Econometrica, Vol. 89, No. 6 (November, 2021), 2637–2657

TEAM PLAYERS: HOW SOCIAL SKILLS IMPROVE TEAM PERFORMANCE

BEN WEIDMANN
Harvard Kennedy School, Harvard University

DAVID J. DEMING
Harvard Kennedy School, Harvard University and NBER

Most jobs require teamwork. Are some people good team players? In this paper, we design and test a new method for identifying individual contributions to team production. We randomly assign people to multiple teams and predict team performance based on previously assessed individual skills. Some people consistently cause their team to exceed its predicted performance. We call these individuals "team players." Team players score significantly higher on a well-established measure of social intelligence, but do not differ across a variety of other dimensions, including IQ, personality, education, and gender. *Social skills*—defined as a single latent factor that combines social intelligence scores with the team player effect—improve team performance about as much as IQ. We find suggestive evidence that team players increase effort among teammates.

KEYWORDS: Skills, human capital, design of experiments.

1. INTRODUCTION

TEAMWORK IS INCREASINGLY IMPORTANT in the modern economy. In 2017, 78 percent of U.S. employment was in occupations where group work was judged either a "very" or "extremely" important part of the job (O*NET (2020)). Employer surveys consistently find that collaboration, communication, and ability to work in a team are among the most desired attributes of new hires (e.g., National Association of Colleges and Employers (NACE) (2019)). Since 1980, occupations requiring high levels of social interaction have grown nearly 12 percentage points as a share of all jobs in the U.S. economy, and have experienced faster wage growth at the same time (Deming (2017)).

The economic payoff to social skills arises because teams often operate more efficiently than people working in isolation (e.g., Lindbeck and Snower (2000), Hamilton, Nickerson, and Owan (2003), Lazear and Shaw (2007), Bloom and van Reenen (2011)). Yet while teamwork skills are highly valuable in principle, in practice it is difficult to isolate individual contributions to team performance. A large literature in economics estimates productivity spillovers across workers and peers (e.g., Falk and Ichino (2006), Mas and Moretti (2009), Arcidiacono, Foster, Goodpaster, and Kinsler (2012), Herbst and Mas (2015), Cornelissen, Dustmann, and Schönberg (2017), Isphording and Zölitz (2019)). Yet this evidence is only useful for the relatively small number of jobs in which individual productivity can be reliably measured. In contrast, while there are many studies of

Ben Weidmann: benweidmann@g.harvard.edu
David J. Deming: david_deming@harvard.edu
We would like to thank Maria Bertling for indispensable help in running the experiment, along with excellent support from Alki Iliopoulou, Gabe Mansur, and the Harvard Decision Science Lab. Thanks also to Doug Staiger, Derek Neal, David Autor, Alex Mas, Angela Duckworth, Steven Sloman, Duncan Watts, Erika Weisz, Mina Cikara and participants in the Cikara lab work-in-progress seminar, the Duckworth Lab, the E-con of Education seminar, NBER Labor Studies, Cornell, Yale, University of Chile, and Harvard Business School for invaluable feedback. Thanks to Victoria Johnson for superb research assistance, and to Brad Duchaine and Karen Sullivan for generously sharing access to materials for the memory task. All errors are our own.

 https://doi.org/10.3982/ECTA18461

the determinants of team success, team-level performance differences are not easily attributed to individual members of the group.[1] How do we know which people are good team players?

In this paper, we design and test a new experimental method for identifying individual contributions to team performance.[2] We first assess individual performance on several different tasks. We then randomly assign individuals to multiple teams, and we measure each team's performance on tasks that are identical or very similar to those that were administered individually. We use the individual scores to generate a prediction for the performance of each team. We then ask whether some teams consistently outperform their prediction when an individual is randomly assigned to them. We call these individuals *team players*.

Team players improve group performance, conditional on their own skill in the task at hand. If we added a chess grand master to a chess-playing team, that person would clearly increase team performance but would not necessarily be a team player by our definition. Instead, team players are individuals who consistently cause their team to produce more than the sum of its parts.

Our first finding is that team players exist. In our pre-registered model, an individual who scores one standard deviation higher on the estimated team player index increases team performance by 0.13 standard deviations. This effect is about 65 percent as large as the impact of individual task-specific skill. We validate the existence of the team player effect by showing that team players improve team performance on a novel, out-of-sample problem-solving task. Our results are robust to a variety of alternative ways of measuring the team player effect and are consistent across task types. We also explore the robustness of our findings to measurement error in the assessment of individual skills.

Our second finding is that team players score significantly higher on the Reading the Mind in the Eyes Test (RMET), a well-established and psychometrically valid measure of social intelligence (Baron-Cohen, Wheelwright, Hill, Raste, and Plumb (2001), Woolley et al. (2010), Baker, Peterson, Pulos, and Kirkland (2014), Engel et al. (2014)). After controlling for task-specific skills, IQ does *not* predict whether someone is a good team player. The team player effect is also uncorrelated with gender, age, education, ethnicity, and scores on the "Big 5" personality factors. Each of these tests was part of our pre-analysis plan, and we report all those results in the first part of the paper before moving on to exploratory analyses.

If we treat our estimated team player index and the RMET as two noisy measures of the same construct, that construct—which we will call *social skill*—predicts team performance about as much as IQ. Consistent with the theoretical model in Deming (2017), social skills improve the productivity of teams and thus are more valuable in workplace settings where more teamwork is required.

[1]An important exception is Almaatouq, Yin, and Watts (2020), who measured individual skill and then used it as a mediator to understand variation in group performance. More broadly, a large literature in organizational psychology studies the determinants of effective teamwork. For an overview, see Driskell, Salas, and Driskell (2018). Characteristics such as group average IQ, personality, and knowledge and experience of and attitudes toward teamwork are all positively correlated with team performance (Morgeson, Reider, and Campion (2005), Bell (2007), Driskell, Salas, and Hughes (2010)). Of particular interest is the literature on "collective intelligence" (CI), which identifies a common factor predicting group performance across a wide range of tasks (Woolley, Chabris, Pentland, Hashmi, and Malone (2010), Engel, Woolley, Jing, Chabris, and Malone (2014)). Woolley et al. (2010) found that CI is predicted by the group's average emotional perceptiveness, conversational turn-taking, and the share of the group that is female.

[2]Our experimental design, statistical analysis plan, and main outcomes of interest were pre-registered with the American Economic Association Randomized Controlled Trial registry as AEARCTR-0002896.

Our experiment is designed to establish the existence of individual differences in the ability to contribute to team production. However, our results are consistent with multiple theoretical models of team production. Many studies treat social or "non-cognitive" skills as additively separable contributions to a skill vector in a Mincerian earnings regression (e.g., Heckman, Stixrud, and Urzua (2006)). In Deming (2017), social skill reduces coordination frictions in team production, which implies that cognitive skill and social skill are complements in a wage equation. We are unable to fully adjudicate between different mechanisms for the impact of being a good team player, including improved communication and integrative thinking, increased allocative efficiency of participants to tasks, and others.

We provide two pieces of suggestive evidence that team players increase effort among teammates. First, groups with good team players are more likely to persist on a task and use their full allotment of time, which is positively correlated with team performance. Second, the team player effect holds even when sub-tasks are performed separately by individual team members, with little direct interaction. This suggests that team players might motivate teammates to exert more individual effort. However, we emphasize that the effort channel may operate alongside other mechanisms, which should be the subject of future study.

Our paper makes three main contributions. First, we develop a new methodology for estimating individual contributions to group performance. We show that *repeated* random assignment is necessary to estimate individual contributions to team performance. Additionally, isolating the "team player" effect from other factors requires conditioning on individual skill in closely related tasks. While the lab setting helped us carefully control these conditions, our experimental approach generalizes to the field and to more complicated real-world tasks (Falk and Heckman (2009), Charness and Kuhn (2011)). Our work is similar in spirit to the literature in economics which estimates productivity by separately identifying worker and firm effects on wages (e.g., Abowd, Kramarz, and Margolis (1999), Card, Heining, and Kline (2013), Cornelissen, Dustmann, and Schönberg (2017)).

