The Impact of Digital Nudging Techniques on the Formation of Self-Assembled Crowd Project Teams

Federica Lucia Vinella
Utrecht University
Utrecht, The Netherlands
f.l.vinella@uu.nl

Rosa Mosch
Utrecht University
Utrecht, The Netherlands
r.l.mosch@uu.nl

Ioanna Lykourentzou
Utrecht University
Utrecht, The Netherlands
f.l.vinella@uu.nl

Judith Masthoff
Utrecht University
Utrecht, The Netherlands
f.l.vinella@uu.nl

ABSTRACT

Self-assembling team formation systems, where online users can select their teammates, are gaining research and industry interest. Still, the benefits of diversity remain frequently untapped for these teams, as people tend to choose others similar to them. In this study, we examine whether making users aware of the team's diversity can impact their selections. In a study involving 120 crowd participants, working on the scenario of a crowdsourced innovation project, we tested the effects of two choice architecture and nudging techniques. The first technique displayed explicit personalized diversity information in the form of the current team diversity score and diversity recommendations. The second technique used diversity priming, in the form of counter-stereotypes and all-inclusive multiculturalism. Our results indicate that, while priming deterred participants from picking teammates of different regions, displaying diversity information was the only factor to positively enhance diverse choices. These results were not moderated by the users' ''need to belong" levels, an intrinsic motivation justifying one's need to form social ties. Other factors which we also find to predict selection behavior were the participants' region of origin, participants' gender, teammates' functional backgrounds, and teammates' order of appearance. In light of these findings, we suggest that nudging techniques need to be cautiously applied to online team formation as the different techniques differ in their ability to evoke diversity among intrinsically diverse crowds, and that personalised displaying of diversity information seems most promising.

CCS CONCEPTS

• **Human-centered computing → Collaborative and social computing**; **Social networking sites**.

KEYWORDS

diversity, teams, self-assembly, nudging, priming, counter-stereotypes

UMAP '22, July 4–7, 2022, Barcelona, Spain

ACM ISBN 978-1-4503-9207-5/22/07…$15.00
https://doi.org/10.1145/3503252.3531298

ACM Reference Format:
Federica Lucia Vinella, Rosa Mosch, Ioanna Lykourentzou, and Judith Masthoff. 2022. The Impact of Digital Nudging Techniques on the Formation of Self-Assembled Crowd Project Teams. In *Proceedings of the 30th ACM Conference on User Modeling, Adaptation and Personalization (UMAP '22), July 4–7, 2022, Barcelona, Spain.* ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/3503252.3531298

1 INTRODUCTION

With a growing international outlook to doing business and outsourcing innovation, diversity and inclusiveness have become substantial parts of most companies' assessments and progress reports [29, 49] while pro-diversity managerial practices are also on the rise [77]. Still, employers can mistakenly overlook employees' homophilic preferences for collaborators. Outsourced crowdsourcing teams can also be subject to homophilic biases and stereotypes while self-assembling and self-organizing [29]. Persisting biases can trigger practices responsible for marginalizing contributors from different backgrounds. Yet, team diversity – especially in open collaboration and crowdsourced innovation projects (CIPs) — is often one of the best assets of crowd collaborative labor [14, 61]. Teams heterogeneous in skills, tenure, and geo-location tend to outperform homogeneous ones in complex and creative tasks [14, 45]. Ideologically polarized Wikipedia teams, such as those composed of the most diverse political slants, are substantially more constructive, competitive, and focused than ideologically homogeneous ones [61]. Despite communication-inhibiting factors [14], diversity aids creative and innovative solutions to complex, open-ended problems [14, 61]. Considering several advantages of team diversity within crowd collaboration [14, 45] and the capacity of digital interfaces to connect diverse collaborators, we ask the following: *how can open collaborative tools support the formation of more diverse crowd project teams?*

Interfaces are known to condition users' choices [46]. The very presence of information while making decisions online can prime users to deviate their behaviour toward an intended outcome. Images and content prime people into building up assumptions and expectations that guide their thought associations. Gómez-Zará et al. show that profiles with high diversity scores are less likely to be chosen by university students forming online teams [20]. These results indicate that empirical research is often fundamentally needed to test the effects of otherwise intuitive – yet unproven

– digital nudging approaches to diversity policies. Combining a growing managerial emphasis on organizational diversity with the growth of open collaborations and CIPs, we identify a gap in the literature regarding interventions designed to safeguard diversity among self-assembled crowd teams. By self-assembled teams, we mean those teams generated through a bottom-up process where actors self-organize [20, 55]. In a scenario where people choose *"the best person for the job"* [20, 25], we aim to observe to what degree participants made choices based upon surface-level diversity of their teammates (complexion, gender) versus their deep-level traits (skills and level of education).

We present a study on the impact of priming and diversity info (two digital nudging techniques) on the formation of teams for an outsourced CIP focusing on a creative complex task representative of crowdsourcing open, diverse ideation and design thinking [40]. Our research questions and hypotheses can be summarized as follows. First, we ask: *how do priming and displaying diversity info affect the diversity of crowd users' choice of teammates?* This question concerns separately two nudging techniques: *priming interventions* (RQ1) and *displaying diversity info* (RQ2). Next, we assess how the combination of priming and diversity info affects users' diverse choices (RQ3) and hypothesize that certain types of Priming (AIM and counter-stereotypes) incentivize crowds in choosing diverse teammates (H1). Based on Gómez-Zará et al.'s work [20], we also hypothesize that displaying diversity info (DI) deters users from choosing more diverse teammates (H2). Furthermore, we hypothesize that one's tendency to choose less diverse teammates – when in the presence of DI – is moderated by their need to belong (NTB) (H3). This last hypothesis concerns the moderating influence of NTB in settings where diversity is encouraged through a choice architecture [39]. Lastly, we hypothesize that a combination of Priming and Diversity Info (Priming + DI) nudges participants toward more diverse teams (H4).

We analyzed the effects of three conditions plus control. These conditions are: 1) a Diversity Info condition showing personalized diversity scores and recommendations (DI), 2) a Priming condition containing All-Inclusive Multiculturalism (AIM) and counter-stereotypes designed to nudge diverse choices called Priming, 3) a condition combining diversity scores, recommendations, and priming together (Priming + DI). We recruited 120 crowd participants to autonomously assemble virtual teams comprising of two teammates (plus themselves). Crowd users could select from a list of thirty international and heterogeneous profiles designed to represent candidate teammates whose profiles showed their gender, ethnicity, age, education, and functional background, and region. Participants were tested for their understanding of the task and completeness of their information (manipulation checks) and their levels of need to belong [42]. We found that personalized Diversity Info (DI) positively affected heterogeneity, in contrast with the findings from Gómez-Zará et al. [20]. Nonetheless, participants still chose primarily according to homophilic preferences of gender and region of origin, while the type of task (creating a slogan for a coffee company) yet seemingly drove crowd participants to choose specific functional backgrounds over others (above all sales). This study offers insights into how socio-technical team formation systems can contribute to more diverse teams in open collaboration for CIPs. It builds upon previous research on digital diversity interventions [20] and aims to shed light on how technology can play a role in attentively stimulating diversity among crowd collaborators. Furthermore, it identifies which digital interventions among priming techniques and diversity info (including recommendations) could adversely affect diverse choices. The rest of the paper is as follows: Section 2 covers the theoretical framework on diversity with a justification of the research questions and corresponding hypotheses. Section 3 presents the study design. Section 4 analyses the results and Section 5 discusses these along with system design recommendations gathered from the study. Section 6 concludes the paper.

