The Impact of Displaying Diversity Information on the Formation of Self-assembling Teams

Diego Gómez-Zará Northwestern University dgomezara@u.northwestern.edu

Mengzi Guo Northwestern University mengziguo2014@u.northwestern.edu

Leslie A. DeChurch Northwestern University dechurch@northwestern.edu

Noshir Contractor Northwestern University nosh@northwestern.edu

ABSTRACT
Despite the benefits of team diversity, individuals often choose to work with similar others. Online team formation systems have the potential to help people assemble diverse teams. Systems can connect people to collaborators outside their networks, and features can quantify and raise the salience of diversity to users as they search for prospective teammates. But if we build a feature indicating diversity into the tool, how will people react to it? Two experiments manipulating the presence or absence of a "diversity score" feature within a teammate recommender demonstrate that, when present, individuals avoid collaborators who would increase team diversity in favor of those who lower team diversity. These results have important practical implications. Though the increased access to diverse teammates provided by recommender systems may benefit diversity, designers are cautioned against creating features that raise the salience of diversity as this information may undermine diversity.

Author Keywords
diversity; teams; team formation; social recommenders; mixed-effect logistic regressions

CCS Concepts
•**Human-centered computing** → *Empirical studies in collaborative and social computing;* Empirical studies in HCI;

INTRODUCTION
In early December 2018, one of the most prestigious law firms in the U.S. proudly announced its new partner class on LinkedIn. The post ignited outrage on social media as people reacted to the image of the 12 smiling attorneys who made partner—all of them white, one woman [79]. Though societal norms advocate diversity and inclusion, hardwired tendencies create an attraction to people who are similar and familiar [63, 64, 2]. This poses a challenge to organizations seeking to leverage the benefits of diversity in teams [34]. Despite the expanded training and diversity programs in organizations, workers' attitudes and behaviors towards diversity are hard to change [24, 17, 69]. Research across disciplines underscores how homophily—the "love of same"—can have unintended consequences for group compositions, skills, and outcomes [43, 72, 90].

Can technologies help to increase the formation of diverse teams? HCI researchers have developed visualizations and interfaces to promote diversity in information-related problems (e.g., selective exposure, filter-bubbles). These studies show diverse political opinions to users [67], highlighting different individuals [33], mitigating intercultural conflict through automated feedback [38], and using different interfaces to provide more diverse recommendations [85]. Taken together, these applications illustrate the potential for technology to serve diversity goals by exposing people to ideas and people they may not otherwise come in contact with. Can this visibility approach hold the key to forming more diverse teams in organizational settings by circumventing innate homophily tendencies? Previous research focuses more on user satisfaction than decision behaviors. Given that self-assembling teams are proliferating in a myriad of online work environments (e.g., startups, open-source projects, hackathons, social movements, online communities, crowdsourcing projects), it is important to explore ways in which sociotechnical systems can help individuals choose and work with others who may increase their group's diversity [35].

Toward that aim, this paper explores the effect of highlighting the diversity of potential collaborators on the decisions people make about whom to work with. We report two studies examining this question. Study 1 is a pre-post intervention study in an onsite class, with 46 college students in the US. Participants self-assembled into teams to work on a series of projects. They used an online team formation system to search for and invite potential collaborators to join teams. A feature was built into the system that could display a "diversity score" next to each potential teammate (Figure 1). This score presented the person running the search query to what extent adding the prospective teammate would change the diversity of the person (and, if present, their committed teammates). This feature was turned off for their first project ("pre" control condition), and then on, for their second project ("post" treatment condition). We examined the effects of displaying this information on team invitation behavior. We discovered that participants were less likely to choose diverse collaborators when the diversity score feature was "on" than when it was "off." In Study 2, we replicated this study using random assignment to the control and treatment group. Study 2 was a randomized experiment conducted with 70 faculty members from a university in Argentina who enrolled in an online class. They used the system

CHI '20, April 25–30, 2020, Honolulu, HI, USA.

ACM ISBN 978-1-4503-6708-0/20/04 ...$15.00.
DOI: https://dx.doi.org/10.1145/3313831.3376654

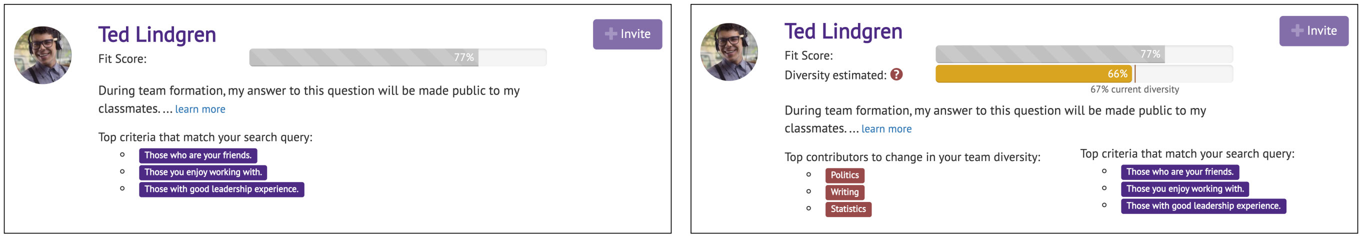

Figure 1. Experimental manipulation. On the left, the control condition where no diversity is information displayed. On the right, the treatment condition where diversity information is displayed.

to form teams to work on technology applications. Faculty members were randomly assigned to conditions, with approximately half using a version of the team formation system with the diversity score feature turned on, and the other half using a version with the diversity score feature turned off. Study 2 replicates the main finding of Study 1: Participants in the treatment condition were less likely than those in the control condition to invite collaborators who would raise the diversity of their team. These findings suggest that when systems provide information about individuals' differences, they are less likely to work with those who are different from them.

This paper makes two contributions. First, it extends insights about diversity in organizations. Displaying diversity information and increasing their perception of dissimilarity negatively affects teammate selection in terms of diversity. Second, it provides new insights on how system interfaces signaling users' perceived diversity during team formation processes influence their choices to work with others. We conclude this paper with theoretical and design implications for future team formation systems to promote diversity considering users' variation and willingness to work with others who are different from them.

LITERATURE REVIEW

We situate our work in the context of prior studies of diversity across the HCI and organizational behavior literature.

Diversity

Diversity is an important topic in several disciplines, including management, biology, psychology, sociology, and information sciences. Most definitions agree that diversity refers to salient differences among individuals. For example, van Knippenberg and Schippers [90] defined diversity as the "differences between individuals on any attribute that may lead to the perception that another person is different from self." Individuals can become aware of their similarities and differences as they face or interact with others. Harrison and Klein [36] emphasized that diversity was a group attribute rather than an individual attribute. They used diversity "to describe the distribution of differences among the members of a unit with respect to a common attribute." Harrison et at. [37] defined diversity at two levels: the *surface-level*, which includes characteristics that are overt, immediately observable, and reflected in physical features; and the *deep-level*, which includes non-observable differences among members, such as attitudes, beliefs, and values. While age, gender, race, and ethnicity are members' surface characteristics, experience, skills, religion, tenure, and status are at the deep-level.

Three main behavioral theories explain individuals' awareness of their differences when they are circumscribed in groups (See [94] for a review). First, *social-categorization theory* posits that individuals categorize themselves and others into groups, distinguishing between similar in-group members and dissimilar out-group members [83, 86]. Second, *similarity/attraction theory* posits that people like and are attracted to others who are similar, rather than dissimilar [15]. Lastly, the *information decision-making theory* proposes that diversity positively influences group processes and group functioning through the increase of information, skills, ability, and knowledge that diversity may bring [94, 87]. Overall, literature provides several conceptualizations of diversity, in which their support has been mixed [43].

Diversity has drawn the attention of several scholars since it may affect group effectiveness, processes, and outcomes [62, 43]. Most results are mixed, and even contradictory. On the one hand, heterogeneous groups can have an outstanding performance since they combine multiple points of view, backgrounds, and knowledge to solve problems [88]. Burt [13, 14] found groups with more diverse social networks had access to more resources than those groups with high redundant social connections. Wang et al. [93] found that teams with higher cultural diversity produced more creative results than those culturally homogeneous. Studying GitHub teams, Vasilescu et al. [91] found that gender diversity and tenure diversity were positive and significant predictors of productivity. On the other hand, literature has shown that diversity faultlines elicit subgroup categorization among members ("us-them" distinction), which disrupts group processes resulting in communication problems [53], lack of cohesion [37], coordination issues, and lack of trust [30]. In response to the mixed results, scholars have been exploring variables that moderate and/or mediate the effect of diversity on group processes. For example, Ye and Robert [99] found that high levels of collectivism facilitated creativity when participants perceived higher diversity in their teams, Ren and Yan [70] found that communication processes moderated positively the effect of diversity on performance, and Azary et al. [4] found that task conflict moderates the effect of cognitive diversity on the quality of teams' outcomes.

FTF vs CMC environments

In contrast to face-to-face (FTF) environments, HCI and CSCW researchers have explored whether the effects of diversity are similar or not in computer-mediated communication (CMC) environments. Research has arrived at a consensus that demographic aspects can be less evident in online environments than offline environments [8, 31]. Carte and Chidambaram [16] proposed that online technologies can reduce the negative effects of diversity by playing down members' surface-level characteristics in the early life of a diverse team whereby participants spend more time communicating and forming ties based on their deep-level characteristics. However, studies have shown that users can infer surface-level attributes through communication processes [23] and self-presentation on online platforms [12, 76]. CHI proceedings have highlighted the overwhelming use of gender and socio-economic class to differentiate systems' users [40, 80].