Second, we uncover a direct mechanism for the economic payoff to social skills in the labor market. Workers with higher social skills causally improve team performance, beyond what their individual task-specific skills would suggest.[3] Our findings are consistent with many other studies showing labor market returns to social skills and "non-cognitive" skills (e.g., Kuhn and Weinberger (2005), Heckman, Stixrud, and Urzua (2006), Borghans, Duckworth, Heckman, and Weel (2008), Almlund, Duckworth, Heckman, and Kautz (2011), Lindqvist and Vestman (2011), Heckman and Kautz (2012), Deming (2017)). A closely related body of work in economics and psychology finds that prosociality is associated with positive labor market outcomes (Kosse, Deckers, Pinger, Schildberg-Hörisch, and Falk (2020)). These studies most often estimate wage differences for individuals with different skill endowments but cannot directly link skills to job performance.

Our third contribution is practical—large productivity gains are possible for employers who can accurately identify and recruit team players. While our experiment is conducted in a lab, there are several reasons to believe that the results might generalize to more realistic settings. Herbst and Mas (2015) reviewed the literature on productivity spillovers and found that lab and field experiments yield strikingly similar magnitudes. Woolley et al. (2010) and Engel et al. (2014) found that average social intelligence predicts group performance, while Deming (2017) found that individual social skills increase

[3]An expert panel judged the RMET to be one of the best measures of the ability to recognize emotions in adults (Pinkham, Penn, Green, Buck, Healey, and Harvey (2014)), and group average scores on the RMET have been shown to predict team performance across a range of tasks (Woolley et al. (2010)).

earnings and lead to sorting into teamwork-intensive jobs. Several studies highlight the role of individual scientists in team production of research (Azoulay, Graff Zivin, and Wang (2010), Oettl (2012), Jaravel, Petkova, and Bell (2018)). Arcidiacono, Kinsler, and Price (2017) and Devereux (2018) estimated individual spillovers onto team performance in professional sports, while many other studies investigate the contribution of teamwork and team-specific capital to team performance (e.g., Wuchty, Jones, and Uzzi (2007), Neffke (2019)).

Our lab tasks are relatively simple, requiring only basic coordination among teammates, and there is almost no scope for repeated interactions. If anything, the lab results might understate the full impact of being a team player. Nonetheless, we find that the team player effect is about 65 percent as important as individual skills in explaining group performance. We also find that social skills have roughly the same predictive power as IQ for team success. This suggests that the individual assessments used in nearly all educational and employment settings miss a lot of information about worker productivity. To identify good team players, you must measure performance in group settings.

The remainder of the paper proceeds as follows. Section 2 describes the experiment and the data. Section 3 outlines our measurement framework. Section 4 presents our main, pre-registered results. Section 5 explores mechanisms, and Section 6 concludes.

2. DESCRIPTION OF EXPERIMENT AND DATA

2.1. *Overview of Experiment*

Our experiment had two phases, summarized in Figure 1. In the first phase, participants completed a series of online tests to measure their individual skill at three problem-solving tasks: Memory, Optimization, and Shapes. Section 2.2 describes these tasks. We also assessed participants' social intelligence / emotional perceptiveness (using a shortened version of the Reading the Mind in the Eyes Test, described in Baron-Cohen et al. (2001))[4] and personalities (using a short version of the Big 5 inventory, from Goldberg (1992)).

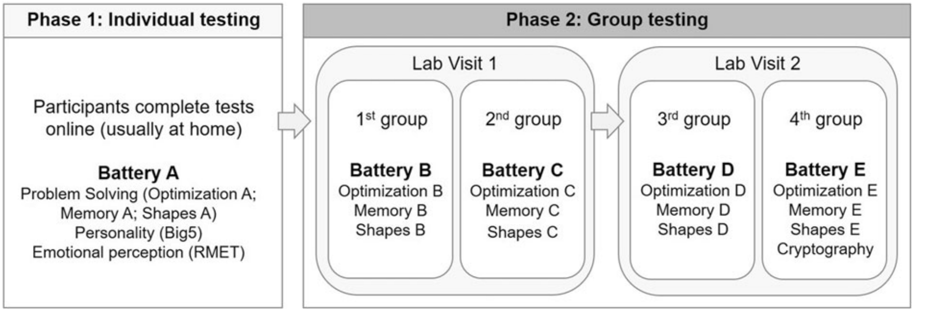


FIGURE 1.—Overview of experiment. *Note*: This figure describes the experimental design from an individual participant's perspective. Tasks are described in Section 2.2. Task batteries represent unique sequences of tasks. Participants never see the same exact task more than once. Lab visits involve 6, 9, or 12 participants, who were randomly allocated to groups of 3 people—see Sections 2.3 and 2.4 for details.

[4]To limit the length of our test battery, we included 26 of the 36 original items. The items we removed were an equal balance of male and female faces.

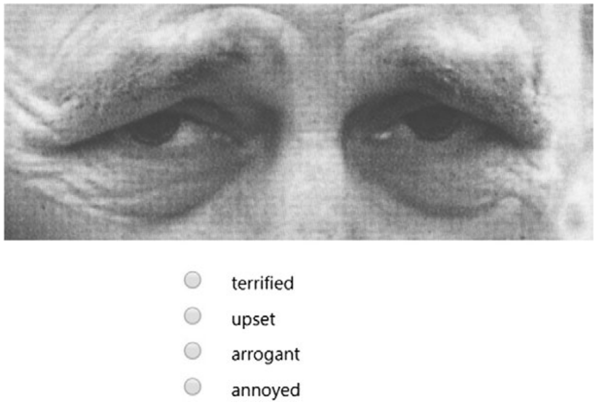


FIGURE 2.—Example item from the Reading the Mind in the Eyes Test (RMET). *Note*: This is an example item from the Reading the Mind in the Eyes Test (RMET), a well-established and psychometrically validated test of emotion recognition and social intelligence (Baron-Cohen et al. (2001)). We administered a 26-item version of the test. The mean score on the test was 18.5 out of 26. The answer to the question above is "upset."

The Reading the Mind in the Eyes Test (RMET) measures participants' ability to recognize emotions in others and, more broadly, their "theory of mind" (i.e., their ability to reason about the mental state of others (Baron-Cohen et al. (2001))). Relative to other measures of social intelligence, the main value of the RMET is that it has right and wrong answers, has relatively high test-retest reliability, and can be quickly and reliably administered (Pinkham et al. (2014)). The test presents participants with photos of faces, cropped so that only the eyes are visible (see example in Figure 2). For each set of eyes, participants are asked to choose which emotion, from four options, best describes the person in the image. We made definitions of all the words available via links to an online dictionary.

Lab participants were also assessed on three dimensions of the Big 5 personality inventory that are positively associated with group performance in other studies—Conscientiousness, Extraversion, and Agreeableness (Bell (2007)). Conscientiousness is often used as a measure of "non-cognitive" skills in economics and is positively associated with employment and earnings (e.g., Almlund et al. (2011), Heckman and Kautz (2012)). We administered the 10-item version of each personality sub-scale, based on Goldberg (1992) and available at IPIP (2018).

The second phase of the experiment focused on testing participants in groups. Participants came to the lab and were randomly assigned to groups of three people. Each group completed a collective version of the individual problem-solving tasks. Participants visited the lab twice. During each visit, participants worked in two separate groups. Over the course of the experiment, participants were allocated to a total of four groups. The average time between the individual assessment and the first lab visit was 11 days. The average time between the first and second lab visit was also 11 days.

To ensure that participants never saw the same problem twice, we incorporated five different versions of each task type and deterministically grouped them into "batteries" A through E, as shown in Figure 1.

Each group of three people worked face-to-face in a single room. The tasks were computer based and each participant was provided with a laptop. We began each group session by asking group members to introduce themselves. Then, groups were required to nominate a "Reporter." The Reporter was responsible for entering their group's answers.

Participants also gathered around the Reporter's laptop for some tasks.[5] Before each problem-solving task, groups were prompted to discuss their strategy. In batteries B and D, groups completed practice versions of each task.

2.2. *Individual and Group Tasks*

We chose tasks to satisfy three criteria. First, tasks must be feasible to administer to both individuals and groups, with only minor modifications between the individual and group versions. This enabled us to estimate group performance *controlling for individual task-specific skill*. Second, tasks needed to be objective in the sense that we could easily rank performance across individuals and groups. Third, since we are interested in studying teamwork, we looked for tasks where cooperation among group members would plausibly improve performance.

The three tasks we use to estimate our "team player" effects—Optimization, Memory, and Shapes—meet each of these three criteria, as does the Cryptography task, which we use for out-of-sample validation. We provide only a short description of the tasks here, with more details available in the Supplemental Material (Weidmann and Deming (2021)).