2 THEORETICAL FRAMEWORK AND HYPOTHESES

2.1 Diversity in crowdsourcing teams

Crowd teams formed through CIPs share qualities that can be summarized as follows: 1) They are competitive and collaborative [52], 2) have no size limit unless specified by the requester, 3) have no hierarchical structure [78], 4) have voluntary ad-hoc membership with fluid boundaries [59], and 5) have no necessary division of duties since collaborators self-coordinate in a fully autonomous fashion [22]. Through CIPs, the crowd is in charge of finding collaborators and is expected to generate innovative solutions. Furthermore, Open collaboration CIPs advance from the knowledge that open-ended problems are better suited to diverse expertise and skills most commonly deriving from a diverse crowd. Analyzing Wikipedia talk pages on a large scale, Shi et al. [61] found that articles with higher debate intensity, lexical and semantic diversity were prevailingly authored by politically polarised – hence highly diverse – groups of collaborators. Editing contested topics required a balancing act from politically polarized contributors that ameliorated conflicting points of view. It is precisely this active engagement of politically diverse opinions that is often lacking in more homogeneous communities such as echo chamber platforms [68]. However, not all kinds of diversity are advantageous to teams as there are dimensions of collaboration that benefit from homogeneity. While most deep-level diversity, including skill sets and tenure, facilitate creative-problem solving in crowdsourcing environments [14], team homophily (or similarity) regarding other attributes (language, geographical proximity, familiarity) helps with communication and coordination [14]. The advantage of homophily may be due to shared and acquired characteristics – such as having worked together in the past or having common language and customs – encouraging team synergy around communication and coordination.

2.2 Nudging through priming

Digital nudging is a subset of persuasive/computing technologies making use of digital interfaces to deliver suggestions and positive reinforcements as ways to direct individuals' behavior toward an intended outcome [67]. Several types of digital nudging techniques have been designed to sway attention and direct users' behavior such as default options, positioning, explanations, and decoys [76]. While digital nudging is present in numerous contexts such as sustainability and well-being, little research has covered its effects within online diversity settings. Aside from the study by Gómez-Zará et al. [20] on displaying diversity scores in an online student team formation scenario, we identify a research gap concerning digital nudging techniques in open collaboration and CIPs. Our research question is as follows:

RQ: How do Priming and Diversity Info affect the diversity of crowd users' team members choice?

We evaluate two digital choice architecture techniques: diversity priming and explicit information. We test their effectiveness at nudging online crowd participants according to the diversity levels of their chosen teammates compared to no nudging. Furthermore, we select only a subset of techniques for each nudging intervention. Priming, in the context of digital nudging for diversity, is herein achieved through the exposure of *counter-stereotypes* and *all-inclusive multiculturalism.* Conversely, diversity info is the *display of diverse attributes* and *recommendations of diverse teammates.* As part of the main research question on the effects of nudging techniques we ask the following:

RQ1: (How) does Priming affect the diversity of the members that crowd users select for their team?

This question singles out digital diversity priming as a way of modifying online users' behavior by directing their choices toward more diverse teammates. More generally, priming is a technique consisting of *"[...] activating memory contents by experimental stimuli (or primes) that are unobtrusively (or even subliminally) presented to respondents in experiments"* [37]. Priming is the use of initial stimuli to condition individuals before a task. Controlled stimuli are designed to 'prime' one's behavior to act in a certain way. Priming can be either subliminal (the subject is not aware of being primed) or informed (acknowledged by the subject). In this research, we focus on priming as conceptual stimuli in the form of contextualization of information to trigger positive associations toward diverse individuals. By exposing crowd users to diversity as explicit info and positive representations of a work culture, we expect them to favor diversity while searching for teammates online. Priming for diversity takes different shapes: from drawing attention to historical injustices to making people recognize their implicit biases while motivating them to act more ethically [10, 73].While in the context of diversity implementing priming techniques has only been hinted at in the past [20, 29], we intend to enlarge the discussion by looking specifically at two main priming techniques designed to increase diversity: **exposure to counter-stereotypes** and the **all-inclusive multiculturalism approach**.

Counter-stereotypes, meaning exposing people to positive examples from under-represented social groups, is known to be particularly effective at curbing biases and stereotypes [3, 8, 11, 12, 17, 18, 23]. An example of an effective counter-stereotype is displaying images of female scientists in STEM textbooks. Female students showed higher comprehension of science lessons after exposure to this counter-stereotype compared to reading texts with gender stereotypical images of male scientists [21]. Even exposing people to the thought of counter-stereotypes has been seen to compel them to abandon the use of categorical labels [26, 30, 38], and develop cognitive flexibility, and creativity [19, 38]. Counter-stereotypes benefit not only the reduction of one's access to stereotypical thoughts [26], but also to reduce stereotype threats, meaning the pernicious effects that stereotypes have on the performance of the subject of stereotypes [9]. Exposing individuals to counter-stereotypical images of underrepresented groups can trigger positive automatic associations that can increase positive feelings toward diverse cultures, ethnicity, and genders [1].

All-inclusive multiculturalism (AIM), on the contrary, is an approach to mitigate the effects of stereotypes by explicitly mentioning both majority and minority groups [32, 56]. AIM is also a response to the limitations faced by the two most common initiatives against stereotyping at work, that is, ethnic color-blindness and multiculturalism. To circumvent the feelings of exclusion that other organizational members might feel in front of multicultural and color-blind agendas, the AIM model proposes that diversity includes *all* employees [65] An example of AIM is explicitly affirming the inclusion of non-minorities (e.g., Dutch workers in a Dutch company) within a general multiculturalism ideology [32]. On the one hand, the AIM approach celebrates differences between individuals and social groups, which is part of acknowledging minorities; on the other hand, by explicitly mentioning the essential role that non-minorities play in the workplace, it limits feelings of exclusion and preferential treatments [65]. Following related work on the effectiveness of priming techniques, we hypothesize that:

H1: Priming crowd participants toward diversity leads them to select more diverse team members.

