment, the minimum hourly wage in Turkey was 8.46 TL and the exchange rate was approximately 3.6 TL to one US dollar.
- ⁵ The instructions in a session were read in three parts. First, the first stage instructions were read and participants completed the piece rate stage. Then, instructions for the common parts of the second and third stages (i.e., the production environment and contest rules, described below) and the type of matching for the second stage (i.e., either endogenous or exogenous matching) were read. Finally, the instructions specific to the matching for the third stage were read after the second stage was completed. Full instructions for the stages can be found in the Appendix.
- ⁶ As we use the term "skills" to refer to these induced productivities and "stronger" to refer to participants who can produce more in a given task, these abilities are fully controlled by the computer program.
- ⁷ Participants also were given a 30-s practice round with randomly placed balls before the piece rate stage, to minimize the likelihood of errors. Participants collected close to the maximum number of balls in the piece rate stage (see Table A1).
- ⁸ In Büyükboyacı and Robbett (2017), there were two types: One that could catch 30 red and 60 blue in 90 s and the other that could catch 100 red and 25 blue in 90 s. Participants were always matched with the opposite type.
- ⁹ We chose these specific types and this particular distribution of types for several reasons. First, it allows for the possibility that perfectly symmetric teams will form. Second, for all types, the total number of balls caught in the piece rate stage ($\alpha_i + \beta_i$) is 100. This ensures that all participants have the same overall productivity profile and that each has earned an identical amount in the piece rate stage. Third, we assigned the productivities such that the production of the Moderate types working alone (21) was roughly in between those of the Balanced (25) and Extreme (16) types, while still retaining round numbers and keeping $\alpha_i + \beta_i = 100$. Finally, an individual's preference over teammates flips from preferring balanced to extreme types if they anticipate specialization, as described in Section 2.3.