The Optimization task requires participants to find the maximum of a complex function by entering guesses and having outputs returned to them. Participants received 15 guesses in the individual stage. In the group stage, each participant received 5 guesses which they entered on their own laptop, and then the Reporter was asked to enter the group's final answer. Each group solved the Optimization task twice, and each time they received a new function.

The Memory task asked individuals in phase one to memorize target lists of words, images, and stories over periods of 20–40 seconds. After the memorization period for each sub-task, individuals were presented with new stimuli and asked to identify which, if any, were in the target list (in the case of stories, participants were asked questions about the content of the stories). Once the participants completed the three individual memory tasks, we provided them with feedback about their results, including on which sub-tasks they scored highest. Our goal was to provide them with information that they could use to divide up sub-tasks in the group phase. In the group phase, we gave each team 40 seconds to *collectively* remember words, images, and stories, and we prompted the groups to discuss their strategy before the memorization period began. Each group member viewed their own laptop and could view any of the three stimuli at any point during the 40-second period. At the end of the group memory task, all three team members collectively answered a set of 24 questions (8 for each sub-task).

The Shapes task was a modified version of two well-established measures of fluid intelligence: the Culture Fair Intelligence Test (CFIT, Scale 3) and the Raven's Advanced Progressive Matrices. In the individual testing phase, participants completed 14 items, which emphasize pattern recognition and spatial reasoning. In the group version, all three team members gathered around a single laptop and collectively decided on the group's answer for each item. No individual or group received a perfect score.

Finally, we chose a fourth task—Cryptography—to provide out-of-sample validation of the team player effect. The Cryptography task is a decoding problem in which each letter

[5]We deliberately framed the role of the Reporter as one in which people follow a 'collaborative' rather than a 'consultative' approach to help facilitate teamwork. In a pre-specified secondary analysis, we examined whether there was a relationship between being nominated as a Reporter and the Team player index. We found no evidence of an association between the team player index and whether someone was nominated to be the Reporter for the group.

from A to J represents a unique number from 0 to 9. Groups were asked to decode the value of each letter by entering mathematical expressions that would return an output. The goal of the task is to find the value of each letter in the fewest number of steps.

Figures A.1–A.4 in the Supplemental Material present a visual depiction of each of our tasks, and the Supplemental Material provides additional detail on the tasks themselves. In each case, individuals could perform the tasks on their own, but working together with their team members in the second phase was important for improving performance.

2.3. *Recruitment and Sample*

We recruited our sample from the Harvard Decision Science Lab participant pool. The pool comprised 41% undergraduate students, 25% graduate students, and 34% non-students. There were two exclusion criteria: participants needed to be under 60 years old, and to be fluent in English. These restrictions were based on several pilot sessions and were intended to minimize the risk of floor effects with our tasks. Participants who completed the study were paid a total of $100: $10 for completing the individual tests; $30 for lab visit 1; and $60 for lab visit 2. We elected not to explicitly pay groups for better performance, relying instead on intrinsic motivation through priming.

Table I shows descriptive statistics of the study population. Relative to the United States, our final analysis sample was younger, more female, and less white. Four hundred thirty-four participants successfully completed the individual tests.[6] Three hundred thirty-two of these participants attended the first lab session. Of these, 274 came to a second lab session. We excluded from the sample 19 groups where at least one participant was ineligible due to lab error, or other technical issues. Thus, the final sample was the 255 participants who were observed in at least 3 successful groups. Figure A.5 of the Supplemental Material presents a simple participant flow diagram.

TABLE I

DESCRIPTIVE STATISTICS.

	Study Sample ($n = 255$)	U.S. Population
Age (median)	26	38
Female%	57%	51%
Latino/Hispanic%	13%	18%
Black%	11%	13%
Asian%	27%	6%
White%	36%	60%

Note: The table presents descriptive statistics for the study population (Column 1), relative to the U.S. population (Column 2). Data for the U.S. population are taken from the 2017 American Community Survey and 2018 estimates from the U.S. Census Bureau.

[6]In addition to these 434 participants, 152 people completed some part, but not all, of the online tasks. Nine participants completed all the tasks but were ineligible for the study, based on a series of attention checks. These were included because participants completed the individual tasks outside of the lab, and we wanted to ensure that participants carefully read the instructions.

2.4. *Randomization*

Lab sessions typically consisted of $n = 9$ or $n = 12$ participants.[7] After participants completed the individual tasks, they were immediately eligible to sign up for group sessions. When participants did not divide evenly into three-person groups due to no-shows, "extra" participants were paid a small show-up fee and asked to come back again. Lab sessions were evenly spaced between August and November of 2019, and participants signed up 6 days ahead on average.

The session signup process was haphazard but not explicitly randomized. However, there is no clear pattern of average participant scores on the RMET or Ravens over time. Reassuringly, when we compare group means in terms of average Ravens or RMET scores, we fail to reject the null hypothesis that they are equal ($p = 0.72$ and $p = 0.30$, respectively).

We randomly assigned participants to teams, but with two deviations from simple randomization. First, we wanted to maximize the importance of any team player effect by creating groups with similar levels of individual skill. Second, we wanted to minimize scenarios where the same people worked together multiple times, so that our results would not be contaminated by familiarity among teammates.

At the start of each session, we conducted a blocked randomization procedure. Participants were ordered according to their mean performance across the individual problem-solving tasks (Memory, Optimization, and Shapes). From this ordering, we formed three blocks: higher-skill; medium-skill; lower-skill. Each block had the same number of participants.

We then randomly generated groups of three people. Each group had a member from each block. Participants randomly drew balls from bags, under the supervision of the experimenters. For example, in a session with $n = 9$ participants, groups from the higher-skill block each randomly drew a ball from the set {A, B, C}; participants from the medium-skill block randomly drew a selection from the set {D, E, F}, and the participants from the lower-skilled block randomly drew a ball from the set {G, H, I}. During each lab session, participants were randomly assigned to two groups. Each participant's randomly assigned letter defined both of their groups. Figure A.6 of the Supplemental Material provides an example of the randomization scheme.

3. IDENTIFICATION STRATEGY

This section describes our conceptual framework and empirical approach. Our analysis strategy was pre-registered at the AEA RCT registry.[8] Deviations are noted in the footnotes.

Let individuals be indexed by $i = 1, \dots, n$. We allocated individuals to groups of three people, where groups are indexed by g, with n_g groups. Let I_g^i be an indicator of whether participant i is in group g. I_g^i is a vector of length n_g, where

$$I_g^i = \begin{cases} 1 & \text{if } i \text{ is in } g, \\ 0 & \text{otherwise.} \end{cases}$$

[7]From a total of 343 groups, 147 were formed in sessions of 12 participants; 145 were formed in sessions of 9; 51 were formed in sessions of 6.

[8]The trial number is AEARCTR-0002896.

Next, we have a set of variables describing task performance. Let X_{ik} denote the performance of individual i on task type $k \in$ {Optimization; Memory; Shapes}. Similarly, let G_{gk} denote the performance of group g on task. We rescale group scores G_{gk} for each task to account for potential differences in task difficulty. Let b indicate task battery, $b \in \{B, C, D, E\}$. Batteries B, C, D, and E each include three tasks: Optimization, Memory, and Shapes.[9]

Rescaled scores are calculated as $\tilde{G}_{gkb} = \frac{G_{gkb} - \hat{\mu}_{kb}}{\hat{\sigma}_{kb}}$, where $\hat{\mu}_{kb}$ and $\hat{\sigma}_{kb}$ are the sample mean and standard deviation for task k in battery b.[10] $\tilde{G}_{gk}$ is our main measure of group performance.

Some groups may perform better on tasks purely because groups have different endowments of task-specific skill. Consider the following model for how well group g performs on task k:[11]

$$\begin{aligned} \tilde{G}_{gk} &= \alpha_k \sum_i I_g^i X_{ik} + \epsilon_{gk}, \\ \epsilon_{gk} &\sim N(0, \sigma_G^2). \end{aligned} \tag{1}$$

The term $\alpha_k \sum_i I_g^i X_{ik}$ measures group g's endowment of individual skill on task type k. The individual scores X_{ik} come directly from the tests administered to participants in phase 1 of the experiment, as shown in Figure 1.