Our choice of priming mechanisms, counter-stereotypes and AIM, tackles potential obstacles for choosing more diverse team members. Counter-stereotypes focus on avoiding negative implicit attitudes and activating positive implicit attitudes. AIM focuses more on the feeling of exclusion of majority demographics caused by the framing of diversity policies. Through combining these techniques, we expect digital priming to be effective for motivating crowd users to choose diverse teammates.

2.3 Nudging through explicit information

While users cannot be forced to choose diversely, and companies and systems' owners should refrain from censorship, the choice of how and what info gets displayed through interfaces can greatly affect decision-making processes [27]. We consider the overt display of diversity info (DI) as a possible way to nudge users [27]. Thus, our second sub-question is:

RQ2: (How) does displaying DI affect the diversity of the members that crowd users select for their teams?

In this question, we consider nudging by way of eliciting information regarding diversity. Information elicitation – or explicit information – discloses, or heightens, topical information intended to change users' awareness of issues while nudging their choices toward the desired outcome. For this study, we define two explicit information techniques namely **exposure of attributes** (or display of users' information), and **recommendations** (a concept borrowed from person-to-person recommender systems). By *display of attributes* we intend the presentation and framing of information addressing online social stereotypes and users' homophilic tendencies [58, 63]. Gómez-Zará et al. [20]'s work demonstrates that the presence of diversity scores within teammates' recommender systems can be disadvantageous to diversity, as collaborators favor others who are similar to them, more so than in scenarios where no diversity scores are given. Aside from these results [20], very few other researchers focus on diversity when looking at the repercussions of the exposure of the users' attributes on their diversity choices.

However, displaying attributes that highlight racial minority and cultural facets might not always be detrimental to diversity choices. [1] Walker et al.'s study on online hiring decisions shows that personal references such as video testimonials were fruitful at yielding more diverse employees [75]. In terms of racial cues, Walker et al. demonstrate that recruitment websites containing racial diversity cues were more extensively browsed and remembered – particularly by Black participants – than those that lacked diversity references [74]. In the light of these divergent findings, our study brings back the focus on the effects of displaying DI in terms of percentage scores, as done by Gómez-Zará et al. [20] (shown as aggregate measures of the teammates' attributes), yet focusing on open collaboration through CIPs and compared to digital priming. Therefore, for the second hypothesis, we expect to observe the negative effects of displaying DI (as seen in the study by Gómez-Zará et al. [20]) in our DI condition. We formulate H2 as:

H2: Displaying explicit DI leads crowd users to select less diverse team members.

Aside from using nudging techniques to foster inclusiveness, we consider possible moderating effects deriving from people's predispositions to social affiliation. We expect that some people adapt to the company's values and norms, even against their judgement, as long as it improves their likelihood of acceptance and inclusion [43]. To test this, we look at levels of *need to belong* (NTB) understood as an essential need to form and maintain social relationships [70]. Thus, we expect that using explicit diversity info on team formation systems persuades people with high need to belong to comply with the company's values. Our third hypothesis is as following:

H3: Crowd participants with a high NTB choose more diverse team members when DI is present.

Contiguous to diversity info, we include *recommendations as digital nudging interventions. By recommendations, we indent personalized suggestions* given to users, which in our case are the diversity traits of the teammate's profiles. Recommendations are common among online dating websites (and mobile apps), where users get recommended to matches based on their preferences (content-based filtering), their similarity of choices (collaborative filtering), their similarity of attributes (demographic filtering), or a combination of those techniques. Recommendations prominence (how bigger or brighter they are compared to the rest of the options), and their position in the list (ranking), are two of the most common methods used when presenting highly recommended choices [36]. Using explanations is also extensively used in person-to-product recommender systems [69], albeit less so for person-to-person recommendations [35]. The use of explanations in reciprocal environments such as recruitment sites, or dating apps, have been noticed to be as persuasive as the order of the presentation of the items [35]; this is particularly true when the costs associated with making that choice

[1] In their study, Walker et al. manipulated four employees on a hypothetical organization's recruitment website as either all White (no racial diversity cue) or two White and two Black (racial diversity cue) whilst holding the gender ratio constant (2 men and 2 women).

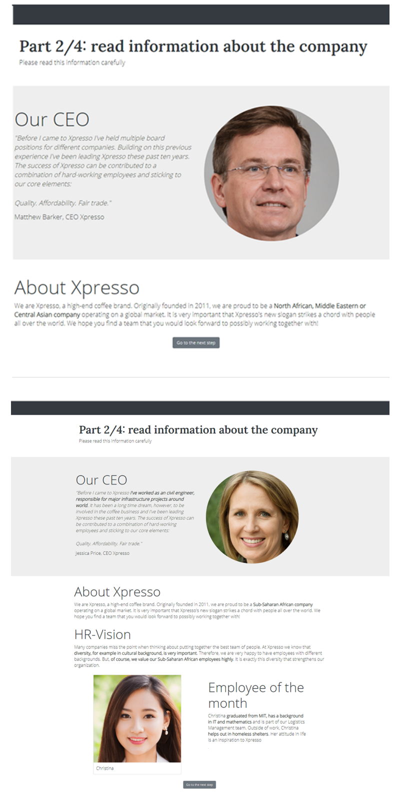

Figure 1: Overview of the information pages about the Xpresso company across Conditions. To the left,the Control and DI: information page with a white male CEO. To the right, Priming and Priming + DI: information page adjusted to crowd users' demographics.

is significant to that user (in the form of a monetary, or emotional investment).

After looking at priming and explicit information as two ways of nudging crowd teams toward inclusiveness, we summarize the effects of these techniques into one research question paired with the corresponding hypothesis:

RQ3: (How) does the combination of Priming and DI (Priming + DI) affect the diversity of team members that crowd users select for their teams? Addressing this final comparative question, and following the logic presented for the previous hypotheses, we expect that combining nudging techniques is sufficient at increasing crowd users' propensity to choose diverse teams. Thus, our last hypothesis reads as follows.

H4: Priming crowd participants in combination with diversity info (Priming + DI) lead users to select more diverse team members compared to no diversity info and priming.

3 STUDY DESIGN

The procedure contained seven steps: 1) Informed consent and task description, 2) registration, 3) information about the requester

(Xpresso company), 4) manipulation check, 5) teammates selection from a collection of profiles (presented in a different random order per participant), 6) need to belong questionnaire, 7) end of the task and thank you page. The task requested participants to form teams to write a coffee slogan: "*We are Xpresso, a coffee company, and we are looking for a new company slogan. We need new, fresh ideas, which is why we decided to outsource this project. Your task is to select two team members from a list of previously registered people to form a team. [...]*". The slogan-task was inspired by previous research on crowdsourced team formation [44].