- ¹⁰ As described in Section 2.1, the team's points from the game were determined according to the minimum of red and blue balls caught by the team in input-substitutable treatment and according to the sum of minimum of red and blue balls caught by each team member in output-substitutable treatment.
- ¹¹ The benefit of the linear contest success function is that participants are incentivized to produce output provided the per-person prize/100 is greater than the marginal cost. This implies that teammates do not have to coordinate on how much output to produce overall (as in a Tullock contest) and can focus on how to produce output. Note that the expected marginal costs and benefits are equivalent to a pure team production setting with each teammate paid based on team output. We use the contest setting to tie the experiment to the bulk of the related literature and to focus participants' attention on the objective of winning the contest by maximizing output, rather than nonmonetary objectives, such as equalizing teammates' earnings or output. This design choice is discussed further in conclusion.
- ¹² Because it is not possible for each of the Balanced types to be matched with all four distinct other types (whereas each of those types is also matched with all the other types), this matching necessitated Balanced types to be matched with other Balanced more than once; however, for Balanced types, there is no difference between being matched with someone who is stronger at Task A versus Task B.
- ¹³ For instance, a team with a Balanced type and an A-Extreme would compete against either another Balanced/A-Extreme team or the equivalent Balanced/B-Extreme team. A team with an A-Extreme and B-Moderate would compete against either another A-Extreme/B-Moderate team or a symmetric team of B-Extreme/A-Moderate, and so forth. In all cases, the predicted output for the two competing teams is identical.
- ¹⁴ Participants were able to use any number from 1 to 100 to allow for greater flexibility during the ranking process. This way, participants could easily slot each potential teammate into the right position as they were going down the list, with plenty of unique integers to use. Restricting the ranking range to be between 1 and 11 may be more intuitive, but it could require more trial and error on the participants' part to ensure that each person was ranked correctly in this range, and potentially be less accurate if they did not wish to revise their rankings as they went down the list to fit this requirement.
- ¹⁵ For instance, if two people were given the most preferred ranking, both were coded as 1.5 (averaging the two first-place spots). If three people were tied for fourth place, all three were coded as 5 (averaging the fourth, fifth, and sixth place spots). Although agents typically do not have a strong incentive to rank potential partners who they are indifferent between as tied rather than giving them consecutive rankings, both ranking strategies yield the same rankings in expectation.
- ¹⁶ We use this matching method rather than a decentralized matching system or a ranking system, in which players are randomly selected to choose among the remaining players because it allows us to elicit a full ranking over all 11 potential partners and it implies that individuals who mutually prefer each other to all other available partners will be matched. We note that truthful revelation of preferences is not a dominant strategy for all possible rankings that the others could submit and that the ranking should be thought of as the priority, in which the participant wishes to seek out potential partners, rather than perfectly capturing their underlying preference ordering.
- ¹⁷ They were able to figure out whom they were matched with from their preference list with that information.
- ¹⁸ As not a major hypothesis of this study, Prediction H1 is essential for generating the other treatment-related predictions to follow: If participants are not able to coordinate on a more efficient equilibrium when specialization is possible, then we do not anticipate productivity benefits of team formation or differences in teammate preferences across treatments.
- ¹⁹ If, instead, teammates anticipate that each partner will specialize by working only on one task, Balanced types become indifferent across all partners (as they can achieve 50 units of output regardless of their partner's type) and Moderates become indifferent between A-Extreme and A-Moderates and between B-Extreme and B-Moderates. Extreme types' preferences would remain strict.
- ²⁰ We note, however, it is possible that participants may respond positively to being able to control their team (similar to Herbst et al., 2015), in which case output would be higher and the output-substitutable contests serve as a control for assessing the impact of endogenous team formation when theory predicts no effect.
- ²¹ Specifically, we expect two teams of two Balanced types producing 50 each, two teams of A-Moderates matched with B-Moderates, producing 70 each, and two teams of A-Extremes matched with B-Extremes producing 80 each.
- ²² Taking the session as the unit of observation, in output-substitutable contests, Balanced types achieve significantly higher output than either the Moderates or Extremes both when teams are formed endogenously ($p = 0.023$ and 0.0003) or exogenously ($p = 0.0012$ and 0.0002). In input-substitutable contests with endogenous team formation, Balanced types achieve significantly less output than Moderates and Extremes ($p = 0.0126$ and 0.0002).
- ²³ Taking the session as the unit of observation, Moderate ($p = 0.008$) and Extreme ($p = 0.0002$) types in input-substitutable contests achieve significantly higher output with endogenous team formation.
- ²⁴ At the individual level, the Spearman correlation coefficients are significantly positive in each treatment ($p < 0.004$). At the session level, the correlation between period and output is not significant in the exogenous input-substitutable contests ($p = 0.18$) and but significantly positive in all other treatments ($p < 0.07$).
- ²⁵ The portion of the 60-s work period that participants spend not working ranges from 6.1% in the exogenous output-substitutable contests to 6.9% in the endogenous output-substitutable contests, and there are no significant differences across treatments. Taking the session as the level of observation, $p > 0.76$ in all pairwise tests, while using the individual as the unit of observation, $p > 0.34$ in all pairwise tests.
- ²⁶ Taking either the individual or the session as the level of observation, the difference is significant at $p \leq 0.007$ level for both Moderates and Extremes in both the endogenous and exogenous treatments.
- ²⁷ Taking either the individual or the session as the level of observation, the difference between the amount of time spent on the stronger tasks is significantly different from the prediction at $p < 0.006$ level or better for each type in each treatment.
- ²⁸ Comparing the Moderate and Extreme types, with the individual as the unit of observation, $p < 0.0001$ for both the endogenous and exogenous treatments and, with the session as the unit of observation, $p < 0.002$ for the exogenous and $p < 0.082$ for the endogenous treatments. Comparing the exogenous and endogenous treatment of each type with the individual or session as the unit of observation, $p > 0.78$ for Moderates and $p > 0.58$ for Extreme types.
- ²⁹ The amount of time spent on the stronger task is significantly greater in the input-substitutable treatment with exogenously formed teams than in either of the output-substitutable treatments for both types. With the individual as the unit of observation, $p < 0.0001$ for each type and treatment comparison. With the session as the unit of observation, $p < 0.0002$.