Define T_g as a measure of *group-level* performance, adjusted for differences in individual task-specific skill. The residuals $\hat{\epsilon}_{gk}$ from equation (1) provide an estimate of whether each group under- or over-performed on task k relative to the prediction based on task-specific skills. Averaging this residual performance across tasks gives us

$$\hat{T}_g = \frac{1}{3} \sum_k \hat{\epsilon}_{gk}. \tag{2}$$

With only a single randomization, it is impossible to determine whether variation in $\hat{T}_g$ arises from unmeasured *individual* attributes of team members, or from group dynamics. However, with repeated random assignment, we can assess whether $\hat{T}_g$ is correlated for individuals as they join different teams. For each participant, we estimate the *team player index* $\hat{\beta}_i$ as the average $\hat{T}_g$ across all groups that i participated in (up to 4):

$$\hat{\beta}_i = \frac{1}{4} \sum_g I_g^i \hat{T}_g. \tag{3}$$

In our framework, $\hat{\beta}_i$ is an estimate of the causal contribution of individual i to team performance. With enough randomizations, we could precisely estimate β_i for each participant. However, with only four team assignments, $\hat{\beta}_i$ is relatively noisy at the individual

[9]Battery E also includes Cryptography, which is our validation task.

[10]After rescaling, we suppress the b subscript for clarity.

[11]Our pre-analysis plan also included in equation (1) indicators for whether group g contained participants who knew each other from outside the experiment (5 percent of the sample) and for whether groups contained participants who had previously been assigned to the same team by chance (41 percent of the sample). Neither of these nuisance controls has any substantive impact on the main results, so we dropped them from the main specification for clarity. The results with these variables included are presented in Supplemental Material Table A.I. We thank the editors for this suggestion.

level. Thus, we focus on σ_β, the standard deviation of the β estimates. We estimate σ_β using a multilevel model:[12]

$$\begin{aligned} \hat{T}_{gi} &= \beta_i + e_{gi}, \\ \beta_i &\sim N\big(0, \sigma_\beta^2\big), \\ e_{gi} &\sim N\big(0, \sigma^2\big). \end{aligned} \tag{4}$$

In equation (4), $\hat{T}_{gi}$ is a 1 x n_g vector of skill-adjusted group performance scores, from equation (2). We include the i subscript to indicate that this part of the analysis examines variation at the level of individual participants. $\hat{T}_{gi}$ is constant for all members of group g. β_i is a random effect for individual i on group g and e_{gi} is residual error.[13]

We fit model (4) to estimate the team player effect ($\hat{\sigma}_\beta$) and evaluate our results against the null that this effect—conditional on individual skill, as in equation (1)—is equal to zero. Because of the unusual nature of our repeated random assignment procedure, we elected in our pre-analysis plan to calculate p-values using randomization inference. However, we also report results using a normal approximation and a profile likelihood procedure. Profile likelihood confidence intervals are based on the chi-squared distribution of the log-likelihood ratio test statistic, and thus may be a better fit in this small, non-normal sample (Venson and Moolgavkar (1988)).

The randomization inference proceeds in four steps. First, we control for individual skill, by fitting model (1). Second, we simulate five thousand allocations of individuals to groups, blocking on task battery so that, in every simulated allocation, we observe each participant the same number of times as we did in the actual experiment. Third, we fit model (4) for each simulation and estimate $\sigma^2_{\beta(\mathrm{NULL})}$. Fourth, we compare our observed team player effect to the simulated distribution under the null, calculating how often the null distribution provides a more extreme value than $\hat{\sigma}_\beta$, that is, $P(\sigma_{\beta(\mathrm{NULL})} > \hat{\sigma}_\beta)$. This is our p-value (Ernst (2004)).

4. MAIN RESULTS

In this section, we report results from our pre-specified models only, and we explicitly note any deviation from the pre-analysis plan. Section 5 reports post hoc exploratory analyses and evidence for mechanisms.

[12]This represents a slight deviation from our pre-registered analysis plan, in which we planned to estimate σ^2_β as $\hat{\sigma}^2_\beta = \hat{\mathrm{var}}(\hat{\boldsymbol{\beta}})$, where $\hat{\boldsymbol{\beta}}$ is a $(1 \times N)$ vector of team player estimates from equation (3). However, $\hat{\mathrm{var}}(\hat{\boldsymbol{\beta}}) = \hat{\sigma}^2_{\hat{\beta}}$, so reporting this figure would overstate the magnitude of the team player effect. This motivates the use of model (4). Supplemental Material Figure A.7 shows that the distribution of the raw $\hat{\beta}_i$'s from equation (3) closely matches a normal approximation, which justifies the normality assumption. We also note that our model implies e_{gi} are uncorrelated within group g, a simplification that is unlikely to hold in all cases.

[13]We can also obtain estimates of the team player effect in a single step using individual fixed effects plus controls for task-specific skills, for example, Model 1(b): $\tilde{G}_{gk} = \alpha_k \Sigma_i I^i_g X_{ik} + \Sigma_i I^i_g \beta_i + \epsilon_{gk}$. We could also use a single-step approach to estimate, for example, Model 2(b): $\tilde{G}_{igk} = \alpha_k \Sigma_i I^i_g X_{ik} + \beta_i + \epsilon_{igk}$; $\beta_i \sim N(0, \sigma^2_\beta)$; $\epsilon_{igk} \sim N(0, \sigma^2_g)$. These single-step procedures allow for a potential correlation between X_{ik} and β_i. Table A.I and Figure A.8 of the Supplemental Material show that the results are extremely similar using these approaches.

4.1. *Are Some People Good Team Players?*

Table II reports estimates of the team player effect: the standard deviation of the team player index $\hat{\sigma}_\beta$. Below the estimate of the team player effect, we report p-values for each inference method. Coefficient standard errors are presented in parentheses below control variables.

In our preferred model, the variance of the team player effect is 0.129 standard deviations. The coefficients on each of the task-specific skills in $\sum X_{ik}$ are highly statistically significant ($p < 0.01$), suggesting that a team's endowment of task-specific skill is a strong predictor of group performance. The average magnitude of the task coefficients is about 0.2, which suggests that the contribution of the team player effect (0.129) is worth about 65 percent as much as individual task-specific skill.

Column 1 reports p-values from each of our three different approaches to developing confidence intervals. Using randomization inference, the actual team player estimate exceeds the simulated estimates in more than 97 percent of the cases ($p = 0.026$). The p-value is similar when we compute confidence intervals using profile likelihood ($p = 0.034$). The standard normal approximation yields much tighter confidence intervals (p < 0.001).

Column 2 of Table II estimates the team player effect with no controls at all, which increases the magnitude of $\hat{\sigma}_\beta$ to 0.244. We do not think of our estimate of $\hat{\sigma}_\beta$ in column 2 as the impact of being a good team player, but rather the *total* causal impact on group performance of receiving a more talented teammate. In that sense, comparing column 1 to column 2 suggests that about half of the variance in an individual's measured causal contribution to the team can be explained by their individual skill in the task at hand. Column 3 replaces task-specific controls (X_{ik}) with IQ, as measured by the Ravens test.

TABLE II

ARE SOME PEOPLE GOOD TEAM PLAYERS?

	Dependent Variable: Group Performance $\tilde{G}_{gk}$		
	(1)	(2)	(3)
Teamplayer Effect $\hat{\sigma}_\beta$	0.129	0.244	0.175
(randomization inference)	($p = 0.026$)	($p < 0.001$)	($p = 0.001$)
[profile likelihood]	[$p = 0.034$]	[$p < 0.001$]	[$p < 0.001$]
{normal approximation}	{$p < 0.001$}	{$p < 0.001$}	{$p < 0.001$}
Task-specific skills			
Memory°	0.166		
	(0.032)		
Optimization°	0.125		
	(0.031)		
Ravens (Shapes)°	0.302		0.161
	(0.030)		(0.018)
Number of groups	343	343	343
Number of participants	255	255	255

Note: °Indicates group-level sum. "Task-specific skills" means that $\tilde{G}_{gk}$ is conditioned on the mean performance of group $g's$ individuals on task k, that is, $\overline{X}_{gk} = \frac{1}{3}\sum_i I_g^i X_{gk}$. This gives us three parameters, one each for shapes, optimization, and memory. Covariate coefficients have standard errors in parentheses. Estimates of the team player effect ($\hat{\sigma}_\beta$) have p-values from: randomization inference (in parentheses); profile likelihood [in brackets]; and Wald test {in braces}; the null hypothesis being tested is that $\sigma_\beta = 0$.