Studies in online environments including crowd-sourcing platforms—such as Wikipedia [70, 65, 21, 49] and GitHub [5, 92, 44]—have tested how diversity affects group performance. For example, Lionel and Romero [74, 72] found that diversity was beneficial for large crowds with low member retention rates but not when crowds were small. Chen et al. [21] found that increasing tenure diversity led to better group outcomes but that at very high levels of tenure-diversity led to negative group outcomes.

Displaying diversity information

Studies that explore the consequences of team diversity on processes and outcomes [43, 10, 9] generally take diversity—or the lack thereof—as a *fait accompli*. An important open question is: when members have the option to form their teams, would they be open to choose diverse teammates? Previous studies on social recommenders suggest diversity is unlikely to emerge as people's selections are based on similarity and common social connections [20, 51, 7]. But these studies explore friendship relationships—rather than work collaborations—and they do not explore whether highlighting information about others' diversity increases the likelihood of forming new relationships with diverse users.

Displaying information to users has been tested by HCI researchers to stimulate creativity, healthy behaviors, and task learning [55, 68, 19], but few studies have explored how displaying information about others' diversity affects users' behaviors. Most related studies conclude that users who tend to connect with like-minded others may feel threatened by unfamiliar information, and avoid interacting with others who have different opinions or qualities [56, 33, 28, 95]. Munson and Resnick [67] found that showing diverse political opinions did not make users more accepting of opinions that differed from their own. Graells-Garrido et al. [33] found that users would eschew algorithmic recommendations of a system and form ties with others of similar political views. An et al. [3] found that users did not value seeing information they were exposed to that was different from their beliefs. They found that users' prior beliefs, emotional state, social context were likely to explain when they were adverse or favorable to unfamiliar information. Notwithstanding the availability of more advanced visualizations options to find more diverse options, Tsai and Brusilovsky [85] found that users tended to use the visualization that enabled them to find similar people. Overall, these studies focus on users' satisfaction with the recommendations provided by systems, but not on whether users' behaviors in choosing options that are more diverse to them.

Our work fills this gap by exploring how showing diversity information can affect users' behavior at the moment of choosing teammates. Specifically, we hypothesize that displaying diversity information will induce users to choose others who are more similar to them, rather than different. Research has shown that people are likely to choose as teammates those who are similar to them [41]; and users are likely to choose teammates based on how they self-present on these platforms or how they perceive their differences based on pictures, names, common social connections, or descriptions [32, 12, 11, 98]. Additionally, people may have *unconscious bias* against working with people who are different from them [75]. Choosing similar others allows people to reduce the uncertainty associated with work with people with different characteristics and attitudes [34]. By displaying diversity information, users' differences will become more salient, and those who are averse to diversity would tend to choose people who are more similar to them. Therefore, we posit that:

H1: Displaying diversity information negatively affects selecting more diverse teammates.

To test this hypothesis, we conducted two studies examining how displaying diversity information affects users' teammate selection. Study 1 was conducted in an onsite class with 46 students who used a team formation system to assemble into teams for two projects. We showed them diversity information only when they were assembling teams for the second project. Study 2 was a field experiment where 70 faculty members were randomly assigned into two sections (control and treatment) of an online class requiring them to work on team projects. While the diversity information was not displayed to participants using the team formation system in the control group, the diversity information was displayed to participants in the treatment condition. By conducting these two studies—one carried out in an FTF setting and another in a CMC setting—we test our hypothesis and demonstrate to what extent the effects of displaying diversity information on team formation processes are reliable and consistent.

STUDY 1: PRE-POST INTERVENTION STUDY

This first study examines how presenting diversity information to users affects their judgment and selection of potential teammates. Participants carried out two team projects in sequence. For each project, participants decided whom to ask to join their teams. We call this request to join a team an "invitation." To study participants' invitations, we provided them a team formation system called *MyDreamTeam* [22], which enables users to self-assemble into teams. Participants were able to ask others to join their teams, and they could accept or decline the invitation. To test our hypothesis, we compare whether displaying diversity information affected participants' teammate selection using a pre-post design: for the first project, the system calculated a diversity score but did not display it to the participants; and for the second project, the system calculated and displayed the diversity information.

Participants

We conducted this study in an undergraduate onsite class at a university in the US. The course required students to assemble into teams of approximately seven members. Forty-six students attended this class, 22 were female and 24 were male. The mean age was 21.11 (SD=1.23), and 9% of the students were international. Students participated voluntarily and provided informed consent. No incentives were provided. Students had the option of assembling teams without using the system—in which case the instructor would assign them randomly into an existing team—and it was explained that doing so would not affect the grade.

Procedure

In the first team project, participants developed an innovation strategy for an entrepreneurial start-up company called "Happy Earth." The team deliverable was a 10-minute presentation of their innovation strategy. In the second project, participants played a social dilemma game [50], and then formed teams to discuss their results and reflect upon their experiences. Teams presented a 5-minute presentation summarizing their experiences and analytical reflections.

For each project, the instructor created the team assembly sessions on the system and added participants' email addresses. Each participant received an email with the access information. Participants then created profiles on the system, which included details about their background information. The system allowed participants to display public information in their profiles, such as their background, hobbies, and motivations. The system displayed participants' names, which identified them to other participants throughout the entire team formation stage. Participants also answered a survey to assess their personality [25], social networks, project skills, leadership experience [66], social skills [29], creativity [84], and psychological collectivism [45]. These answers were confidential but used by the team formation system to generate recommendations of potential teammates.

After completing their profiles and surveys, participants were prompted to fill out a search query to find potential collaborators using search attributes based on the initial survey questions. The query prompted participants to rate these attributes on a 7-point Likert scale, ranging from "Not important at all" (-3), to "Don't care" (0), and "Yes, for sure" (+3). The search attributes were based on the initial survey questions. After the participant established the search criteria, the system generated a rank-sorted list of teammate recommendations that best matched the participant's search query. For each recommendation, the system displayed participants' full name, the percentage of how well they matched with the participant's query (labeled on the system as "fit score"), a link to their public profile, their current teammates, and an "Invite" button (see Figures in supplementary materials). If the participant (sender) decided to invite that potential teammate (recipient), the system would send an invitation message to that person.

Participants who receive an invitation could accept it, decline it, or ignore it. If the recipient accepts the invitation, the sender and the recipient would be in a team. If the recipient or sender were already part of a team, their teams would merge. The system only allowed teams to merge if the final team size was less than or equal to the maximum size (of seven) allowed. Participants were also able to switch or leave their teams. The instructor allowed participants one week to assemble their teams. All participants assembled teams through the system, and no information about their diversity was provided outside of the system.

Measurements

Dependent Variable: Teammate selections

To compare how teams were assembled in both projects, we use participants' invitations on the system as the dependent measure. The system tracked when participants invited a specific person based on a recommendation result (1=selected, 0=not selected). Since participants typically do not scroll all the way down the recommendation results, we ran the analysis by including the top 15, 20, 25, and 30; and the results were qualitatively the same. We report results based on the top 20 since most of those invited to join teams appeared as a top 20 search result (84.5%).

Independent variables

Diversity score: The system calculates how the diversity of the participant's current team will be affected by adding the potential teammate. Participants can increase (or decrease) diversity by adding more diverse (or similar) people to their teams. Based on Harrison and Klein's diversity typology [36], we operationalize the diversity contributed by the potential teammate using the Blau index for categorical attributes. This is defined as $1-\sum p_i^2$, where p is the proportion of team members who fall into a particular category i. A low score means that all members fall into the same category, and a high score means that members fall into different categories. For numerical variables, we operationalize the diversity contributed by the potential teammate using the coefficient of variation. It is defined as the ratio of the standard deviation to the mean of a variable x, and its formula is $\sqrt{\sum(x_i - x_{mean})^2/n}/x_{mean}$. A low score means all members have similar levels of the attribute, and a high score means that members have different levels of the attribute.

In the study's initial survey, we assessed the following participants' characteristics—which are the criteria most frequently considered by participants when assembling teams—to calculate their teams' diversity:

- *Age diversity*: This is computed using the coefficient of variation of the self-reported ages of team members.
- *Gender diversity*: In the initial survey, participants self-reported their gender identity as "Male," "Female," or "Other.' This is computed using the Blau index.
- *Cultural diversity*: We used nationality as a proxy for cultural diversity [82]. In the initial survey, we asked participants "Which is your country of nationality citizenship?" and they answered using a list of countries provided by the system. This metric is calculated using the Blau index.
- *Project skills diversity*: In the initial survey, participants self-reported their expertise on six project skills relevant to the course using a 5-point Likert scale ranging from "Not at all skilled" to "Extremely skilled." This is calculated using the coefficient of variation for each project skills and then averaged across all six.

While the first two attributes are surface-level attributes, the last two attributes are deep-level attributes. Each diversity metric ranged from 0 to 1 since we normalized them based on their theoretical maximum.

To test our hypothesis—whether displaying diversity information negatively affects selecting more diverse teammates—we averaged these diversity metrics to compute a "diversity score" as a single independent variable for each potential teammate. This allows us to measure to what extent the potential change in diversity—measured on a continuous scale—affected participants' team member selections. The system calculated these metrics for every recommendation provided to users, whether or not the user saw this score. The supplementary material includes the equations for these metrics.