- ³⁰ Recall that the incentive to spend time on one's stronger task varies depending on whom an individual is matched with. With the individual as the unit of observation, the difference in time spent on stronger task between endogenous and exogenous teams is significant at $p < 0.0001$ for each type. With the session as the unit of observation, $p = 0.0041$ for Moderates and $p = 0.0002$ for Extreme types.
- ³¹ At the participant level, the differences in rankings of Balanced and Moderates, Balanced and Extremes, and Moderates and Extremes are all significant at $p < 0.0001$ ($Z = 7.49, 10.024,$ and 6.40), respectively. At the session level, the differences are all significant at $p < 0.0006$ ($Z = 4.09, 4.31,$ and 3.44).
- ³² A possible explanation for this pattern may be that participants may consider the likelihood of the match being realized. If moderate types realize that the Balanced types will prefer to be together, they may "skip" over them in their rankings. A similar phenomenon was documented by Echenique, Wilson, and Yariv (2016) in the context of a two-sided matching experiment using the deferred acceptance algorithm (Gale and Shapley, 1962). It is, however, then surprising that the B-Moderates rank the other B-Moderate worse than Balanced. Furthermore, we see no evidence of Extremes engaging in this strategy.
- ³³ $Z > 2.38, p < 0.02$ in all comparisons at either the individual or the session-type level.
- ³⁴ At the participant level (session level), the difference is significant at $p < 0.007$ ($p < 0.02$) for all three types.
- ³⁵ Note that this cannot be directly read from Figure 7, which shows the overall realized teams, not the frequency with which each type is matched with each other type. To calculate these relative frequencies from the figure, we must double count homogeneous group observations to account for both members being in the predicted match.
- ³⁶ This is simply due to the fact that there are twice as many other agents in the population with exactly complementary skills than identical skills, and so we expect complementary teams to be twice as likely to form.
- ³⁷ Specifically, the indicator "Similar Mod." is equal to 1 if the team consists of two Moderates with the same skills, "Comp. Mod" indicates a team of two Moderates with complementary skills, "Similar Ext." indicates two Extremes with the same skills, "Comp. Ext." indicates two Extremes with complementary skills, "Similar Mod. & Ext." indicates a Moderate and an Extreme who are strong at the same task, "Comp. Mod. & Ext." indicates a Moderate and an Extreme who are strong at different tasks, "Balanced & Mod." indicates a team with one Balanced and one Moderate, and "Balanced & Ext." indicates a team with one Balanced and one Extreme. When combined with the omitted condition of two Balanced team members, these nine possible team compositions align with the nine boxes in Figure 7.
- ³⁸ Additionally, clustering the standard errors at the session level yields similar conclusions about the significance of the coefficients, as reported in Table A2 in the Appendix.

ORCID

Mürüvvet Büyükboyacı  <http://orcid.org/0000-0002-0697-1479>

REFERENCES

- Ahn, T. K., Isaac, R. M., & Salmon, T. C. (2008). Endogenous group formation. *Journal of Public Economic Theory*, 10(2), 171–194.
- Ahn, T. K., Isaac, R. M., & Salmon, T. C. (2009). Coming and going: Experiments on endogenous group sizes for excludable public goods. *Journal of Public Economics*, 93(1-2), 336–351.
- Baccara, M., & Yariv, L. (2016). Choosing peers: Homophily and polarization in groups. *Journal of Economic Theory*, 165, 152–178.
- Bandiera, O., Barankay, I., & Rasul, I. (2013). Team incentives: Evidence from a firm level experiment. *Journal of the European Economic Association*, 11(5), 1079–1114.
- Brookins, P., Lightle, J. P., & Ryvkin, D. (2015). An experimental study of sorting in group contests. *Labour Economics*, 35, 16–25.
- Brookins, P., Lightle, J. P., & Ryvkin, D. (2017). Sorting and communication in weak-link group contests. *Journal of Economic Behavior and Organization*, Vol. 152, pp. 64–80.
- Buser, T., & Peter, N. (2012). Multitasking. *Experimental Economics*, 15, 641–655.
- Büyükboyacı, M., & Robbett, A. (2017). Collaboration and free-riding in team contests. *Labour Economics*, 49, 162–178.
- Cason, T. N., Sheremeta, R. M., & Zhang, J. (2012). Communication and efficiency in competitive coordination games. *Games and Economic Behavior*, 76, 26–43.
- Cason, T. N., Sheremeta, R. M., & Zhang, J. (2017). Asymmetric and endogenous communication in competition between groups. *Experimental Economics*, 20(4), 946–972.
- Charness, G., & Yang, C. L. (2014). Starting small toward voluntary formation of efficient large groups in public goods provision. *Journal of Economic Behavior and Organization*, 102, 119–132.
- Chaudhuri, A. (2011). Sustaining cooperation in laboratory public goods experiments: A selective survey of the literature. *Experimental Economics*, 14(1), 47–83.
- Chen, R. (2017). Coordination with endogenous groups. *Journal of Economic Behavior and Organization*, 141, 177–187.
- Cinyabuguma, M., Page, T., & Putterman, L. (2005). Cooperation under the threat of expulsion in a public goods experiment. *Journal of Public Economics*, 89, 1421–1435.
- Cooper, D. J., & Sutter, M., (2013). Role selection and team performance. IZA Discussion Paper No. 5892.
- Echenique, F., Wilson, A., & Yariv, L. (2016). Clearinghouses for two-sided matching: An experimental study. *Quantitative Economics*, 7(2), 449–482.
- Ehrhart, K-M., & Keser, C. (1999). Mobility and cooperation: On the run. CIRANO Working Paper.
- Fischbacher, U. (2007). z-Tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics*, 10, 171–178.