Controlling for general IQ rather than task-specific skills increases the magnitude of the team player effect from 0.129 to 0.175, suggesting that task-specific skills have additional predictive power.

A more general issue is that measurement error in individual skills might lead us to overestimate the team player effect. We account for the possibility of measurement error in three ways. First, Table A.II of the Supplemental Material presents additional results where we control for a group's endowment of individual skills in a variety of ways—including the maximum and minimum of task-specific skill, RMET, and IQ, along with flexible functions of individual skill (rather than linear relationships). None of our results are sensitive to these different specifications. In a fully flexible "kitchen sink" prediction of group performance, we still estimate a team player effect of 0.120 ($p = 0.044$).

Second, we estimate a modified version of our main analysis that controls for all three measures of individual skills in equation (1), which implicitly assumes that individual contributions to group performance are determined by a single latent construct that is noisily measured by each individual task.[14] This attenuates the team player effect slightly, to 0.106, although it remains statistically significant at the 10 percent level. However, in this case, we must also account for measurement error in the individual skills themselves. Table A.IV of the Supplemental Material shows results of simulations where different amounts of measurement error are added to the individual task scores. If we adjust for measurement error using the correlation between test items for an individual (e.g., Cronbach's α, which averages 0.63 across the three tests), we obtain a team player effect of 0.159 ($p = 0.013$), slightly larger than our baseline results.

Third, we validate the existence of the individual team player index $\hat{\beta}_i$ by asking whether it predicts group-level performance on a fourth out-of-sample task, Cryptography. We chose the Cryptography task for out-of-sample validation because the literature on teamwork has shown that it rewards teamwork in the sense that groups typically perform better than the sum of their parts (Larson (2010)).[15] Eighty-five groups completed the Cryptography task, always in the last task battery (Battery E).

Unlike the other three group tasks, Cryptography performance was not used to estimate the team player effects shown in Table II. Rather, we regress the group's Cryptography score on the group's average team player index $\overline{\hat{\beta}_i}$, as specified in our pre-analysis plan. We also estimate the correlation between Cryptography task performance and other group characteristics such as average IQ and average RMET, as well as other combinations such as the maximum or minimum of each measure.

The results are in Table III. Column 1 presents results from our pre-specified model, a bivariate correlation between Cryptography task performance and the team's average team player index $\overline{\hat{\beta}_i}$. The correlation is positive—a 1-standard-deviation increase in $\overline{\hat{\beta}_i}$ increases Cryptography task performance by 0.145 standard deviations—but it is not statistically significant ($p = 0.185$).

Columns 2 and 3 show the same results, except with $\hat{\beta}_{\max}$ and $\hat{\beta}_{\min}$ as the predictors, respectively. The maximum team player index in a group predicts similarly well to the

[14]Table A.III of the Supplemental Material shows that for each group task, individual skill *in that task* is a better predictor of group performance than individual skill in the other two tasks. This is evidence against the notion that we are capturing a single underlying factor of "being good at problem-solving."

[15]According to Larson (2010), the primary advantage of group work in the Cryptography task is that some team members can execute the current strategy (e.g., figuring out what the next equation should be given the output of the current equation) while others can consider new strategies. It is extremely challenging for individuals to simultaneously execute a strategy and consider a new one, perhaps due to attention and working memory constraints (Larson (2010)).

TABLE III

VALIDATING THE TEAM PLAYER EFFECT WITH OUT-OF-SAMPLE TASK PERFORMANCE.

	Dependent Variable: Score on Cryptography Task C_{gk}					
	(1)	(2)	(3)	(4)	(5)	(6)
Team player index (β)						
Group Mean	0.145					
		(0.109)				
Group Max	0.162					
		(0.108)				
Group Min			0.084			
			(0.109)			
Group contains someone with $\beta > 1\sigma$				0.223		
				(0.107)		
RMET group mean (mean_g RMET_i)					0.192	
					(0.108)	
Ravens group mean (mean_g Ravens_i)						0.217
						(0.107)

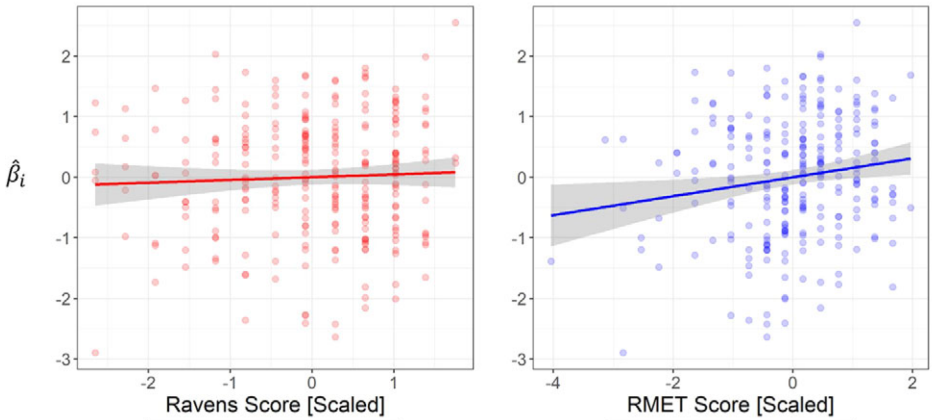
Note: Results from the Cryptography task were available for $N = 85$ groups. See Section 2.2 and the Supplemental Material for details about the task. The team player index comes from our pre-registered model described in Section 3. Ravens is a well-established measure of IQ or fluid intelligence. RMET is the Reading the Mind in the Eyes Test, a well-established test of emotion perception and social intelligence. Coefficients have standard errors in parentheses. All variables were standardized to have mean = 0 and sd = 1. The results in Columns 1, 5, and 6 are from our pre-registered analysis plan.

mean (0.162, $p = 0.140$) while the minimum performs worse (0.084, $p = 0.443$). Column 4 reports the correlation between having a good team player (defined as someone whose score is $>1\sigma$ above average) and a group's Cryptography score. This is a relatively strong predictor (0.223, $p = 0.041$). Column 5 shows the correlation between a group's average RMET score and Cryptography task performance. A 1-standard-deviation increase in mean RMET score increases performance by 0.192 standard deviations ($p = 0.079$). Column 6 shows a positive and statistically significant relationship with group average IQ (0.217, $p = 0.047$). Overall, we cannot reject the hypothesis that all these results are equal to each other (with the exception of the minimum team player score in Column 3) and all of them positively predict group performance out-of-sample. In Section 5.1, we explore combining these measures together to maximize the predictive power of individual characteristics for group performance.

4.2. *What Predicts Being a Good Team Player?*

Figure 3 presents two scatterplots, where each dot is an individual, and the team player index $\hat{\beta}_i$ is on the vertical axis. Below each figure, we show the bivariate correlations and p-values from our pre-specified model estimate of $\hat{\beta}_i$ (the first column), and a model showing the total effect with no controls (second column). The scatterplot always shows results from the first column.

The left panel of Figure 3 shows results for IQ. There is no significant association between being a good team player and IQ ($\hat{\rho} = 0.045$, $p = 0.468$) in our baseline preregistered model. However, IQ is strongly correlated with $\hat{\beta}_i$ when group performance is estimated with no controls ($\hat{\rho} = 0.374$, $p < 0.001$). An individual's *total* causal impact on group performance is strongly related to IQ, but nearly all of the impact of IQ is mediated by task-specific individual skill.



(1)	(2)
0.045	0.374
(p=0.468)	(p<0.001)

(1)	(2)
0.156	0.300
(p=0.012)	(p<0.001)

FIGURE 3.—The Team Player Index is correlated with Social Intelligence, but not IQ. *Note*: Each panel presents a scatterplot of an individual's estimated team player index $\hat{\beta}_i$ against their individual Ravens score (left panel) and their individual RMET score (right panel). In both cases $\hat{\beta}_i$ shown in the figures is estimated based on the model in equations (1) through (4), as described in Section 3 and detailed in our pre-registered analysis plan. Ravens is a well-established measure of IQ or fluid intelligence. RMET is the Reading the Mind in the Eyes Test, a well-established test of emotion perception and social intelligence. Beneath each panel, we show coefficients from two different estimates of $\hat{\beta}_i$: (1) our pre-specified model, with controls for task-specific skills and indicators for group familiarity; (2) no controls. See the text for details. The scatterplot always shows estimates from (1), which comes from equation (3) in the paper. The same sample was used for all analysis: 1029 group-task observations, 343 groups, 255 participants.