Finally, we conducted a pilot study with ten students from our university to evaluate interfaces used by the system to display participants' diversity information. Based on their feedback, the system displays this score as a percentage: A score of 0% means that all the team members are similar, and 100% means that all team members have different attributes. The system uses bars to show the participant how his/her current team's diversity would be affected by incorporating the potential teammate. Additionally, the system displays the potential teammate's most salient contributors to change participants' team diversity (Figure 1).

Gender homophily: Although gender was included in the diversity score, we also added gender homophily as an independent variable (i.e., sending invitations to those of the same gender) to check whether this effect was above and beyond displaying diversity information to participants. This dummy variable indicated "1" if both users had the same gender.

Control variables

To test our hypothesis on the impact of displaying diversity information, we controlled for factors we expect would influence participants' decisions to invite others: their prior relations with others and the rank of others in a list of recommendations in response to a search query.

Prior relations: In the initial survey, we gave participants a full list of the people in their class and asked them to identify who (1) they had previously collaborated with and (2) they have enjoyed socializing with to measure two social networks (collaboration and friendship). We consolidated each participant's responses by assigning a relationship between two participants if at least one involved participant reported a connection to the other [52].

Recommendation rank: We used the rank of the recommendation on the results page. The system computed this ordinal variable from the "fit score," which is the percentage of how well they matched with the participant's query. For example, "1" means that it was the first recommendation provided, "2" means it was in the second position, and a larger number indicates that it was listed toward the end.

Manipulation check

To test whether participants paid attention to the diversity score, we included a manipulation check in a final online survey. This voluntary survey was conducted after all team project work was completed and turned in (52.2% completion rate). Participants were asked: "Did you notice the diversity score displayed for each potential teammate?" Overall, 71.4% of the participants reported seeing the diversity score.

Methodology

To test our hypothesis, we use mixed-effects logistic regression [1] to predict teammate selection. This method models the log-odds of a binary outcome as a linear combination of the predictor variables using both fixed and random effects. We model senders as random effects—as each participant could search multiple times—and thus, invitations are nested within senders. In other words, the senders are the Level-1 unit of this study. To test whether displaying diversity information affected teammate selection (H1), we add an interaction effect between the treatment condition (i.e., whether the system displayed diversity information or not) and the diversity score. We conducted the analysis hierarchically. First, we begin with a model that only includes control variables: the recommendation rank and prior relations (Model 1). We then include the treatment condition variable, the diversity score, and their interaction terms (Model 2a). To test the effects of gender homophily, from Model 1 we include the treatment condition variable, the gender homophily effect, and their interaction terms (Model 2b). Finally, we test a model with the control variables, the independent variables, and the interaction terms (Model 3). Model statistics were compared at each step to determine if the additional parameters were significant (using likelihood ratio tests) and improved the overall fit (comparing AIC/BIC). To check goodness-of-fit, we test whether the model suffers from overdispersion by comparing Pearson's residual to the residual degrees of freedom using Pearson's chi-squared test. We also check for multicollinearity by calculating the variance inflation factors (VIF) and checked that the values were less than 10. We use the R package *lme4* to test the proposed models. More details can be found in the supplementary materials.

Results

We provide descriptive statistics for the variables used in the analyses. From the initial survey, the average of the participants' project skills was 3.07 (SD=0.63), the median number of previous collaborations reported by participants was 2 (M=3.04, SD=3.14), and the median number of friends was 3 (M=3.91, SD=3.72). In the first project, 33 of 46 participants searched for potential teammates, the number of searches per participant was 1.94 (SD=1.68), the system generated 1,261 recommendations, 63 were selected by participants, and the average number of recommendations selected per user was 1.91 (SD=2.42). From the second project, 20 participants searched for potential teammates, the number of searches per participant was 1.30 (SD=0.47), the system generated 520 recommendations, 26 were selected by participants, and the average number of recommendations selected per user was 1.30 (SD=1.13). Fewer participants searched for teammates in the second project. However, the proportion of invitations per participants' searches in the second project was similar to that in the first project. Correlation analysis found that the

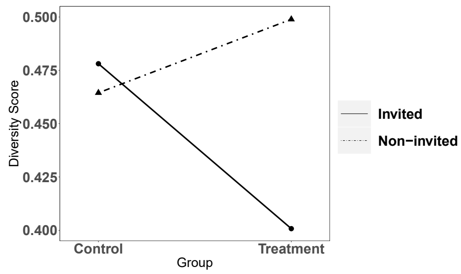

Figure 2. Showing the diversity score reduces the diversity of invited teammates

friendship between two users was moderately correlated with their previous collaboration ($r = 0.513, p < .001$). The recommendation rank of the potential teammates displayed on the results page was weakly correlated with previous collaboration ($r = -0.230, p < .001$) and friendship ($r = -0.202, p < .001$). Furthermore, the diversity score was weakly correlated to gender homophily ($r = -0.198, p < .001$). To determine the amount of membership change between the first and second projects, we classify each dyad as being on the same team or not during the first and second projects: 50.7% of the dyads remained the same between the two projects.

Invitations sent by participants in the treatment group had lower diversity scores

Since the dependent variable is categorical, we first check the average diversity scores of the recommendations generated by the system and compare those that were selected to those that were not. We find that users who were shown diversity information (treatment-group) invited others who were more similar to them. Figure 2 displays the average diversity scores for recommended teammates who were selected and those who were not by treatment condition. For the control group, the mean diversity score is 0.478 (SD=0.173) from the selected recommendations and 0.464 (SD=0.149) from the not-selected recommendations. For the treatment group, the mean diversity score is 0.401 (SD=0.166) from the selected recommendations and 0.499 (SD=0.160) from not-selected recommendations. The difference in the average diversity score between selected and not-selected recommendations is not statistically significant in the control group (two-sided unpaired t-test, $t = -0.62, p > .05$), but statistically significant and less diverse in the treatment group (two-sided unpaired t-test, $t = 2.94, p < .01$).

Displaying diversity information affects teammate selection

Table 1 shows the odd-ratios and standard errors of the mixed-effect logistic regression models, which regress users' decisions to invite (or not) a potential teammate. Model 1—with the control variables—finds that the recommendations' rank and prior relations significantly affected teammate selection. We find that participants were 61.2% times more likely to be invited if they were listed at the beginning rather than at the end ($OR = 0.388, p < .001$). Participants were 5.456 times more likely to send an invitation to a friend ($p < .001$) and 2.03 times more likely to send an invitation to a previous collaborator ($p > .05$).

Adding treatment, diversity score, and their interaction (i.e., showing diversity information) as independent variables in Model 2a does not significantly improve over the first model ($\chi^2(3) = 5.59, p > .10$). The interaction term shows that participants who saw the diversity score were 44.9% less likely to invite someone who would increase the team's diversity than someone who would not increase it ($p < .05$). The treatment condition itself does not affect the number of sent invitations ($OR = 0.991, p > .05$), nor does the diversity score ($OR = 1.127, p > .05$). This suggests that a simple manipulation as showing a user how a prospective teammate would affect their team's diversity has significant (interaction) effect on decisions about whom to collaborate with.

Since gender was included in the diversity score, Model 2b explores the addition of gender homophily as an alternative explanation. We find that this variable significantly affects participants' teammate selection ($\chi^2(3) = 16.04, p < .01$). Participants were 1.86 times more likely to send an invitation to someone of the same gender ($p < .01$). In the treatment condition, the effect of gender homophily was larger, but this difference was not statistically significant ($OR = 3.408, p > .05$).

Finally, Model 3 includes all independent and control variables. Comparing Model 3 to Model 1, we find that including treatment, gender homophily, diversity score, and their interactions significantly affects teammate selection ($\chi^2(5) = 20.215, p < .01$). Specifically, those who appeared at the top of the ranked recommendations were 62.3% more likely to receive an invitation ($OR = 0.377, p < .001$), friends were 3.831 times more likely to receive an invitation ($p < .01$), individuals who would increase team diversity were 41.6% less likely to be invited when the diversity score was displayed ($OR = 0.584, p < .05$), and same-gender individuals were 2.02 times more likely to receive an invitation than different-gender individuals ($p < .05$). The significant interaction between the diversity score and treatment condition supports H1.

The gender homophily effect was larger in the treatment condition, but this difference was not significant. It is notable that interaction term between treatment and diversity score is still significant while accounting for gender homophily. This suggests the interaction is driven more by the remaining components of the diversity score—age, cultural diversity, and project skill diversity—than by gender diversity.

The AIC values indicate that Model 3 provides the best fit for the data. None of the models were overdispersed since all three p-values from the Pearson's chi-squared tests were larger than 0.05. Multicollinearity was not an issue: for all the models presented, no factor had a VIF value greater than 3.8.

Although this study shows that displaying diversity information influenced teammate selection, the pre-post design is limited by confounding treatment and maturation effects; the use of a student sample is another limitation. We conducted a second randomized experiment in a field setting to address these two limitations.