- Gale, D., & Shapley, L. S. (1962). College Admissions and the Stability of Marriage. *The American Mathematical Monthly*, 69(1), 9–15.
- Gächter, S., Huang, L., & Sefton, M. (2016). Combining 'real effort' with induced effort costs: The ball catching task. *Experimental Economics*, 19(4), 687–712.
- Gill, D., & Prose, V. (2012). A structural analysis of disappointment aversion in a real effort competition. *American Economic Review*, 102(1), 469–503.
- Guido, A., Robbett, A., & Romaniuc, R. (2018). Group formation and cooperation in social dilemmas: A survey and meta-analytic evidence.
- Gunthorsdottir, A., Houser, D., & McCabe, K. (2007). Disposition, history and contributions in public goods experiments. *Journal of Economic Behavior and Organization*, 62(2), 304–315.
- Hamilton, B. H., Nickerson, J. A., & Owan, H. (2003). Team incentives and worker heterogeneity: An empirical analysis of the impact of teams on productivity and participation. *Journal of Political Economy*, 111, 465–497.
- Herbst, L., Konrad, K. A., & Morath, F. (2015). Endogenous group formation in experimental contests. *European Economic Review*, 74, 163–189.
- Johnson, N. F., Xu, C., Zhao, Z., Ducheneaut, N., Yee, N., Tita, G., & Hui, P. M. (2009). Human group formation in online guilds and offline gangs driven by a common team dynamic. *Physical Review*, 79, 066117.
- Lazear, E. (1998). *Personnel economics for managers*. New York: John Wiley & Sons.
- Lazear, E. (2000). Personnel economics and economic approaches to incentives. *HKCER Letters*, 61, 1–8.
- Lazear, E., & Shaw, K. (2007). Personnel economics: The economists view of human resources. *Journal of Economic Perspectives*, 21(4), 91–114.
- Leibbrandt, A., & Sääksvuori, L. (2012). Communication in intergroup conflicts. *European Economic Review*, 56, 1136–1147.
- McPherson, M., Smith-Lovin, L., & Cook, J. M. (2001). Birds of a feather: Homophily in social networks. *Annual Review of Sociology*, 27, 415–444.
- Page, T., Putterman, L., & Unel, B. (2005). Voluntary association in public goods experiments: Reciprocity, mimicry, and efficiency. *Economic Journal*, 115, 1032–1053.
- Riedl, A., Rohde, I., & Strobel, M. (2016). Efficient coordination in weakest link games. *Review of Economic Studies*, 83(2), 737–767.
- Robbett, A. (2016). Community dynamics in the lab. *Social Choice and Welfare*, 46(3), 543–568.
- Samuelson, W. F., & Bazerman, M. H. (1985). Negotiating under the winner's curse. In V. Smith (Ed.), *Research in Experimental Economics* (Vol. 3, pp. 105–137). Greenwich, CT: JAI Press.
- Sheremeta, R. M. (2018). Behavior in group contests: A review of experimental research. *Journal of Economic Surveys*, 32(3), 683–704.
- Sutter, M., & Strassmair, C. (2009). Communication, cooperation, and collusion in team tournaments—an experimental study. *Games and Economic Behavior*, 66, 506–525.
- Xie, W.-J., Li, M.-X., Jiang, Z.-Q., Tan, Q.-Z., Podobnik, B., Zhou, W.-X., & Stanley, H. E. (2016). Skill complementarity enhances heterophily in collaboration networks. *Scientific Reports*, 6(1), 47.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Büyükboyacı M, Robbett A. Team formation with complementary skills. *J Econ Manage Strat*. 2019;28:713–733. <https://doi.org/10.1111/jems.12296>