The right panel of Figure 3 shows the correlation between $\hat{\beta}_i$ and an individual's score on the RMET. We find a clear and statistically significant correlation between being a good team player and emotional perceptiveness ($\hat{\rho} = 0.156$, $p = 0.012$). This contrasts notably with IQ and suggests that RMET adds substantial predictive power beyond the impact of individual task-specific skill. The correlation between RMET and $\hat{\beta}_i$ is even stronger when group performance is estimated with no controls ($\hat{\rho} = 0.300$, $p < 0.001$).

Table IV shows correlations between the team player index and demographic characteristics, as well as three of the Big 5 personality factors (Agreeableness, Conscientiousness, and Extraversion). We find no evidence of gender differences in $\hat{\beta}_i$.[16] We also find no statistically significant association between the team player index and age, race/ethnicity, or years of completed education, although our standard errors are too large to rule out small differences for any of these characteristics. Finally, we also find no relationship between personality scores and the team player index. Figure A.9 of the Supplemental Material presents scatterplots of the relationship between each personality score and the team player effect.

Tables A.V, A.VI, and A.VII of the Supplemental Material report additional results from our pre-analysis plan, including variation by task type and associations between team

[16]This contrasts somewhat with Woolley et al. (2010), who found that teams with more women perform better on group tasks. In Woolley et al. (2010), gender differences in group performance were mediated by gender differences in performance on the RMET. However, we find only a small difference in RMET scores by gender, with women scoring 18.7 on average compared to 18.0 for men ($p = 0.11$).

TABLE IV

THE TEAM PLAYER INDEX IS UNCORRELATED WITH DEMOGRAPHIC CHARACTERISTICS.

	Dependent Variable: Team Player Index $\hat{\beta}_i$				
	(1)	(2)	(3)	(4)	(5)
Female	0.001				
	(0.063)				
Years of education		0.031			
		(0.063)			
Age			−0.073		
			(0.063)		
Under-represented minority°				−0.092	
				(0.146)	
Personality					
Agreeableness					0.072
					(0.069)
Conscientiousness					0.017
					(0.065)
Extraversion					−0.023
					(0.067)
Observations	252	254	252	254	255

Note: Each column presents bivariate correlations between estimates of $\hat{\beta}_i$ from our pre-registered model described in Section 3 and the indicated demographic characteristics. °Under-represented minorities are participants who identified as African-American, Latino/Hispanic, or Native American. Column 5 reports results for the Agreeableness, Conscientiousness, and Extraversion scales of the Big 5 Personality inventory. Standard errors are presented in parentheses. All variables were standardized to have mean = 0 and sd = 1.

performance and diversity. We find no statistically significant relationships along any of these dimensions.

Overall, we find strong evidence that—holding task-specific individual skills constant—some people consistently improve their group's performance. These individuals are good "team players." The one characteristic that consistently predicts who will be a good team player is that individual's score on the RMET, a well-established test of social intelligence and emotional perceptiveness.

5. MECHANISMS

The results are consistent with several different potential mechanisms, and our experiment was not designed to disentangle them. Team players might increase effort among teammates, improve communication about comparative advantage in team production, or encourage better group problem-solving through improved dialogue and integrative thinking. Moreover, these explanations are not mutually exclusive, and team players may improve group performance in multiple ways.

5.1. *Social Skills*

Across all individual characteristics, only the Reading the Mind in the Eyes Test (RMET) score predicts whether someone is a team player. The RMET was originally designed by autism researchers to diagnose deficits in the capability to reason about the mental state of others (Baron-Cohen et al. (2001)). However, it also has explanatory power in a general, non-impaired population and has become a well-established measure of social intelligence (Baron-Cohen et al. (2001), Baker et al. (2014)).

TABLE V

PREDICTIVE POWER OF SOCIAL SKILLS AND OTHER FACTORS ON THE CRYPTOGRAPHY TASK.

	Dependent Variable: $\hat{C}_g$ (Cryptography Score)			
	(1)	(2)	(3)	(4)
$\overline{\text{RMET}}$	0.192			
	(0.108)			
S (Social skills)				
Group mean		0.214		0.171
		(0.107)		(0.110)
Ravens (IQ)				
Group mean			0.217	0.175
			(0.107)	(0.110)
Observations	85	85	85	85
R^2	0.037	0.046	0.047	0.074
Adjusted R^2	0.025	0.034	0.035	0.052

Note: Results from the Cryptography task were available for $N = 85$ groups. See Section 2.2 and the Supplemental Material for details about the task. $\overline{\text{RMET}}$ is the group's average score on the Reading the Mind in the Eyes Test, a well-established test of emotion perception and social intelligence. S (Social Skills) is the average of each participant's mean standardized score on the RMET and their estimated team player index. To avoid double-counting, the team player index used in this average controls from group endowments of RMET. Ravens is a well-established measure of IQ or fluid intelligence. Covariate coefficients have standard errors in parentheses. All variables were standardized to have mean $= 0$ and sd $= 1$.

Since effective teamwork probably requires individuals to read their teammates' emotional states, it is sensible that the RMET would predict whether someone is a good team player. Group average scores on the RMET have been shown to predict team performance across a range of tasks (Woolley et al. (2010), Engel et al. (2014)). Table A.VIII of the Supplemental Material shows that the correlation between the RMET and the team player index $\hat{\beta}_i$ persists after controlling for an increasingly rich set of individual characteristics.

This suggests that the team player index $\hat{\beta}_i$ and the RMET might both measure the same underlying construct, which we call *social skills* (S_i). To explore this construct, we average each participant's team player effect $\hat{\beta}_i$ and RMET score,[17] and test the predictive power of this composite measure $\hat{S}_i$ on group performance in the Cryptography task.

The results are in Table V. The first column repeats the results from Column 5 of Table III, which shows the bivariate correlation between Cryptography task performance and the group's mean score on the RMET (0.192, $p = 0.079$). Column 2 shows the correlation between Cryptography performance and the group's average score on the social skills composite $\overline{\hat{S}_i}$. A 1-standard-deviation increase in $\overline{\hat{S}_i}$ is associated with a 0.214-standard-deviation increase in the Cryptography task ($p = 0.050$). Column 3 replaces $\overline{\hat{S}_i}$ with $\overline{\text{IQ}_i}$. The coefficient on group average IQ is 0.217 ($p = 0.047$), which suggests that average IQ and average social skills are about equally predictive of group performance on an out-of-sample validation task.[18] Column 4 shows that $\overline{\hat{S}_i}$ and $\overline{\text{IQ}_i}$ are about equally predictive when included in the same specification, and adding each of them increases the R-squared of the prediction by about 60 percent relative to only including one. This

[17]To avoid double-counting RMET, we use $\hat{\beta}_i$ from a model that conditions on group average RMET.

[18]The results are very similar when we substitute the max for the mean, or when we add additional controls such as personality scores.

suggests that each makes an independent contribution to predicting group performance. Social skills have substantial predictive power for group performance, about the same as IQ.

5.2. *Effort and Motivation*

Good team players may increase group performance by encouraging their teammates to increase effort. We test this in two ways. First, we examine group-level variation in one measure of effort—whether the team used their full allotted time for a task.

Eighty-two percent of groups took their full allotment of time in the Shapes task. Of the groups that finished before time, 10 percent "rushed," which we define arbitrarily as groups who submitted answers with more than 15 seconds to spare. There is a negative association between "rushing" and performance. Groups that "rushed" answered 54 percent of items correctly, compared to 63 percent among non-rushed groups ($p = 0.005$). Groups that had a "good team player"—an individual with $\hat{\beta}_i > 1\sigma$ above the average—rushed 5 percent of the time, compared to 13 percent for all other groups ($p = 0.02$). We also find that group average endowments of $\hat{\beta}_i$, RMET, and $\hat{S}_i$ are negatively associated with rushing, as is the group's average score on the Conscientiousness personality factor (even though it is not related to overall performance). This provides suggestive evidence that team players encourage their group to exert more effort on the Shapes task.

Our second test for the importance of effort involves the impact of social skills on group performance in the Memory task, where group performance is almost always the sum of individual contributions. Recall that, in the Memory task, there were three types of stimuli—words, images, and stories. Although group answers were recorded on a single laptop at the end of the task, each member had their own laptop during the memorization period. Ninety-two percent of our groups divided responsibilities such that each member only looked at a single stimulus category for the entire memorization period.