3.1 Participants

The study drew from a sample of 150 people of which 30 were excluded[2]. With the intent to capture a diverse pool of crowd workers, we hired participants from two of the most popular online crowdsourcing platforms namely Amazon Mechanical Turk [51] (*n*=57) and Prolific [53] (*n*=60). The remaining 3 participants were recruited via personal invite. Most participants were Western European (*n*=38), North American (*n*=29), or South Asian (*n*=27). Others were from Eastern Europe (*n*=9), Southern-Europe (*n*=6), South-East Europe (*n*=4), South America (*n*=2). Only one participant was from the remaining zones of origin [3]. The sample was predominately male (*n*=77). All participants provided informed consent and received 5 USD[4].

3.2 Research design

To test whether participants could be primed to make more diverse choices, we used a 2x2 factorial design. Participants were randomly assigned to one of the conditions. The factorial design allowed to observe the independent and interaction effects of priming and displaying DI on the choices of teammates [5]. The independent variables were **Priming**, and **DI**; each could be present (Applied) or not (None). The factorial design resulted in the following conditions.

3.2.1 Control. Displays neither the DI nor priming information toward diversity. In the information page about the company, the CEO of the company, Matthew Barker (average white, male, see Figure 1a) is described by his position and expectations with regards to the task.

3.2.2 DI. As the control except for changes in the team selection page, once participants had added a member to their team. Firstly, a progress bar showed the team's diversity as an aggregate measure of the Blau score (Section 3.4) calculated on a scale of 1-100 (see Figure 2a). Secondly, dummy profiles were recommended via a banner: *Add me for a more diverse team* (see Figure 2b). A dummy was recommended if adding it to the team pushed diversity above 75%. DI scores adjusted according to the users' and teammates' attributes.

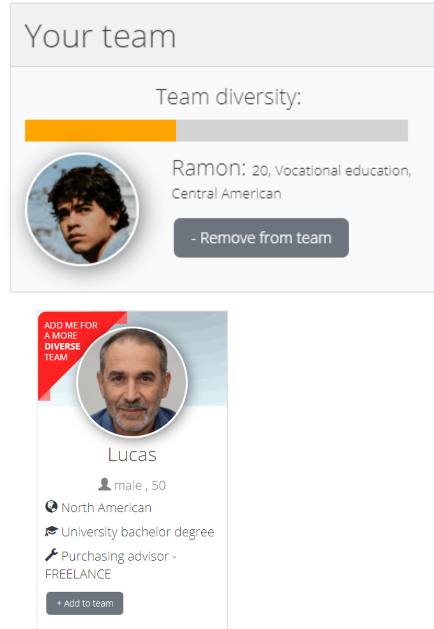

Figure 2: Overview of the teammates profiles showing the diversity score as progress bar (left fig.) and explicit recommendation of a teammate in the form of a red banner (right fig.) on the top-left side of the profile. Two diversity nudging interventions part of DI and Priming + DI.

3.2.3 Priming. As the control except that it implements counter-stereotypes (female CEO, minority employee of the month[5]) and AIM (HR-vision statement[6]) (see Figure 1b). We consider these techniques as priming since they expose users to conceptual stimuli (alternative role models, cultural inclusiveness) that may influence their responses to subsequent stimuli (choice of diverse attributes). The CEO and employee of the month pictures were on top of the information page, the manipulation check page, and the team selection page. The HR-vision statement was on the information page, and on top of the manipulation check and team selection page. Priming interventions adjusted to the users' attributes (e.g., if the participant was White, the counter-stereotype showed a non-White ethnic background).

3.2.4 Priming + DI. Combines Priming and DI, with the same diversity information on the team selection page as DI, and the same counter-stereotypes and AIM characteristics as Priming.

[2]5 excluded participants belonged to the control condition, 5 to DI, 7 to Priming, 8 to Priming+DI. Criteria for exclusion were: 1) incomplete submission, 2) incorrect answers to all manipulation checks, 3) clear lack of engagement (only clicking on top of list). Although costly, the latter criteria was intended to exclude outliers from the results from those participants that did not browse through the whole list of teammates.
[3]The zones of origin are intended as geographic regions and do not represent the participants' race.
[4]the average payment for crowd-sourced work [6] and meeting ethical minimum wage requirements

[5]The minority employee was a counter-stereotypical women in science, whose ethnicity was adapted to the participant, to be visibly different than their own (so, in the minority when combined with the HR statement)
[6]HR statement: "*Many companies miss the point when thinking about putting together the best team of people. At Xpresso we know that diversity, for example in cultural background, is very important. Therefore, we are very happy to have employees with different backgrounds. But, of course, we value our [PARTICIPANT's REGION] employees highly. It is exactly this diversity that strengthens our organization*". This statement was adapted to the participants' region, so that they would see themselves as part of the majority

3.3 Materials

We created 30 realistic-looking teammate dummy profiles (see example in Figure 2b). These were each assigned the relevant attribute characteristics [7], gender (male (*n=14*), female (*n=15*), other (*n=1*)), functional background (10 types (*n=3*)), region (European (*n=7*), North African, Middle Eastern or Central Asia (*n=1*), Latin American (*n=3*), East Asian (*n=3*), South and South-East Asian (*n=7*), Caribean (*n=1*), Sub-Saharan African (*n=1*), North American and Australasian (*n=7*)), and ethnicity (White (*n=12*), Black (*n=4*), Asian (*n=11*), Latino (*n=3*)). For an overview of the attributes, see Section 3.4). The profile pictures were partly AI-generated [33] and partly acquired as royalty-free pictures [72]. Between 30-40% of the photos were distorted or colorized to resemble as closely as possible the level of variance and individuality of profiles that one would expect from real-life matchmaking platforms. The dummy names were common for the region they supposedly came from [41]. Dummy attributes, such as age, region, and ethnicity, were not equally distributed but assigned to the 30 profiles based on the population statistics of workers from crowdsourcing platforms such as Amazon Mechanical Turk, predominantly showing Indians and North-Americans from the millennial generation [13]. Furthermore, limiting the number of profiles to 30 was the result of a trade-off between representing as many combinations of diversity attributes and ensuring participants had the opportunity to look at and assess all the profiles.

3.4 Metrics

3.4.1 Dependent measure. Our metrics are based upon Gómez-Zará et al. [20]'s study design calculating team diversity as an aggregate measure. Teams consisted of two dummy profiles chosen by participants plus themselves. We chose to study diversity among crowd teams of size three since we wanted to provide an initial analysis of a basic team unit but also avoided studying dyads as often a debated subject in the field of CSCW research on group formation [47].

Team diversity for each attribute was calculated using the commonly used Blau index [4, 20, 34], which is formulated as: $1 - \sum_{i=1}^{k} Pi^2$. The Pi corresponds to the proportion of team members in the $i-th$ category, k refers to the number of categories of that particular attribute [62]. Blau's index calculates a diversity score between 0 and 1 for each attribute and is a measure for diversity as variety [24]. After calculating Blau's score for each attribute. We normalized these scores by multiplying the scores with their respective maximum categories ($\frac{K}{K-1}$). We derived diversity from the sum of the normalized scores divided by the total number of the included attributes. While most diversity traits were categorical, age was calculated as range or generations to circumvent the problem of a sparse matrix of continuous data [50].