STUDY 2: RANDOMIZED FIELD EXPERIMENT

We conducted a second study to test whether this effect is consistent in a randomized experimental context. Instead

	Model 1		Model 2a		Model 2b		Model 3	
Fixed effects	Odds Ratio	(SE)	Odds Ratio	(SE)	Odds Ratio	(SE)	Odds Ratio	(SE)
Intercept	0.017	(0.316)***	0.017	(0.339)***	0.0119	(0.392)***	0.0113	(0.397)***
Recommendation rank	0.388	(0.155)***	0.391	(0.156)***	0.373	(0.158)***	0.377	(0.157)***
Previous collaboration	2.031	-0.443	2.143	-0.452	1.888	-0.472	2.08	-0.48
Friend	5.456	(0.421)***	4.993	(0.429)***	4.081	(0.452)**	3.831	(0.459)**
Treatment	-	-	0.991	-0.328	0.472	-0.596	0.509	-0.598
Diversity score	-	-	1.127	-0.156	-	-	1.204	-0.156
Diversity score x Treatment	-	-	0.551	(0.262)*	-	-	0.584	(0.266)*
Gender homophily	-	-	-	-	1.855	(0.308)*	2.02	(0.313)*
Gender homophily x Treatment	-	-	-	-	3.408	-0.658	2.733	-0.671
Random Effects	Variance	(SD)	Variance	(SD)	Variance	(SD)	Variance	(SD)
Variance	1.44	-1.2	1.48	-1.215	1.516	-1.231	1.522	-1.234
Log Likelihood	-278.45		-275.65		-270.43		-268.34	
AIC	566.9		567.3		556.86		556.69	

Significance code: * $p < .05$, ** $p < .01$, *** $p < .001$. Number of senders: 36. Number of observations: 1,781.

Table 1. Study 1. Mixed-effects logistic regression predicting participants' invitations

of testing a single group in a pre-post design, we randomly assigned participants to two groups (control and treatment) to assemble teams either with or without seeing a diversity score associated with teammate recommendations.

Participants

We recruited 70 faculty members at a public university in Argentina. Participants enrolled voluntarily in a five-week online course to explore the use of technologies in education. More females than males participated (78.5%), the average age was 43.26 (SD=11.28), and no international participants took this course. The online course, provided by a university center for teaching, was open and free for all the university's faculty members. One instructor was responsible for teaching this course, along with two teaching assistants. The research group only provided access to the team formation system and was not involved in the course's learning activities and evaluations. Using the team formation system was voluntary, and participants provided informed consent. If participants did not want to use the system, they could manually assemble their teams. The participants were assured that use of this team formation system was not going to affect their experience in the course and their grades. We randomly assigned each participant to one of the two diversity-information conditions, resulting in 36 participants in the control group and 34 in the treatment group (two dropped the course). In total, participants assembled 13 teams (6 in the control group and 7 in the treatment group) ranging from 2 to 5 members per team.

Procedure

After we assigned participants to each treatment condition, the instructor created two separate sections. The first included the control group participants and the second the treatment group. In this online class, participants reflected on how to promote cooperation and collaboration strategies in their courses. Before the course started, the instructor invited each course section to the team formation system. Participants received an email with login instructions, completed an initial survey to assess their traits and social networks. The team formation procedure and system were the same as in Study 1. All participants assembled teams through the system, and no information about their diversity was provided outside of the system.

As part of the course, participants watched videos, discussed them in their teams, and elaborated a report synthesizing their main takeaways. Since all course activities were online, participants interacted with each other using the course website. They were also allowed to communicate with their teammates using email and instant-messaging technologies, such as WhatsApp, Facebook, or Google Hangouts. We did not control whether participants worked together in face-to-face conditions after they assembled their teams. The final team deliverable was a 2-page document that included the team's analysis of the videos and reflections.

Measurements

Different from Study 1, we included additional deep-level attributes to the diversity score relevant to the academic context of this course. Specifically, we added:

- *Faculty diversity*: At this university, faculty is a division comprising one subject area or a group of related subject areas, which is similar to colleges or schools in the US. Participants reported their primary affiliation in the initial survey. Faculty diversity was calculated using the Blau index to capture differences in the proportion of team members who were from the same or different faculties.
- *Work status diversity*: It was calculated using the Blau index, as the proportion of team members' time dedication (i.e., full- or part-time).
- *Academic position diversity*: It was calculated using the Blau index based on tenure at the university.

Manipulation check

To check if participants paid attention to the diversity score displayed by the system, we included a manipulation check in a voluntary online survey. Participants completed this survey after they submitted their team reports. We asked participants in the treatment condition who searched for teammates about whether or not they saw the diversity score. From the treatment group, 26 participants completed the survey, 22 of whom had used the system to find potential teammates. Of these, 83.3% of the participants reported seeing the diversity score.

Analysis and Results

We provide descriptive statistics for the main variables used in this study. From the initial surveys, we do not find significant differences in participants' project skills between the control group (M=3.26, SD=0.65) and the treatment group (M=3.29,

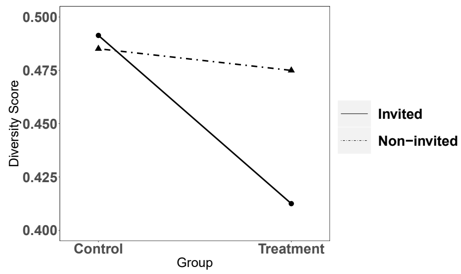

Figure 3. Showing the diversity score reduces the diversity of invited teammates

SD=0.45). In the control condition, we find that the median of friends reported was 2 (M=4.4, SD=11.30) and the median of collaborators reported was 2 (M=4.55, SD=11.26). In the treatment group, the median of friends reported by participants was 6 (M=12.47, SD=17.06), which was higher than in the treatment group, and previous collaborators were 2 (M=5.52, SD=10.98). From the control group, 23 participants searched for potential teammates, the number of searches per participant was 2.13 (SD=1.46), the system generated 944 recommendations, and 66 were selected by participants, and the average number of invitations sent by each participant was 2.86 (SD=3.02). From the treatment condition, 24 participants searched for potential teammates, the number of searches per participant was 3.00 (SD=2.21), the system generated 1,408 recommendations, 72 were selected by participants, and the average number of invitations sent by each participant was 3.00 (SD=3.36). Among the independent variables, friendship was moderately correlated with previous collaboration ($r = 0.549, p < .001$). The diversity score of each person recommended to the searcher was highly correlated with gender homophily ($r = -0.718, p < .001$), weakly correlated with their friendship ($r = 0.239, p < .001$) and their previous collaboration ($r = 0.182, p < .001$).

Users who saw a diversity score were less likely to invite teammates who would diversify their team

Figure 3 shows the same pattern that was found in Study 1. Users who were shown the diversity score were less likely to invite teammates who would increase the team's diversity. For the control group, the mean diversity score is 0.491 (SD=0.156) from the selected recommendations and 0.485 (SD=0.150) from the not-selected recommendations. For the treatment group, the mean diversity score is 0.413 (SD=0.165) from the selected recommendations and 0.475 (SD=0.159) from not-selected recommendations. Replicating Study 1, the difference between not-selected and selected recommendations in the treatment group was statistically significant (two-sided unpaired t-test, $t = 3.14, p < .001$).

Displaying diversity information affects teammate selection

We replicated the analysis from Study 1, using a mixed-effect logistic regression to determine whether displaying diversity information affects participants' teammate selection (Table 2). The null model controls for the non-independence created by having multiple invitations per participant. Model 1 is a baseline, controlling for three variables known to affect teammate preferences. Consistent with the first study, participants were 44.7% times more likely to be invited when they were displayed at the beginning of the list rather than at the end ($OR = 0.553, p < .001$), 2.649 times more likely if they worked together in the past ($p < .05$), and 2.647 times more likely if they were friends ($p < .01$).

Model 2a includes the diversity score treatment condition, diversity score, and their interaction as independent variables. We find these variables significantly affect participants' teammate selection ($\chi^2(5) = 17.63, p < .001$). The treatment condition and displaying diversity information are statistically significant, where participants were 48.4% times less likely to invite someone who would make the team more diverse than someone who would make it less diverse ($p < .01$). Interestingly, this model shows that participants in the treatment condition were less likely to send an invitation than those in the control condition ($p < 0.05$).

Model 2b explores the addition of gender homophily as an alternative explanatory variable. In contrast to Study 1, we did not find a significant gender homophily effect in Study 2 ($\chi^2(3) = 7.30, p > .05$). Neither gender homophily nor its interaction with the treatment variable was significant. However, the gender homophily effect is considerably larger in the treatment than in the control condition, and it is marginally significant ($p < .10$).

Model 3 includes all independent and control variables. Adding treatment, diversity score, and gender homophily variables significantly affect participants' teammate selection ($\chi^2(5) = 17.94, p < .001$). Displaying diversity information is significant, even accounting for the diversity score itself and the other controls known to affect teammate selection like recommendation rank, prior collaboration, and friendship. When controlling these other effects, participants in the treatment condition were 48.7% less likely to invite teammates who would diversify their team. Therefore, H1 is supported.

None of the models were over-dispersed since all three *p*-values from the Pearson's chi-squared tests were larger than 0.05. Multicollinearity was not an issue: for all the models presented, no factor had a VIF value greater than 5.5.

DISCUSSION

In this paper, we examine whether and how displaying diversity information to individuals as they form teams affects their willingness to choose to work with others who are different from them. These two studies confirm that individuals who were shown diversity information were less likely to choose to work with others who differ from them (H1). Whom do individuals choose as teammates? They prefer those who appear higher on a recommendation list, those who are friends, those who are prior collaborators, and those who appear similar to them. Taken together, these findings show the negative consequences of highlighting the differences among members on forming diverse teams. It also underscores how interfaces may generate bias on people's teammate selection. These findings lead to four conclusions.