In cases where each group member memorized stimuli that their teammates did not see, we can measure whether good team players improve their teammates' performance *despite not being directly involved in the sub-task*. We estimate the impact of being randomly assigned to a teammate with high social skills using the following model:

$$G_{gt[i]} = \alpha + \delta \overline{\text{SRMET}_{g[-i]}} + \phi X_{ti} + \gamma \,\text{Round}_g + \lambda \,\text{Type}_t + \epsilon_{tg[i]}. \tag{5}$$

$G_{gt[i]}$ is group g's average score on sub-task t (words, images, stories) of the Memory task, for groups where only one person looked at each stimulus type—which allows us to attribute the score to individual i (in brackets). $\overline{\text{SRMET}}_{g[-i]}$ is the mean social skills of individual i's randomly assigned teammates, leaving out i. X_{ti} is i's individual score on the memory sub-task t, assessed during phase 1 of the experiment. Round_g and MemType_t are controls for the ordering of the task batteries and the type of memory task, respectively, to remove practice effects and to control for baseline differences in memory task difficulty. δ is the spillover effect of being assigned to a group with a 1-standard-deviation higher mean RMET score. Because the team player effect $\hat{\beta}_i$ is estimated from our data, we use only the RMET as a measure of social skill, to ensure that our results are not mechanical.

Column 1 of Table VI shows baseline estimates of equation (5). After controlling for an individual's own skill on the Memory task, a 1-standard-deviation increase in the average RMET score of their randomly assigned teammates improves their performance by 0.092 standard deviations, an increase that is statistically significant at the less than 1 percent level.

TABLE VI

TEAM PLAYERS IMPROVE TEAMMATE PERFORMANCE ON SUB-TASKS PERFORMED SEPARATELY.

	Dependent Variable: $G_{gt[i]}$ (Memory Score)	
	(1)	(2)
Individual Memory ($X_{i,\text{memory},t}$)	0.184	0.157
	(0.030)	(0.030)
Mean RMET in i's group ($\overline{\text{RMET}}_{g[-i]}$)	0.092	0.088
	(0.030)	(0.030)
Individual RMET (RMET_i)		0.085
		(0.030)
Controls for memory battery (Round)	✓	✓
Controls for memory type (Type)	✓	✓
Observations	921	921

Note: This table presents results from a regression of the group's average score on a memory sub-task t (words, images, stories) on the average RMET score of participant i's teammates. The model is fit to the 92% of cases in which only one individual looked at each stimulus type during the memorization period. In these teams, individuals memorized separate material, yet having good teammates still improves performance. We also include controls for individual memory scores on task type t as well as fixed effects for task battery and memory type. Covariate coefficients have standard errors in parentheses. Variables were standardized to have mean = 0 and sd = 1.

Teammates with higher social skills, as measured by RMET, causally improve an individual's performance on a memory sub-task, even though their teammates view different stimuli and cannot directly help them answer the recall questions. Column 2 adds a control for the individual's own RMET score. Even though RMET is a strong independent predictor of performance, the impact of teammates' RMET only slightly decreases. In both models, we find that teammates with high social skills—as measured by RMET—improve an individual's scores on tasks that are performed independently. This strongly suggests that one mechanism for the team player effect may be increased effort and/or motivation.

5.3. *Allocative Efficiency*

Another way that social skills might affect group performance is by facilitating a more efficient allocation of group members to tasks in which they have a comparative advantage. Gains from "trading tasks" was the key mechanism in the model in Deming (2017).

We test this mechanism by again focusing on the memory task. Using individual memory sub-task scores from the first phase of the experiment, we generate an expected score for each group assuming that they adopt the most efficient allocation of people to sub-tasks. We then compare this expected score to a prediction that is based on the actual stimuli that individuals were assigned to memorize by their group. We do this in two ways. First, we create a distance measure that takes the difference between the predicted score of the optimal and actual strategies. Groups that chose the efficient allocation have a distance of zero. Second, we simply order all six possible strategies from best to worst and assign to each group the rank of the strategy they actually choose.

Teams with more efficient allocations have modestly better performance ($\sigma = 0.11$, $p = 0.06$) for the rank measure, and ($\sigma = 0.07$, $p = 0.20$) for the distance measure. However, we find no evidence that groups with higher social skill formed better strategies. The correlations between both rank and distance and the average social skills of the group $\overline{S_g}$ are both positive, but very small and not statistically distinguishable from zero.

Finally, we test the hypothesis that cognitive skill and social skill are complements in team production, which is an indirect implication of the allocative efficiency mechanism in Deming (2017). In Deming (2017), social skill improves productivity by lowering communication frictions between teammates, allowing them to more efficiently specialize in their comparative advantage(s). This implies that cognitive skill and social skill are complements, and that the interaction term would be positive in a regression where productivity (or wages) is the outcome. Table A.IX of the Supplemental Material presents results from a regression of the raw team player index (estimated without controls, as in column 2 of Table II) on cognitive skill (as measured both by Ravens and task-specific skill), social skill, and the interaction terms. We find no evidence of complementarity in any specifications.

Overall, while allocative efficiency is a potential mechanism by which social skills could affect group performance, we did not find any evidence for it. That said, given that relationship between allocative efficiency and performance was only modestly positive in the first place, this is not a well-powered test. Complementarity might be a more important mechanism as team production becomes increasingly complex, and our tasks were relatively simple, with little scope for allocative efficiency gains.

6. CONCLUSION

In this paper, we develop a new method for estimating individual contributions to team performance. We repeatedly randomly assign people to teams and find that some people consistently cause their teams to exceed predicted performance. These people are good "team players."

The team player effect is not predicted by demographic characteristics such as age, gender, education, ethnicity, or IQ. Yet it is strongly related to individual scores on the Reading the Mind in the Eyes test, a widely used measure of social intelligence. Social intelligence requires the ability to read others' emotional states, which is probably a necessary—but not sufficient—condition for being a good teammate. Social skills predict group performance about as well as IQ. This suggests that being a good team player is an important skill that is distinct from general ability.

Our results uncover a direct mechanism for the growing body of evidence on the importance of "non-cognitive" skills in the labor market (e.g., Kuhn and Weinberger (2005), Heckman, Stixrud, and Urzua (2006), Borghans et al. (2008), Almlund et al. (2011), Lindqvist and Vestman (2011), Heckman and Kautz (2012), Deming (2017)). We also find suggestive evidence that team players improve group performance by encouraging effort among their teammates.

Our experimental approach highlights one way that organizations can identify good "team players." Future work should focus on scaling up the results of our experiment and testing the viability of measuring the importance of social skills in a variety of other settings.

REFERENCES

ABOWD, J. M., F. KRAMARZ, AND D. N. MARGOLIS (1999): "High Wage Workers and High Wage Firms," *Econometrica*, 67 (2), 251–333. [2639]

ALMAATOUQ, A., M. YIN, AND D. WATTS (2020): "Collective Problem-Solving of Groups Across Tasks of Varying Complexity," Unpublished working paper (found at https://psyarxiv.com/ra9qy). [2638]

ALMLUND, M., A. DUCKWORTH, J. J. HECKMAN, AND T. KAUTZ (2011): "Personality Psychology and Economics," in *Handbook of the Economics of Education* (First Ed.), Vol. 4, ed. by E. A. Hanushek, S. Machin, and L. Woßmann. Amsterdam: Elsevier, 1–181. [2639,2641,2655]

ARCIDIACONO, P., J. FOSTER, N. GOODPASTER, AND J. KINSLER (2012): "Estimating Spillovers Using Panel Data, With an Application to the Classroom," *Quantitative Economics*, 3 (3), 421–470. [2637]

ARCIDIACONO, P., J. KINSLER, AND J. PRICE (2017): "Productivity Spillovers in Team Production: Evidence From Professional Basketball," *Journal of Labor Economics*, 35 (1), 191–225. [2640]

AZOULAY, P., J. S. GRAFF ZIVIN, AND J. WANG (2010): "Superstar Extinction," *Quarterly Journal of Economics*, 125 (2), 549–589. [2640]

BAKER, C. A., E. PETERSON, S. PULOS, AND R. A. KIRKLAND (2014): "Eyes and IQ: A Meta-Analysis of the Relationship Between Intelligence and 'Reading the Mind in the Eyes'," *Intelligence*, 44, 78–92. [2638, 2651]