3.4.2 Independent measures. To measure diversity, we used the following independent measures gathered from the participants: **1) Surface-level traits**: age and gender, **2) Deep-level traits**: functional background (meaning skills and knowledge) and level of education, and **3) Surface and deep-level traits**: cultural background. We also evaluate their **need to belong**.

Surface-level (relations-oriented) traits: age and gender. Age is treated here as a measure for differences of years within a team in terms of variety. As a continuous metric, age first had to be converted into a categorical variable via discretization [50]. Based on Ferrero-Ferrero et al. [15]'s classification of generations, we categorized the participants and the dummies' ages into five generations: Greatest Generation/Silent (aged 76+ in 2021), Boomers (57-75), Generation X (41-56), Millenials (25-40), and Generation Z (18-24). This grouping helped with the clustering of the subjects in terms of their generational differences, especially with regards to their values, trust of authority, and independent thinking [15, 66, 71]. Gender, being by definition categorical, did not require additional discretization.

Deep-level (task-relevant) traits: functional background and level of education. For the classification of the functional backgrounds, we revised the nine categories by Pegels et al. [54] into the following ten: 1) Information systems, 2) Customer service, 3) Sales and marketing, 4) Engineering, R&D, 5) Purchasing/Procurement, 6) Operations, administrations or manufacturing, 7) Consultancy, 8) HR/personnel, 9) General management, 10) Creative sector. To ensure that the potential combination of functional backgrounds could be observed in practice [50], we revised the list by focusing on the (implicit) types of skills required for each sector. Additionally, manufacturing was added to the revised list as it was not present in the original version by Pegels et al. [54]. To validate the applicability of this classification, we checked these revised categories against Indeed [31]'s list of most popular jobs in the United States as of 2020 [31].

For levels of education, we chose the following: (1) Primary education, 2) High school Priming + Diploma, 3) Vocational education, 4) University bachelor degree, and 5) 5-year university degree or Ph.D.

Surface and deep-level (relations-oriented) traits: cultural background. To capture both surface and deep-level attributes of cultural background, we calculated this metric as the mean cultural background diversity of two other features: ethnicity and region. We classified both participants and dummies into one of five categories: Asian, Black, Brown, Latino, and White [16]. For the calculation of the region of origin, we settled for regional data [7, 60] from the *European Standard Classification of Cultural and Ethnic Groups* adopted by Schneider and Heath [60]. This categorization does not center around nationalities, as it combines regional and cultural aspects – such as religions – to provide a rather exclusive and complete list that captures more than just surface-level aspects of cultural background [25, 64].

Need to Belong. This measure describes people's need to feel included [2]. People with a high need to belong may be more susceptible to diversity statements containing, for instance, explicit references to AIM [32]. We measured participants' NTB levels through Leary et al.'s ten statements 5-point Likert scale [42] (used as an ordinal scale). The scale was used to gauge participants' differences in the levels of need to belong that might have moderated their interest in working with others – or for organizations – that expose them to counter-stereotypes and AIM [56].

[7] The dummy profiles characteristics were distributed as follows. Age (Generation Z (*n=9*), Millennials (*n=10*), Generation X (*n=11*))

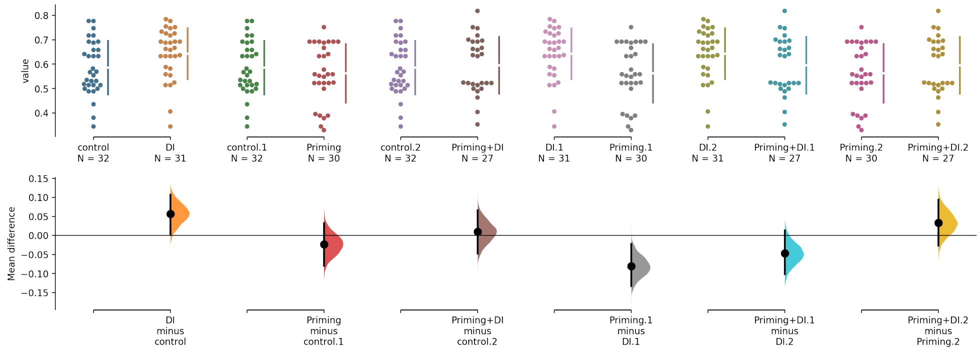

Figure 3: The mean difference for 6 comparisons are shown in the Cumming estimation plot. The raw data is plotted on the upper axes; each mean difference is plotted on the lower axes as a bootstrap sampling distribution. Mean differences are depicted as dots; 95% confidence intervals are indicated by the ends of the vertical error bars.

4 RESULTS

Section 4.1 broadly addresses the main research question (*How does displaying diversity info or priming diversity affect users choose more diverse team members?*); it concerns the testing of all four hypotheses listed in the introduction. Section 4.2 presents a posthoc analysis along with any secondary effects of participants' and dummies' characteristics. We use Kruskal-Wallis tests analyzing independent factors. Particularly, we look at crowd users' region of origin (Section 4.2.1). With an unpaired one-tailed t-test we investigate dummy profiles' attributes that contributed to their popularity (Section 4.2.2). A two-sample paired Wilcoxon test highlighted differences in gender preferences and the presence of gender-driven homophily (Section 4.2.3). Finally, we used a linear regression model to evaluate possible presentation biases in the study design confirmed by dummies' popularity (Section 4.2.4).

4.1 Hypotheses testing

4.1.1 Displaying diversity info positively affects diversity. Given the non-normality of the data (Shapiro-Wilk normality test, p = 7.166e-4) and the factorial design of the study, we could not run a two-way ANOVA but used a Mann Whitney U test instead. DI, addressed by H2, was the only condition to yield statistically significant results (Mann Whitney U=2205.00, p=0.033). This finding implies that the sole display of diversity info is sufficient to impact the diverse choice of teammates. Given that both Priming and Priming + DI did not yield statistically significant results (Mann Whitney U=1542.00, p=0.179), our study cannot confirm potential positive effects of Priming and Priming + DI on diversity.

As there have been drawbacks noted to the use of traditional statistical hypothesis testing, we also provide an estimation statistics analysis which is more focused on effect-sizes. Figure 3 shows the data from a multiple-two group analysis with estimation statistics[8] [28]. We compare conditions (control, DI, Priming, Priming + DI) and display the results through a Cumming estimation plot of several sets of two-groups data, enabling pair-wise comparison of mean differences (Table 1)[9]. This confirms the earlier Mann-Whitney result that DI resulted in more diverse teams. Priming resulted in significantly less diverse teams than DI, and seems to have had a small negative impact on diversity, which only became significant when comparing it to the condition that had a positive impact.