First, this paper provides empirical evidence of how people's proclivity for similarity can be exacerbated by showing them their differences from others. Similar to previous HCI studies on selective exposure and filter-bubbles, this paper highlights

	Model 1		Model 2a		Model 2b		Model 3	
Fixed effects	Odds Ratio	(SE)	Odds Ratio	(SE)	Odds Ratio	(SE)	Odds Ratio	(SE)
Intercept	0.030	(0.241)***	0.041	(0.295)***	0.043	(0.351)***	0.045	(0.409)***
Recommendation Rank	0.553	(0.097)***	0.552	(0.102)***	0.546	(0.098)***	0.552	(0.102)***
Previous collaboration	2.649	(0.393)*	2.516	(0.415)*	2.787	(0.395)**	2.466	(0.416)*
Friend	2.647	(0.310)**	4.337	(0.336)***	3.636	(0.330)***	4.33	(0.340)***
Treatment	-	-	0.431	(0.411)*	0.247	(0.560)*	0.453	-0.63
Diversity score	-	-	1.031	-0.157	-	-	0.972	-0.23
Diversity score x Treatment	-	-	0.516	(0.229)**	-	-	0.513	(0.319)*
Gender Homophily	-	-	-	-	0.901	-0.304	0.851	-0.457
Gender Homophily x Treatment	-	-	-	-	2.525	-0.484	0.944	-0.698
Random Effects	Variance	(SD)	Variance	(SD)	Variance	(SD)	Variance	(SD)
Variance	1.475	-1.214	1.165	-1.08	1.296	-1.138	1.164	-1.079
Log Likelihood	-456.74		-447.92		-453.09		-447.77	
AIC	923.48		911.84		922.18		915.53	

Significance code: * $p < .05$, ** $p < .01$, *** $p < .001$. Number of senders: 47. Number of observations: 2,352.

Table 2. Study 2. Mixed-effects logistic regression predicting participants' invitations

the importance of the information displayed in socio-technical systems and shows how profiles, pictures, and other recommendation metrics can create bias among users. Prior literature has been clear on how social networks, stereotypes, and prejudices affect people's choices to work with others. This paper extends this work by demonstrating how system features can also affect those choices. One possible explanation for these results is that displaying diversity information increases users' perceived diversity, which is the group members' belief that others are different [99, 81]. Perceived diversity depends on how group members recognize their differences and benefit from them. It can be positively related to better teamwork by increasing trust and cohesion, and decreasing conflict [27]. However, when members are not able to take advantage of their differences, the opposite effects are more likely to occur [73]. Individuals who are more aware of their dissimilarities are less likely to be involved in task-related processes [42], identify less with their teams, and are more likely to have conflicts [39]. Since users are highly reliant on systems' design and information displayed in online environments, users' awareness of their differences is more likely to increase, and perceived diversity would induce them to select more similar individuals.

An alternative explanation for this result is the inclusion of surface-level attributes in the diversity score, making them more explicit. As [16] suggests, technologies should emphasize deep-level attributes among team members to increase relationships and trust. Future studies should examine whether highlighting only deep-level characteristics prompt users to seek potential teammates that offer different skills, values, and backgrounds.

Second, the findings for gender homophily were mixed. Gender homophily was influential in Study 1 but not in Study 2. A possible reason is the FTF context in the first study, where participants were more aware of their surface-level differences since they were able to see each other in the classroom. In contrast, the CMC context in Study 2 may have played down gender differences. Since they expected to work online, they could have prioritized other characteristics at the moment of assembling teams—such as project skills. However, this conclusion requires more testing since prior results are mixed [26, 58]. Cultural, age, and environmental differences between the two studies could have been other possible explanations. Like previous studies comparing FTF and CMC teams, future work should address whether gender homophily can be mitigated by the use of communication technologies.

Third, this paper shows the effects of grouping several dimensions of diversity into a single score. This aggregation has effects on (i) how systems calculate diversity among users and (ii) how users will be aware of their differences. Diversity can be operationalized in a more holistic way rather than a sum of attributes [80]. Moreover, aggregating diversity dimensions into a single number generated potentially simplistic evaluations and allowed participants to see which potential teammates were more similar to them. Future studies should test how displaying different metrics of diversity—one for each kind—through the use of badges or bars could enable them of their options and decisions. Systems could detect what kinds of diversity types (e.g., cognitive, cultural, gender) benefits users and make suggestions based on those. For example, a multi-armed bandit approach could allow systems to test different diversity types options by providing recommendations to users and learning from their choices [100].

Fourth, this work shows that the effect of displaying diversity information on individuals' invitation behavior has consequences for overall emergent team composition. We analyzed whether teams in the treatment condition were less diverse than those assembled in the control group condition. We compared the teams' diversity metrics in both conditions computing the Blau index for categorical variables and Gini coefficients for numerical variables. In Study 1, we found the control condition's teams were more diverse in gender (8%) and project skills (2%) than treatment condition's teams. Similarly, we found in Study 2 that control condition's teams were slightly more diverse than treatment condition's teams in school representation (10%), time dedication (5%), project skills (2%), academic positions (2%), and gender (1%). Although these differences were not statistically significant—perhaps due to the small number of teams—they exhibit how the effect of displaying diversity information on individuals' choices at the micro-level manifests in overall team diversity at the emergent macro-level. As such, this paper sheds light on how presenting users' identities affect the composition of online communities, where diversity is more likely to work in large groups since members are able to find similar others within the group [72]. Since most online communities are organically assembled, the way in which systems present users and make visible their attributes has consequences on the composition of smaller communities, and ultimately, in their inclusion, cohesion, and performance [72, 70].

Theoretical implications

This paper advances our understanding of how socio-technical systems organize people. Team formation systems, like the one used in this paper, do far more than inform users about their options and augment their decisions. Their interfaces and features can promote certain interactions among individuals who are similar and different and provide several ways to organize them in groups. This work offers a new view regarding how diversity information affects the composition of teams assembled in socio-technical systems. How can systems provide a better balance? Previous crowd systems have strongly relied on computational mechanisms to assemble more effective teams [59, 77, 78, 71, 101], however, the challenge remains in providing users a certain degree of flexibility and agency. Certainly, there is a trade-off. On the one hand, literature has shown how allowing users to assemble their own teams may lead to better group dynamics and outcomes since members are more committed to the task and the team [6, 18, 61]. On the other hand, segregation, homophily, and lack of diversity are more likely to occur in those teams. Although the system described contributions that diverse people could bring to their teams, most participants were not likely to consider those recommendations. This paper prompts discussion which is fundamental not only for crowd or team formation systems, but is also relevant for social network platforms where homophily and agency are present and coexist together. The emergence of diversity is not a mathematical aggregation process, but fundamentally a social process that requires communication, coordination, and commitment of their members [70].

Design Implications

The results of this paper provide guidance on how HCI practitioners should design systems that promote conscious reflections about diversity among users. Previous literature emphasizes how allowing people to choose by themselves provides better results [24, 54] and how branding diversity in organizations may affect some users' perceptions, generating distrust and stress [47]. In this sense, systems can engage users in choosing diverse teammates by *priming* them with potential benefits of diversity (e.g., showing studies' results) and allow them to make a final decision. The use of alternatives languages—such as framing people's diversity as *strengths* or *complementary skills*—may help users to avoid raising their bias against diversity. Future systems could assess participants' openness to diversity and display diverse information only to those who may seek it [67]. Through the data collection of users' personality and social networks, systems could potentially infer their willingness to seek diversity (or avoid it). For example, people who are more connected with diverse communities could be more likely to receive these recommendations [14]. Since the most effective diversity programs are those that engage people in working for diversity and increase their contact with those who are different [24] systems can facilitate interactions among users who are different through (i) implementing social agents (e.g., bots) who interact with users and become "match-makers" [97], (ii) providing icebreaker exercises to allow direct social contact between them and facilitating the reduction of prejudices [60], (iii) highlighting users who can mediate among diverse people and perform as "bridges" between users with extreme characteristics, and (iv) highlighting the attributes that diverse users may have in common. These ideas require experimentation and testing.

Limitations

It is important to acknowledge the limitations of this paper. First, we did not directly measure participants' perceptions of diversity on teams. To address this issue, future studies must assess individuals' perceived diversity—whose measurement is still under discussion in the academic community [81]—and analyze any latent variables between perceived diversity and diversity information. Second, the characteristics of our context and participants place important boundary conditions on the findings. Additional research is needed with a broader population and different organizational contexts [46]. Having more diverse samples may affect the extent to which diversity levels suppress or trigger social-categorization processes, which in turn would affect the degree to which users are sensitive and/or reactive to features that signal diversity information. Third, most research on group diversity is premised on its impact on group processes and outcomes, such as performance and viability, which were not tested in this paper [48, 96, 89]. Fourth, we did not control any other communication channels that participants could have used, which could have influenced their choices. Finally, testing alternative interface designs could have strengthened this paper. Though our "one single-score" design did not differ from previous HCI studies [56, 57], this could have been too simplistic. Future experiments could include the variation of several interfaces to see which ones are more effective [85].