BARON-COHEN, S., S. WHEELWRIGHT, J. HILL, Y. RASTE, AND I. PLUMB (2001): "The 'Reading the Mind in the Eyes' Test Revised Version: A Study With Normal Adults, and Adults With Asperger Syndrome or High-Functioning Autism," *Journal of Child Psychology and Psychiatry*, 42 (2), 241–251. [2638,2640,2641, 2651]

BELL, S. T. (2007): "Deep-Level Composition Variables as Predictors of Team Performance: A Meta-Analysis," *Journal of Applied Psychology*, 92 (3), 595. [2638,2641]

BLOOM, N., AND J. VAN REENEN (2011): "Human Resource Management and Productivity," in *Handbook of Labor Economics* (First Ed.), Vol. 4, ed. by O. Ashenfelter and D. Card. Amsterdam: Elsevier, 1697–1767. [2637]

BORGHANS, L., A. L. DUCKWORTH, J. J. HECKMAN, AND B. WEEL (2008): "The Economics and Psychology of Personality Traits," *Journal of Human Resources*, 43 (4), 972–1059. [2639,2655]

CARD, D., J. HEINING, AND P. KLINE (2013): "Workplace Heterogeneity and the Rise of West German Wage Inequality," *Quarterly Journal of Economics*, 128 (3), 967–1015. [2639]

CHARNESS, G., AND P. KUHN (2011): "Lab Labor: What Can Labor Economists Learn From the Lab?" in *Handbook of Labor Economics* (First Ed.), Vol. 4, ed. by O. Ashenfelter and D. Card. Amsterdam: Elsevier, 229–330. [2639]

CORNELISSEN, T., C. DUSTMANN, AND U. SCHÖNBERG (2017): "Peer Effects in the Workplace," *American Economic Review*, 107 (2), 425–456. [2637,2639]

DEMING, D. J. (2017): "The Growing Importance of Social Skills in the Labor Market," *Quarterly Journal of Economics*, 132 (4), 1593–1640. [2637-2639,2654,2655]

DEVEREUX, K. (2018): "Identifying the Value of Teamwork: Application to Professional Tennis," Working Paper Series No. 14, Canadian Labour Economics Forum. [2640]

DRISKELL, J. E., E. SALAS, AND T. DRISKELL (2018): "Foundations of Teamwork and Collaboration," *American Psychologist*, 73 (4), 334–348. [2638]

DRISKELL, J. E., E. SALAS, AND S. HUGHES (2010): "Collective Orientation and Team Performance: Development of an Individual Differences Measure," *Human Factors*, 52 (2), 316–328. [2638]

ENGEL, D., A. W. WOOLLEY, L. X. JING, C. F. CHABRIS, AND T. W. MALONE (2014): "Reading the Mind in the Eyes or Reading Between the Lines? Theory of Mind Predicts Collective Intelligence Equally Well Online and Face-to-Face," *PloS One*, 9 (12), e115212. [2638,2639,2652]

ERNST, M. D. (2004): "Permutation Methods: A Basis for Exact Inference," *Statistical Science*, 19 (4), 676–685. [2646]

FALK, A., AND J. J. HECKMAN (2009): "Lab Experiments Are a Major Source of Knowledge in the Social Sciences," *Science*, 326 (5952), 535–538. [2639]

FALK, A., AND A. ICHINO (2006): "Clean Evidence on Peer Effects," *Journal of Labor Economics*, 24 (1), 39–57. [2637]

GOLDBERG, L. R. (1992): "The Development of Markers for the Big-Five Factor Structure," *Psychological Assessment*, 4 (1), 26. [2640,2641]

HAMILTON, B. H., J. A. NICKERSON, AND H. OWAN (2003): "Team Incentives and Worker Heterogeneity: An Empirical Analysis of the Impact of Teams on Productivity and Participation," *Journal of Political Economy*, 111 (3), 465–497. [2637]

HECKMAN, J. J., AND T. KAUTZ (2012): "Hard Evidence on Soft Skills," *Labour Economics*, 19 (4), 451–464. [2639,2641,2655]

HECKMAN, J. J., J. STIXRUD, AND S. URZUA (2006): "The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behavior," *Journal of Labor Economics*, 24 (3), 411–482. [2639, 2655]

HERBST, D., AND A. MAS (2015): "Peer Effects on Worker Output in the Laboratory Generalize to the Field," *Science*, 350 (6260), 545–549. [2637,2639]

IPIP (2018): Retrieved from https://ipip.ori.org/New_IPIP-50-item-scale.htm. [2641]

ISPHORDING, I. E., AND U. ZÖLITZ (2019): "The Value of a Peer," Working paper. [2637]

JARAVEL, X., N. PETKOVA, AND A. BELL (2018): "Team-Specific Capital and Innovation," *American Economic Review*, 108 (4–5), 1034–1073. [2640]

KOSSE, F., T. DECKERS, P. PINGER, H. SCHILDBERG-HÖRISCH, AND A. FALK (2020): "The Formation of Prosociality: Causal Evidence on the Role of Social Environment," *Journal of Political Economy*, 128 (2), 434–467. [2639]

KUHN, P., AND C. WEINBERGER (2005): "Leadership Skills and Wages," *Journal of Labor Economics*, 23 (3), 395–436. [2639,2655]

LARSON, J. R. (2010): *In Search of Synergy in Small Group Performance*. Psychology Press. [2648]

LAZEAR, E. P., AND K. L. SHAW (2007): "Personnel Economics: The Economist's View of Human Resources," *Journal of Economic Perspectives*, 21 (4), 91–114. [2637]

LINDBECK, A., AND D. J. SNOWER (2000): "Multitask Learning and the Reorganization of Work: From Tayloristic to Holistic Organization," *Journal of Labor Economics*, 18 (3), 353–376. [2637]

LINDQVIST, E., AND R. VESTMAN (2011): "The Labor Market Returns to Cognitive and Noncognitive Ability: Evidence From the Swedish Enlistment," *American Economic Journal: Applied Economics*, 3 (1), 101–128. [2639,2655]

MAS, A., AND E. MORETTI (2009): "Peers at Work," *American Economic Review*, 99 (1), 112–145. [2637]

MORGESON, F. P., M. H. REIDER, AND M. A. CAMPION (2005): "Selecting Individuals in Team Settings: The Importance of Social Skills, Personality Characteristics, and Teamwork Knowledge," *Personnel Psychology*, 58 (3), 583–611. [2638]

NATIONAL ASSOCIATION OF COLLEGES AND EMPLOYERS (NACE) (2019): "Job Outlook 2020." [2637]

NEFFKE, F. M. (2019): "The Value of Complementary Co-Workers," *Science Advances*, 5 (12), eaax3370. [2640]

O*NET (2020): "Work Context: Work With Work Group or Team," O*NET OnLine. Retrieved from https://www.onetonline.org/find/descriptor/result/4.C.1.b.1.e?s=1&a=1. [2637]

OETTL, A. (2012): "Reconceptualizing Stars: Scientist Helpfulness and Peer Performance," *Management Science*, 58 (6), 1122–1140. [2640]

PINKHAM, A. E., D. L. PENN, M. F. GREEN, B. BUCK, K. HEALEY, AND P. D. HARVEY (2014): "The Social Cognition Psychometric Evaluation Study: Results of the Expert Survey and RAND Panel," *Schizophrenia Bulletin*, 40 (4), 813–823. [2639,2641]

VENSON, D. J., AND S. H. MOOLGAVKAR (1988): "A Method for Computing Profile-Likelihood-Based Confidence Intervals," *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 37 (1), 87–94. [2646]

WEIDMANN, B., D. J. DEMING (2021): "Supplement to 'Team Players: How Social Skills Improve Team Performance'," *Econometrica Supplemental Material*, 89, https://doi.org/10.3982/ECTA18461. [2642]

WOOLLEY, A. W., C. F. CHABRIS, A. PENTLAND, N. HASHMI, AND T. W. MALONE (2010): "Evidence for a Collective Intelligence Factor in the Performance of Human Groups," *Science*, 330 (6004), 686–688. [2638, 2639,2650,2652]

WUCHTY, S., B. F. JONES, AND B. UZZI (2007): "The Increasing Dominance of Teams in Production of Knowledge," *Science*, 316 (5827), 1036–1039. [2640]

Co-editor Guido Imbens handled this manuscript.

Manuscript received 11 May, 2020; final version accepted 10 April, 2021; available online 2 June, 2021.