Additionally, we ran Kruskal-Wallis tests for each separate diversity attribute. We examined whether the treatment conditions significantly affected team diversity, independently. The tests showed no significant differences between the control and treatment conditions for gender diversity (p=0.068), ethnicity diversity (p=0.219), age diversity (p=0.242), education diversity (p=0.546) and functional background (p=0.491), except for the region of origin. We found that differences in the region of origin were negatively affected by Priming, compared to the control (adjusted p= 0.038).

Results on the effects of Priming and DI show that **although Priming did not positively affect diversity choices, the display of diversity info (DI) positively impacted diversity. This result contradicts previous findings [20]. We also observed that through Priming (counter-stereotypes and AIM), users were less likely to choose teammates from other regions.**

4.1.2 Need to belong does not affect the diversity choice. To ground H3 (*People with a high need to belong choose more diverse team members with diversity info.*), we used the results from the NTB scale[10] that checked for biases and tendencies against (or toward) diversity. After transforming the data of the reversed statements, we calculated the median (2.99), the mean (3.1) and the 75th percentile (3.3) of the NTB scores, giving us three ways to classify participants as high or low. We considered participants with a score higher than the median/mean/75th percentile to have high NTB, and others to have low NTB. As normality could not be assumed when looking

[8]This method uses bootstrapping; resampling the distribution of the difference in means approaches a normal distribution, allowing for parametric tests to be used.

[9]5000 bootstrap samples were taken; the confidence interval is bias-corrected and accelerated. The P value(s) reported are the likelihood(s) of observing the effect size(s), if the null hypothesis of zero difference is true. For each permutation P value, 5000 reshuffles of the control and test labels were performed.

[10]Following the 5-point Likert scale. 1 indicates the lowest need to belong; 5 indicates the highest need to belong. Reliability analyses revealed that item 1 decreased the internal consistency of the scale in this study. Specifically, the internal consistency across all ten items was α=.576, whereas the internal consistency across the nine items excluding item 1 was α=.640. Given the internal consistency did not substantially change without item 1, we included all the ten items in the analysis.

Condition 1	*Condition 2*	*Unpaired mean difference*	*95% CI*	*p-value*
Control	DI	0.0569	[0.00221, 0.107]	0.044
Control	Priming	-0.0233	[-0.0794, 0.032]	0.417
Control	Priming + DI	0.0098	[-0.048, 0.0662]	0.741
DI	Priming	-0.0802	[-0.133, -0.0222]	0.0056
DI	Priming + DI	-0.0471	[-0.102, 0.0136]	0.114
Priming	Priming + DI	0.0331	[-0.0263, 0.0944]	0.298

Table 1: Pairwise comparisons of unpaired mean differences with 95% Confidence Interval and p-value of the 2-sided non-parametric permutation t-test

	Adjusted p-value for NTB level based on		
	Median	*Mean*	*75th percentile*
Control	0.894	0.648	0.895
DI	0.426	0.245	0.086
Priming	0.722	1.0	0.603
Priming + DI	0.519	0.423	0.780

Table 2: Adjusted p-values of Kruskal-Wallis comparison of high versus low NTB participants per condition and with different ways to determine NTB level

Comparison	*Z*	*P.unadj*	*P.adj*
European - North-American	2.778	**0.005****	**0.016***
European - South and South-East Asian	2.424	**0.015***	**0.046***
North-American - South and South-East Asian	-0.250	0.802	1.000

Table 3: Kruskal-Wallis: p= 0.005. Comparison of Country of origin and diversity choice

at the data cumulatively, we conducted Kruskal-Wallis tests comparing participants with a low NTB to participants with a high NTB for each condition. None were significant (Table 2), meaning that **for none of the conditions, the NTB level affected team diversity**.

4.2 Post-hoc analysis

The results do not support the hypotheses. They even show some opposite results (diversity scores enhance diversity choices). We deemed it insightful to provide an extensive posthoc analysis for mainly two reasons: 1) to validate the data by confirming different expected behavior, 2) to see what other factors do affect the choice of team members.

4.2.1 Regions of origin perceive team diversity differently. As the participant pool had highly different demographic backgrounds, we looked at significant differences in team diversity between participants from different regions of origin. Europeans, South-Asians, and North-Americans were the majority populations within the participant pool (95%). Those participants were part of the analysis. The South-Asian and North-American groups were normally distributed, but the European group was not (Shapiro-Wilk test for European group: p= 0.017). We therefore opted for the non-parametric Kruskal-Wallis test (Table 3). We compared three regions, with team diversity as the dependent variable. The test showed that the **European participants significantly chose more diverse team members than North-Americans and South-Asians**. As different regions of origin yielded significant differences in team diversity, we examined the effect of the treatment conditions per region. As the European participant pool yielded significantly different results we assessed them separately from the North-American and South-Asian participants. We conducted a two-way ANOVA test including only the European participant population. The assumption of normal residuals (Shapiro-Wilk model residuals: p= 0.418) and homogeneity of the variances (Levene's test: p= 0.074) were met. The two-way ANOVA showed that European participants were positively affected by the display of diversity info (p=0.013). **Europeans are therefore more likely to choose diverse teammates, and are also more positively affected by diversity info than other participants** [11] Conducting a two-way ANOVA including all non-European participants furthermore showed that Priming may actually negatively impact the choice of more diverse team members (p=.011).

4.2.2 Functional background matters when choosing teammates. While the task of creating a coffee slogan for a company does not necessarily require any formal education, we did expect to possibly see a preference for teammates with a sales and marketing background[12]. We compared the means of selected team members with a sales background and without one. Due to assumed normality and homogeneity of variances (Shapiro-Wilk: sales p= 0.536, not-sales p= .099; F-test: p=.919) we conducted an unpaired one-tailed t-test which showed that, indeed, **dummies with a sales background were significantly more frequently selected than those without one** (t-test: p= 1.57e-4).

4.2.3 Same gender matters when choosing teammates. One of the stronger homophilic tendencies is that of gender. The expected behavior is that participants choose significantly more team members of the same gender. Due to the unequal distribution of genders among the dummies, we normalized the scores of same-gender team members and different-gender team members of each participant. We ran a two-sample paired Wilcoxon test (due to non-normality, Shapiro-Wilk: p= 3.263e-11) and found a significant difference between selected same-gender team members and selected different-gender team members (upper-tailed Wilcoxon: p=0.002). We repeated the Wilcoxon test for female participants (n= 44) and male participants (n= 76). Female participants similarly selected significantly more same-gender team members (upper-tailed Wilcoxon:

[11]T.B.N. Diversity is calculated as an aggregate measure of all profiling attributes, not only region.
[12]The different functional backgrounds were equally distributed among the dummy profiles.