CONCLUSION

There is a growing groundswell of societal interest in diversity. Mounting evidence supports the benefits of diversity in teams. As technologies increasingly make it possible for teams to form and collaborate across boundaries, an open question is: Can technologies support diverse team formation? The two studies presented in this paper underscore the complexity of designing for diversity. Whereas a feature can be built to display hypotheticals, individuals' motives may run counter to the intended goal of such features, in this case, team diversity.

ACKNOWLEDGMENTS

This material is based upon work supported by the National Aeronautics and Space Administration under grant NNX15AM32G, and the National Institutes of Health under award number R01GM112938-01. We thank Sabine Brunswicker, Silvia Andreoli, Julia de Souza Faria, and the CITEP Lab from Universidad de Buenos Aires for helping us in conducting these experiments. We also thank the anonymous reviewers for their feedback and suggestions.

REFERENCES

[1] Alan Agresti and Maria Kateri. 2011. *Categorical data analysis*. Springer.

[2] Luca Maria Aiello, Alain Barrat, Rossano Schifanella, Ciro Cattuto, Benjamin Markines, and Filippo Menczer. 2012. Friendship Prediction and Homophily in Social Media. *ACM Trans. Web* 6, 2, Article Article 9 (June 2012), 33 pages. DOI: http://dx.doi.org/10.1145/2180861.2180866

[3] Jisun An, Daniele Quercia, and Jon Crowcroft. 2013. Why Individuals Seek Diverse Opinions (or Why They Don't). (2013), 15–18. DOI: http://dx.doi.org/10.1145/2464464.2464493

[4] Ofer Arazy, Oded Nov, Raymond Patterson, and Lisa Yeo. 2011. Information Quality in Wikipedia: The Effects of Group Composition and Task Conflict. *Journal of Management Information Systems* 27, 4 (April 2011), 71–98.

[5] Joop Aué, Michiel Haisma, Kristín Fjóla Tómasdóttir, and Alberto Bacchelli. 2016. Social Diversity and Growth Levels of Open Source Software Projects on GitHub. In *Proceedings of the 10th ACM/IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM '16)*. ACM, New York, NY, USA, 41:1–41:6.

[6] Donald R Bacon, Kim A Stewart, and William S Silver. 1999. Lessons from the Best and Worst Student Team Experiences: How a Teacher can make the Difference. *Journal of Management Education* 23, 5 (Oct. 1999), 467–488.

[7] George A Barnett and Grace A Benefield. 2017. Predicting international Facebook ties through cultural homophily and other factors. *New Media & Society* 19, 2 (Feb. 2017), 217–239.

[8] Zoe I Barsness, Kristina A Diekmann, and Marc-David L Seidel. 2005. Motivation and Opportunity: The Role of Remote Work, Demographic Dissimilarity, and Social Network Centrality in Impression Management. *AMJ* 48, 3 (June 2005), 401–419.

[9] Suzanne T Bell. 2007. Deep-level composition variables as predictors of team performance: a meta-analysis. *Journal of Applied Psychology* 92, 3 (2007), 595.

[10] Suzanne T. Bell, Anton J. Villado, Marc A. Lukasik, Larisa Belau, and Andrea L. Briggs. 2011. Getting Specific about Demographic Diversity Variable and Team Performance Relationships: A Meta-Analysis. *Journal of Management* 37, 3 (2011), 709–743. DOI: http://dx.doi.org/10.1177/0149206310365001

[11] Jeremy Birnholtz, Colin Fitzpatrick, Mark Handel, and Jed R. Brubaker. 2014. Identity, Identification and Identifiability: The Language of Self-Presentation on a Location-Based Mobile Dating App. In *Proceedings of the 16th International Conference on Human-Computer Interaction with Mobile Devices & Services (MobileHCI '14)*. Association for Computing Machinery, New York, NY, USA, 3–12. DOI: http://dx.doi.org/10.1145/2628363.2628406

[12] Lori Boyer, Brigitta R Brunner, Tiffany Charles, and Patrice Coleman. 2006. Managing Impressions in a Virtual Environment: Is Ethnic Diversity a Self-Presentation Strategy for Colleges and Universities? *J. Comput. Mediat. Commun.* 12, 1 (Oct. 2006), 136–154.

[13] Ronald S Burt. 2005. *Brokerage and closure: An introduction to social capital*. Oxford university press.

[14] Ronald S Burt. 2009. *Structural Holes: The Social Structure of Competition*. Harvard University Press.

[15] Donn Byrne. 1971. *The Attraction Paradigm*.

[16] Traci Carte and Laku Chidambaram. 2004. A Capabilities-Based Theory of Technology Deployment in Diverse Teams: Leapfrogging the Pitfalls of Diversity and Leveraging Its Potential with Collaborative Technology. *Journal of the Association for Information Systems* 5, 11 (Dec. 2004), 4.

[17] Edward H Chang, Katherine L Milkman, Dena M Gromet, Robert W Rebele, Cade Massey, Angela L Duckworth, and Adam M Grant. 2019. The mixed effects of online diversity training. *Proc. Natl. Acad. Sci. U. S. A.* 116, 16 (April 2019), 7778–7783.

[18] Kenneth J Chapman, Matthew Meuter, Dan Toy, and Lauren Wright. 2006. Can't we pick our own groups? The influence of group selection method on group dynamics and outcomes. *Journal of Management Education* 30, 4 (2006), 557–569.

[19] Frank X. Chen, Abby C. King, and Eric B. Hekler. 2014. "healthifying" Exergames: Improving Health Outcomes through Intentional Priming. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '14)*. Association for Computing Machinery, New York, NY, USA, 1855–1864. DOI: http://dx.doi.org/10.1145/2556288.2557246

[20] Jilin Chen, Werner Geyer, Casey Dugan, Michael Muller, and Ido Guy. 2009. Make New Friends, but Keep the Old: Recommending People on Social Networking Sites. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '09)*. Association for Computing Machinery, New York, NY, USA, 201–210. DOI: http://dx.doi.org/10.1145/1518701.1518735

[21] Jilin Chen, Yuqing Ren, and John Riedl. 2010. The Effects of Diversity on Group Productivity and Member Withdrawal in Online Volunteer Groups. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '10)*. Association for Computing Machinery, New York, NY, USA, 821–830. DOI: http://dx.doi.org/10.1145/1753326.1753447

[22] Noshir Contractor, Leslie A. DeChurch, Sawant Anup, and Xiang Li. 2013. My Dream Team Assembler. (2013). http://sonic.northwestern.edu/software/c-iknow-mydreamteam/

[23] Dan Cosley, Pamela Ludford, and Loren Terveen. 2003. Studying the Effect of Similarity in Online Task-Focused Interactions. (2003). DOI: http://dx.doi.org/10.1145/958160.958212

[24] Frank Dobbin and Alexandra Kalev. 2016. Why Diversity Programs Fail. *Harvard Business Review* (July 2016).

[25] M Brent Donnellan, Frederick L Oswald, Brendan M Baird, and Richard E Lucas. 2006. The mini-IPIP scales: tiny-yet-effective measures of the Big Five factors of personality. *Psychol. Assess.* 18, 2 (2006), 192.

[26] Geeta C D'Souza and Stephen M Colarelli. 2010. Team member selection decisions for virtual versus face-to-face teams. *Comput. Human Behav.* 26, 4 (July 2010), 630–635.

[27] Priscilla M Elsass and Laura M Graves. 1997. Demographic Diversity in Decision-Making Groups: The Experiences of Women And People of Color. *AMRO* 22, 4 (Oct. 1997), 946–973.

[28] Siamak Faridani, Ephrat Bitton, Kimiko Ryokai, and Ken Goldberg. 2010. Opinion Space: A Scalable Tool for Browsing Online Comments. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '10)*. Association for Computing Machinery, New York, NY, USA, 1175–1184. DOI: http://dx.doi.org/10.1145/1753326.1753502

[29] Gerald R Ferris, Darren C Treadway, Robert W Kolodinsky, Wayne A Hochwarter, Charles J Kacmar, Ceasar Douglas, and Dwight D Frink. 2005. Development and validation of the political skill inventory. *J. Manage.* 31, 1 (2005), 126–152.

[30] Gary Garrison, Robin L. Wakefield, Xiaobo Xu, and Sang Hyun 'Kim. 2010. Globally Distributed Teams: The Effect of Diversity on Trust, Cohesion and Individual Performance. *SIGMIS Database* 41, 3 (Aug. 2010), 27–48. DOI: http://dx.doi.org/10.1145/1851175.1851178

[31] Robert C Giambatista and Anita D Bhappu. 2010. Diversity's harvest: Interactions of diversity sources and communication technology on creative group performance. *Organ. Behav. Hum. Decis. Process.* 111, 2 (2010), 116–126.