Coefficients	*Estimate*	*Std. Error*	*t-value*	*p-value*
(Intercept)	13.73103	1.27922	10.734	**1.97e-11** ***
Ranking	-0.36974	0.07206	-5.131	**1.94e-05** ***

Table 4: Linear regression order of appearance (Shapiro-Wilk: p-value = 0.1482).

p=0.011). Male participants followed the same pattern (upper-tailed Wilcoxon: p=0.044). These results indicate that **gender homophily is present, regardless of the condition.**

4.2.4 Order of appearance matters when choosing teammates. We examined whether the order of appearance of the dummy profiles influenced the choice of team members. We expected a linear relationship and therefore conducted a linear regression, where the x-axis represented the order of appearance (places 1-30), and the y-axis the selection frequency of the dummies. We assumed normality (Shapiro-Wilk: p= 0.148). We found that the order of appearance of the dummies showed a strong correlation with the selection of team members (p= 1.94e-5). So, **the order of appearance[13] played a role in the choice of team members from a selection of 30 dummies** (Table 4).

5 DISCUSSION

In this study, we have examined how priming and displaying personalized diversity info affects choosing more diverse team members among crowd users in an open collaboration context. More specifically, we examined whether priming and displaying diversity information together increased crowd team diversity. We conclude that: a) Diversity Info (DI) alone positively affects the choice of more diverse crowd teams; b) Priming alone negatively impacts crowd team diversity when compared with the display of DI; c) combining Priming and Diversity Info (Priming + DI) does not effectively nudge diverse choices of crowd teammates. DI was expected to decrease team diversity [20], yet results indicate that there was no significant drop as participants selected more diverse teammates, especially in the diversity score condition. Future work will be needed to disentangle the effect of each DI intervention (the progress bar and the profile recommendations). Priming – expected to increase team diversity – yielded no significant positive effect. Results even indicate the opposite effect may occur, especially in terms of homophilic preference of teammates of the same region. There are several possible causes for our diverging results. Comparing our results concerning diversity info with the ones from Gómez-Zará et al. [20] we detect study design differences that may have contributed to differences in outcome. In particular, their sample differed significantly from that used in this study. Participants from Gómez-Zará et al. [20] were fewer (N=46, of which the greatest part was American) all volunteering and non-paid undergraduate students while we hired a diverse pool of crowd workers (N=120 from three continents) compensated and motivated through a CIP competition, more in line with a real world setting. Finally, we suggest repeating the study with other scenarios such as political writing and analytical problems to evaluate the effects of different task types.

5.1 System design recommendations

Our first recommendation is to *design platforms that openly explain diversity instead of subliminally.* Direct diversity interventions like recommendations of diverse teammates and diversity scores are more effective at nudging toward diversity choices than suggestive and indirect means. Our use of priming interventions might have been unnoticed by users as it was seamlessly integrated with the rest of the task description. Using UI elements distinctly and concisely can be more effective at nudging than more covert digital priming techniques. Combining different kinds of nudging techniques does not seem to yield predictable outcomes. It can even risk confounding information by incorporating conflicting perspectives as diversity nudges carry implicit assumptions. While counter stereotypes and AIM trigger associations to cultural identity and social belonging, combined with other nudging techniques they could activate undesirable reactions to diverse choices. Based on our results, we suggest testing before combining diversity nudging techniques into a single system. Our third recommendation is to avoid overly-generalized inclusive statements and references and focus on *designing a personalized workspace adjusting to characteristics and task objectives.* We noted through our work that personalized recommendations and diversity scores were more effective than priming on targeted traits. Finally, we link the clarity of nudging interventions with the success of explainable AI. This user-centered experience of AI is proving successful at building trust in digital services [57] as it makes algorithmic engagement accessible. Accordingly, diversity interventions that are interpretable can help grow trust in a system.

5.2 Limitations

One limitation of this study is the choice of participants due to different recruitment platforms used, followed by the limited number (*n=120*) determined by strict quality-control procedures. [14] Another limitation is the design of the task and the profiles. Although crowd users thought that the teammates and the outsourcing company were real[15] (which should have made them pick teammates carefully), we did not offer them tangible proof of the validity of this exercise. Other aspects of the teammates' profiles were not taken into account such as the perceived attractiveness of their photos. We limited the number of teammates to a small pool (*n=30*) which, in some cases, would have been larger in real-world settings. It also limited the available combinations of profiles possible with the given attributes. We consider this a limitation that can be addressed in future studies using real-life tasks requiring actual collaborators. Furthermore, this study gathers findings on the effects of displaying

[13]The dummy profiles were in a random order for each participant. No specific dummy profile had an unfair advantage to be selected based on this order in the regression analysis.

[14]Despite the presence of manipulation checks safeguarding the reliability of the results, it was not entirely possible to monitor crowd users' intents and levels of engagement in the task. One of the risks of relying on remote crowd workers is that we could not test whether participants were making decisions entirely based on the given information or primarily focused on making quick decisions to optimize their time on task. Importantly, the order of appearance mattered, demonstrating that users were prone to select early options without necessarily comparing other profiles.

[15]Validated by crowd participants' feedback that they thought profiles and tasks were genuine.

diversity info and priming in a crowdsourcing setting. Nonetheless, we recognize that these specific implementations (all-inclusive multiculturalism, counter-stereotypes, diversity scores with a bar and color) do not cover all possible applications of priming and displaying DI. There may be different interventions that may yield different results. Active learning interventions such as discussion and problem solving instead of priming techniques could be part of alternative studies since priming risks being short-lived and less resilient to arbitrary factors [48]. Active learning would also be more manageable to evaluate than subtle nudging techniques such as the type of priming used in this study. It would place users at the centre of diversity-enhancement, giving them greater responsibility and agency than more subliminal methods of nudging diversity. Moreover, learning activities could engage crowd users in discussions which are currently absent in our digital intervention.

6 CONCLUSION

This study looked at ways that digital nudging improves diversity in open collaboration crowd teams. It shows that designing diversity-enhancing interfaces, particularly for practical implementations, is greatly context and user-target-dependent. We find that displaying diversity info, as a form of nudging, surprisingly enhances diverse choices among remote users hired for a crowdsourced innovation project. On the contrary, priming strengthens homophilic biases toward users' profiles from the same region. Results from testing diversity priming techniques (AIM and counter-stereotypes) even hint at possible adverse effects on the diverse choice of crowd users. Overall, we also observe homophilic tendencies toward the same gender among crowd users choosing teammates online. The need to belong measure did not yield (expected) significant results in moderating diverse choices, while user interfaces, particularly the profile presentation order, influences participants. Online team formation systems for crowd collaboration have many opportunities to alter their users' perception and decision-making processes. Yet, certain types of nudges may trigger adverse reactions towards diversity. Based on our results, displaying personalized diversity information seems most promising.