[32] Diego Gómez-Zará, Matthew Paras, Marlon Twyman, Jacqueline N. Lane, Leslie A. DeChurch, and Noshir S. Contractor. 2019. Who Would You Like to Work With?. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. Association for Computing Machinery, New York, NY, USA, Article Paper 659, 15 pages. DOI: http://dx.doi.org/10.1145/3290605.3300889

[33] Eduardo Graells-Garrido, Mounia Lalmas, and Ricardo Baeza-Yates. 2016. Data Portraits and Intermediary Topics: Encouraging Exploration of Politically Diverse Profiles. In *Proceedings of the 21st International Conference on Intelligent User Interfaces (IUI '16)*. Association for Computing Machinery, New York, NY, USA, 228–240. DOI: http://dx.doi.org/10.1145/2856767.2856776

[34] Justin D. Hackett and Michael A. Hogg. 2014. The diversity paradox: when people who value diversity surround themselves with like-minded others. (2014). DOI:http://dx.doi.org/10.1111/jasp.12233

[35] Alexa M. Harris, Diego Gómez-Zará, Leslie A. DeChurch, and Noshir S. Contractor. 2019. Joining Together Online: The Trajectory of CSCW Scholarship on Group Formation. *Proc. ACM Hum.-Comput. Interact.* 3, CSCW, Article Article 148 (Nov. 2019), 27 pages. DOI:http://dx.doi.org/10.1145/3359250

[36] David A Harrison and Katherine J Klein. 2007. What's the difference? diversity constructs as separation, variety, or disparity in organizations. *Acad. Manage. Rev.* 32, 4 (2007), 1199–1228.

[37] David A Harrison, Kenneth H Price, and Myrtle P Bell. 1998. Beyond Relational Demography: Time and the Effects of Surface- and Deep-Level Diversity on Work Group Cohesion. (1998).

[38] Helen Ai He, Naomi Yamashita, Chat Wacharamanotham, Andrea B. Horn, Jenny Schmid, and Elaine M. Huang. 2017. Two Sides to Every Story: Mitigating Intercultural Conflict through Automated Feedback and Shared Self-Reflections in Global Virtual Teams. *Proc. ACM Hum.-Comput. Interact.* 1, CSCW, Article Article 51 (Dec. 2017), 21 pages. DOI: http://dx.doi.org/10.1145/3134686

[39] Tanja Hentschel, Meir Shemla, Jürgen Wegge, and Eric Kearney. 2013. Perceived Diversity and Team Functioning: The Role of Diversity Beliefs and Affect. *Small Group Research* 44, 1 (Feb. 2013), 33–61.

[40] Julia Himmelsbach, Stephanie Schwarz, Cornelia Gerdenitsch, Beatrix Wais-Zechmann, Jan Bobeth, and Manfred Tscheligi. 2019. Do We Care About Diversity in Human Computer Interaction: A Comprehensive Content Analysis on Diversity Dimensions in Research. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. Association for Computing Machinery, New York, NY, USA, Article Paper 490, 16 pages. DOI: http://dx.doi.org/10.1145/3290605.3300720

[41] Pamela J Hinds, Kathleen M Carley, David Krackhardt, and Doug Wholey. 2000. Choosing work group members: Balancing similarity, competence, and familiarity. *Organ. Behav. Hum. Decis. Process.* 81, 2 (2000), 226–251.

[42] Elizabeth V Hobman, Prashant Bordia, and Cynthia Gallois. 2004. Perceived Dissimilarity and Work Group Involvement: The Moderating Effects of Group Openness to Diversity. *Group & Organization Management* 29, 5 (Oct. 2004), 560–587.

[43] Sujin K Horwitz and Irwin B Horwitz. 2007. The Effects of Team Diversity on Team Outcomes: A Meta-Analytic Review of Team Demography. *J. Manage.* 33, 6 (Dec. 2007), 987–1015.

[44] Nasif Imtiaz, Justin Middleton, Joymallya Chakraborty, Neill Robson, Gina Bai, and Emerson Murphy-Hill. 2019. Investigating the Effects of Gender Bias on GitHub. In *Proceedings of the 41st International Conference on Software Engineering (ICSE '19)*. IEEE Press, Piscataway, NJ, USA, 700–711.

[45] Christine L Jackson, Jason A Colquitt, Michael J Wesson, and Cindy P Zapata-Phelan. 2006. Psychological collectivism: A measurement validation and linkage to group member performance. *J. Appl. Psychol.* 91, 4 (2006), 884.

[46] Susan E Jackson and Aparna Joshi. 2004. Diversity in social context: a multi-attribute, multilevel analysis of team diversity and sales performance. *Journal of Organizational Behavior: The International Journal of Industrial, Occupational and Organizational Psychology and Behavior* 25, 6 (2004), 675–702.

[47] Cheryl R Kaiser, Brenda Major, Ines Jurcevic, Tessa L Dover, Laura M Brady, and Jenessa R Shapiro. 2013. Presumed fair: ironic effects of organizational diversity structures. *J. Pers. Soc. Psychol.* 104, 3 (March 2013), 504–519.

[48] Young Ji Kim, David Engel, Anita Williams Woolley, Jeffrey Yu-Ting Lin, Naomi McArthur, and Thomas W. Malone. 2017. What Makes a Strong Team?: Using Collective Intelligence to Predict Team Performance in League of Legends. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing (CSCW '17)*. ACM, New York, NY, USA, 2316–2329. DOI: http://dx.doi.org/10.1145/2998181.2998185

[49] Aniket Kittur and Robert E Kraut. 2008. Harnessing the Wisdom of Crowds in Wikipedia: Quality Through Coordination. In *Proceedings of the 2008 ACM Conference on Computer Supported Cooperative Work (CSCW '08)*. ACM, New York, NY, USA, 37–46.

[50] Shirli Kopelman, J Mark Weber, and David M Messick. 2002. Factors influencing cooperation in commons dilemmas: A review of experimental psychological research. *The drama of the commons* (2002), 113–156.

[51] Gueorgi Kossinets and Duncan J Watts. 2009. Origins of Homophily in an Evolving Social Network. *Am. J. Sociol.* 115, 2 (Sept. 2009), 405–450.

[52] David Krackhardt. 1987. Cognitive social structures. *Soc. Networks* 9, 2 (June 1987), 109–134.

[53] Dora C Lau and J Keith Murnighan. 2005. Interactions within groups and subgroups: The effects of demographic faultlines. *Acad. Manage. J.* 48 (2005), 645–659.

[54] Lisa Legault, Jennifer N Gutsell, and Michael Inzlicht. 2011. Ironic effects of antiprejudice messages: how motivational interventions can reduce (but also increase) prejudice. *Psychol. Sci.* 22, 12 (Dec. 2011), 1472–1477.

[55] Sheena Lewis, Mira Dontcheva, and Elizabeth Gerber. 2011. Affective Computational Priming and Creativity. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '11)*. Association for Computing Machinery, New York, NY, USA, 735–744. DOI: http://dx.doi.org/10.1145/1978942.1979048

[56] Q. Vera Liao and Wai-Tat Fu. 2013. Beyond the Filter Bubble: Interactive Effects of Perceived Threat and Topic Involvement on Selective Exposure to Information. (2013), 2359–2368. DOI: http://dx.doi.org/10.1145/2470654.2481326

[57] Q. Vera Liao and Wai-Tat Fu. 2014. Can You Hear Me Now? Mitigating the Echo Chamber Effect by Source Position Indicators. In *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing (CSCW '14)*. Association for Computing Machinery, New York, NY, USA, 184–196. DOI:http://dx.doi.org/10.1145/2531602.2531711

[58] M R Lind. 1999. The gender impact of temporary virtual work groups. *IEEE Trans. Prof. Commun.* 42, 4 (Dec. 1999), 276–285.

[59] Ioanna Lykourentzou, Angeliki Antoniou, Yannick Naudet, and Steven P. Dow. 2016. Personality Matters: Balancing for Personality Types Leads to Better Outcomes for Crowd Teams. (2016), 260–273. DOI: http://dx.doi.org/10.1145/2818048.2819979

[60] Ioanna Lykourentzou, Robert E. Kraut, and Steven P. Dow. 2017. Team Dating Leads to Better Online Ad Hoc Collaborations. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing (CSCW '17)*. Association for Computing Machinery, New York, NY, USA, 2330–2343. DOI: http://dx.doi.org/10.1145/2998181.2998322

[61] Sakthi Mahenthiran and Pamela J Rouse. 2000. The impact of group selection on student performance and satisfaction. *International Journal of Educational Management* 14, 6 (2000), 255–265.

[62] John E Mathieu, John R Hollenbeck, Daan van Knippenberg, and Daniel R Ilgen. 2017. A century of work teams in the Journal of Applied Psychology. *J. Appl. Psychol.* 102, 3 (March 2017), 452–467.

[63] Miller McPherson, Lynn Smith-Lovin, and James M Cook. 2001. Birds of a Feather: Homophily in Social Networks. *Annu. Rev. Sociol.* 27, 1 (Aug. 2001), 415–444.

[64] Kelly A Mollica, Barbara Gray, and Linda K Treviño. 2003. Racial Homophily and Its Persistence in Newcomers' Social Networks. *Organization Science* 14, 2 (April 2003), 123–136.

[65] Jonathan T. Morgan, Michael Gilbert, David W. McDonald, and Mark Zachry. 2014. Editing beyond Articles: Diversity & Dynamics of Teamwork in Open Collaborations. In *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing (CSCW '14)*. Association for Computing Machinery, New York, NY, USA, 550–563. DOI:http://dx.doi.org/10.1145/2531602.2531654

[66] Michael D Mumford, Wayne A Baughman, K Victoria Threlfall, Charles E Uhlman, and David P Costanza. 1993. Personality, adaptability, and performance: Performance on well-defined problem sovling tasks. *Hum. Perform.* 6, 3 (1993), 241–285.

[67] Sean A. Munson and Paul Resnick. 2010. Presenting Diverse Political Opinions: How and How Much. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '10)*. Association for Computing Machinery, New York, NY, USA, 1457–1466. DOI: http://dx.doi.org/10.1145/1753326.1753543

[68] Edward Newell and Derek Ruths. 2016. How One Microtask Affects Another. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. Association for Computing Machinery, New York, NY, USA, 3155–3166. DOI: http://dx.doi.org/10.1145/2858036.2858490

[69] Mike Noon. 2018. Pointless Diversity Training: Unconscious Bias, New Racism and Agency. *Work Employ. Soc.* 32, 1 (Feb. 2018), 198–209.

[70] Ruqin Ren and Bei Yan. 2017. Crowd Diversity and Performance in Wikipedia: The Mediating Effects of Task Conflict and Communication. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. Association for Computing Machinery, New York, NY, USA, 6342–6351. DOI: http://dx.doi.org/10.1145/3025453.3025992

[71] Daniela Retelny, Sébastien Robaszkiewicz, Alexandra To, Walter S. Lasecki, Jay Patel, Negar Rahmati, Tulsee Doshi, Melissa Valentine, and Michael S. Bernstein. 2014. Expert Crowdsourcing with Flash Teams. In *Proceedings of the 27th Annual ACM Symposium on User Interface Software and Technology (UIST '14)*. Association for Computing Machinery, New York, NY, USA, 75–85. DOI: http://dx.doi.org/10.1145/2642918.2647409

[72] Lionel Robert and Daniel M. Romero. 2015. Crowd Size, Diversity and Performance. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15)*. Association for Computing Machinery, New York, NY, USA, 1379–1382. DOI: http://dx.doi.org/10.1145/2702123.2702469

[73] Lionel P. Robert. 2016. Far but Near or Near but Far? The Effects of Perceived Distance on the Relationship between Geographic Dispersion and Perceived Diversity. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. Association for Computing Machinery, New York, NY, USA, 2461–2473. DOI: http://dx.doi.org/10.1145/2858036.2858534

[74] Lionel P Robert, Jr and Daniel M Romero. 2017. The influence of diversity and experience on the effects of crowd size: When Does Crowd Size Matter? *J Assn Inf Sci Tec* 68, 2 (Feb. 2017), 321–332.

[75] Howard Ross. 2008. Exploring unconscious bias. *Diversity best Practices* (2008).

[76] Jian Raymond Rui and Michael A Stefanone. 2013. Strategic image management online: Self-presentation, self-esteem and social network perspectives. *Inf. Commun. Soc.* 16, 8 (2013), 1286–1305.

[77] Niloufar Salehi and Michael S. Bernstein. 2018. Hive: Collective Design Through Network Rotation. *Proc. ACM Hum.-Comput. Interact.* 2, CSCW, Article Article 151 (Nov. 2018), 26 pages. DOI: http://dx.doi.org/10.1145/3274420

[78] Niloufar Salehi, Andrew McCabe, Melissa Valentine, and Michael Bernstein. 2017. Huddler: Convening Stable and Familiar Crowd Teams Despite Unpredictable Availability. (2017), 1700–1713. DOI: http://dx.doi.org/10.1145/2998181.2998300

[79] Noam Scheiber and John Eligon. 2019. Elite Law Firm's All-White Partner Class Stirs Debate on Diversity. *The New York Times* (Jan. 2019).

[80] Ari Schlesinger, W. Keith Edwards, and Rebecca E. Grinter. 2017. Intersectional HCI: Engaging Identity through Gender, Race, and Class. (2017), 5412–5427. DOI:http://dx.doi.org/10.1145/3025453.3025766

[81] Meir Shemla, Bertolt Meyer, Lindred Greer, and Karen A Jehn. 2016. A review of perceived diversity in teams: Does how members perceive their team's composition affect team processes and outcomes? *J. Organ. Behav.* 37 (Feb. 2016), S89–S106.

[82] D Sandy Staples and Lina Zhao. 2006. The Effects of Cultural Diversity in Virtual Teams Versus Face-to-Face Teams. *Group Decision and Negotiation* 15, 4 (July 2006), 389–406.

[83] Henri Tajfel and John C Turner. 1979. An integrative theory of intergroup conflict. *The social psychology of intergroup relations* 33 (1979), 74.

[84] Pamela Tierney and Steven M Farmer. 2002. Creative Self-Efficacy: Its Potential Antecedents and Relationship to Creative Performance. *Acad. Manage. J.* 45, 6 (2002), 1137–1148.

[85] Chun-Hua Tsai and Peter Brusilovsky. 2018. Beyond the Ranked List: User-Driven Exploration and Diversification of Social Recommendation. In *23rd International Conference on Intelligent User Interfaces (IUI '18)*. Association for Computing Machinery, New York, NY, USA, 239–250. DOI: http://dx.doi.org/10.1145/3172944.3172959

[86] John C Turner. 1987. *Rediscovering the Social Group: A Self-categorization Theory*. B. Blackwell.

[87] Aharon Tziner and Dov Eden. 1985. Effects of crew composition on crew performance: Does the whole equal the sum of its parts? *J. Appl. Psychol.* 70, 1 (1985), 85.

[88] Brian Uzzi, Satyam Mukherjee, Michael Stringer, and Ben Jones. 2013. Atypical Combinations and Scientific Impact. *Science* 342, 6157 (2013), 468–472.

[89] Daan Van Knippenberg, Carsten K W De Dreu, and Astrid C Homan. 2004. Work group diversity and group performance: an integrative model and research agenda. *J. Appl. Psychol.* 89, 6 (2004), 1008.

[90] Daan van Knippenberg and Michaéla C Schippers. 2007. Work group diversity. *Annu. Rev. Psychol.* 58 (2007), 515–541.

[91] Bogdan Vasilescu, Daryl Posnett, Baishakhi Ray, Mark G.J. van den Brand, Alexander Serebrenik, Premkumar Devanbu, and Vladimir Filkov. 2015a. Gender and Tenure Diversity in GitHub Teams. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15)*. Association for Computing Machinery, New York, NY, USA, 3789–3798. DOI: http://dx.doi.org/10.1145/2702123.2702549

[92] Bogdan Vasilescu, Alexander Serebrenik, and Vladimir Filkov. 2015b. A Data Set for Social Diversity Studies of GitHub Teams. In *Proceedings of the 12th Working Conference on Mining Software Repositories (MSR '15)*. IEEE Press, Piscataway, NJ, USA, 514–517.

[93] Hao-Chuan Wang, Susan R. Fussell, and Dan Cosley. 2011. From Diversity to Creativity: Stimulating Group Brainstorming with Cultural Differences and Conversationally-Retrieved Pictures. In *Proceedings of the ACM 2011 Conference on Computer Supported Cooperative Work (CSCW '11)*. Association for Computing Machinery, New York, NY, USA, 265–274. DOI: http://dx.doi.org/10.1145/1958824.1958864

[94] Katherine Williams Williams and Charles A O'Reilly. 1998. Demography and Diversity in Organisations: A review of 40 years of research in BM Staw and LL Cummings (eds) Research in Organisational Behaviour Vol. 20. *Jai Pres, Connecticut* (1998).

[95] David Wong, Siamak Faridani, Ephrat Bitton, Björn Hartmann, and Ken Goldberg. 2011. The Diversity Donut: Enabling Participant Control over the Diversity of Recommended Responses. (2011), 1471–1476. DOI: http://dx.doi.org/10.1145/1979742.1979793

[96] Anita Williams Woolley, Christopher F Chabris, Alex Pentland, Nada Hashmi, and Thomas W Malone. 2010. Evidence for a Collective Intelligence Factor in the Performance of Human Groups. *Science* 330, 6004 (2010), 686–688.

[97] Ziang Xiao, Michelle X. Zhou, and Wat-Tat Fu. 2019. Who Should Be My Teammates: Using a Conversational Agent to Understand Individuals and Help Teaming. In *Proceedings of the 24th International Conference on Intelligent User Interfaces (IUI '19)*. Association for Computing Machinery, New York, NY, USA, 437–447. DOI: http://dx.doi.org/10.1145/3301275.3302264

[98] Chia-Chen Yang and B Bradford Brown. 2016. Online Self-Presentation on Facebook and Self Development During the College Transition. *J. Youth Adolesc.* 45, 2 (Feb. 2016), 402–416.

[99] Teng Ye and Lionel P. Robert. 2017. Does Collectivism Inhibit Individual Creativity? The Effects of Collectivism and Perceived Diversity on Individual Creativity and Satisfaction in Virtual Ideation Teams. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing (CSCW '17)*. Association for Computing Machinery, New York, NY, USA, 2344–2358. DOI: http://dx.doi.org/10.1145/2998181.2998261

[100] Chunqiu Zeng, Qing Wang, Shekoofeh Mokhtari, and Tao Li. 2016. Online Context-Aware Recommendation with Time Varying Multi-Armed Bandit. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '16)*. Association for Computing Machinery, New York, NY, USA, 2025–2034. DOI: http://dx.doi.org/10.1145/2939672.2939878

[101] Sharon Zhou, Melissa Valentine, and Michael S. Bernstein. 2018. In Search of the Dream Team: Temporally Constrained Multi-Armed Bandits for Identifying Effective Team Structures. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. Association for Computing Machinery, New York, NY, USA, Article Paper 108, 13 pages. DOI: http://dx.doi.org/10.1145/3173574.3173682