REFERENCES

[1] Brenda J Allen and Kavita Garg. 2016. Diversity matters in academic radiology: acknowledging and addressing unconscious bias. *Journal of the American College of Radiology* 13, 12 (2016), 1426–1432.

[2] Roy F Baumeister and Mark R Leary. 1995. The need to belong: desire for interpersonal attachments as a fundamental human motivation. *Psychological bulletin* 117, 3 (1995), 497.

[3] Irene V Blair, Jennifer E Ma, and Alison P Lenton. 2001. Imagining stereotypes away: the moderation of implicit stereotypes through mental imagery. *Journal of personality and social psychology* 81, 5 (2001), 828.

[4] Peter Michael Blau. 1977. *Inequality and heterogeneity: A primitive theory of social structure.* Vol. 7. Free Press New York.

[5] 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.* 1855–1864.

[6] Xiao Chen and Kaiqi Xiong. 2017. A payment scheme in crowdsourcing. In *2017 IEEE International Conference on Communications (ICC).* IEEE, 1–6.

[7] Derrick L Cogburn. 2004. Diversity matters, even at a distance: Evaluating the impact of computer-mediated communication on civil society participation in the World Summit on the Information Society. *Information Technologies & International Development* 1, 3 (2004), pp–15.

[8] Corey Columb and E Ashby Plant. 2011. Revisiting the Obama effect: Exposure to Obama reduces implicit prejudice. *Journal of Experimental Social Psychology* 47, 2 (2011), 499–501.

[9] Richard J Crisp, Laura M Bache, and Angela T Maitner. 2009. Dynamics of social comparison in counter-stereotypic domains: Stereotype boost, not stereotype threat, for women engineering majors. *Social Influence* 4, 3 (2009), 171–184.

[10] Nilanjana Dasgupta. 2004. Implicit ingroup favoritism, outgroup favoritism, and their behavioral manifestations. *Social justice research* 17, 2 (2004), 143–169.

[11] Nilanjana Dasgupta and Shaki Asgari. 2004. Seeing is believing: Exposure to counterstereotypic women leaders and its effect on the malleability of automatic gender stereotyping. *Journal of experimental social psychology* 40, 5 (2004), 642–658.

[12] Nilanjana Dasgupta and Anthony G Greenwald. 2001. On the malleability of automatic attitudes: combating automatic prejudice with images of admired and disliked individuals. *Journal of personality and social psychology* 81, 5 (2001), 800.

[13] Djellel Difallah, Elena Filatova, and Panos Ipeirotis. 2018. Demographics and dynamics of mechanical turk workers. In *Proceedings of the eleventh ACM international conference on web search and data mining.* 135–143.

[14] Indika Dissanayake, Sridhar P Nerur, and Jie Zhang. 2019. Team Formation and Performance in Online Crowdsourcing Competitions: The Role of Homophily and Diversity in Solver Characteristics.. In *ICIS.*

[15] Idoya Ferrero-Ferrero, M Ángeles Fernández-Izquierdo, and M Jesús Muñoz-Torres. 2015. Age diversity: An empirical study in the board of directors. *Cybernetics and Systems* 46, 3-4 (2015), 249–270.

[16] Shandra Forrest-Bank and Jeffrey M Jenson. 2015. Differences in experiences of racial and ethnic microaggression among Asian, Latino/Hispanic, Black, and White young adults. *J. Soc. & Soc. Welfare* 42 (2015), 141.

[17] Lorie A Fridell. 2016. Racial aspects of police shootings: Reducing both bias and counter bias. *Criminology & Pub. Pol'y* 15 (2016), 481.

[18] Bertram Gawronski, Roland Deutsch, Sawsan Mbirkou, Beate Seibt, and Fritz Strack. 2008. When "just say no" is not enough: Affirmation versus negation training and the reduction of automatic stereotype activation. *Journal of experimental social psychology* 44, 2 (2008), 370–377.

[19] Małgorzata A Gocłowska, Richard J Crisp, and Kirsty Labuschagne. 2013. Can counter-stereotypes boost flexible thinking? *Group Processes & Intergroup Relations* 16, 2 (2013), 217–231.

[20] Diego Gómez-Zará, Mengzi Guo, Leslie A DeChurch, and Noshir Contractor. 2020. The Impact of Displaying Diversity Information on the Formation of Self-assembling Teams. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems.* 1–15.

[21] Jessica J Good, Julie A Woodzicka, and Lylan C Wingfield. 2010. The effects of gender stereotypic and counter-stereotypic textbook images on science performance. *The Journal of social psychology* 150, 2 (2010), 132–147.

[22] Mary L Gray, Siddharth Suri, Syed Shoaib Ali, and Deepti Kulkarni. 2016. The crowd is a collaborative network. In *Proceedings of the 19th ACM conference on computer-supported cooperative work & social computing.* 134–147.

[23] Yves RF Guillaume, Jeremy F Dawson, Lilian Otaye-Ebede, Stephen A Woods, and Michael A West. 2017. Harnessing demographic differences in organizations: What moderates the effects of workplace diversity? *Journal of Organizational Behavior* 38, 2 (2017), 276–303.

[24] David A Harrison and Katherine J Klein. 2007. What's the difference? Diversity constructs as separation, variety, or disparity in organizations. *Academy of management review* 32, 4 (2007), 1199–1228.

[25] 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. *Academy of management journal* 41, 1 (1998), 96–107.

[26] Reid Hastie, Colin Schroeder, and Renée Weber. 1990. Creating complex social conjunction categories from simple categories. *Bulletin of the Psychonomic Society* 28, 3 (1990), 242–247.

[27] Natali Helberger. 2011. Diversity by design. *Journal of Information Policy* 1 (2011), 441–469.

[28] Joses Ho, Tayfun Tumkaya, Sameer Aryal, Hyungwon Choi, and Adam Claridge-Chang. 2019. Moving beyond P values: data analysis with estimation graphics. *Nature methods* 16, 7 (2019), 565–566.

[29] Vivian Hunt, Dennis Layton, and Sara Prince. 2015. Diversity matters. *McKinsey & Company* 1, 1 (2015), 15–29.

[30] Russell RC Hutter and Richard J Crisp. 2005. The composition of category conjunctions. *Personality and Social Psychology Bulletin* 31, 5 (2005), 647–657.

[31] Indeed. 2020. 25 Most Common Jobs in America. https://www.indeed.com/career-advice/finding-a-job/most-common-jobs-in-america

[32] Wiebren S Jansen, Sabine Otten, and Karen I van der Zee. 2015. Being part of diversity: The effects of an all-inclusive multicultural diversity approach on majority members' perceived inclusion and support for organizational diversity efforts. *Group Processes & Intergroup Relations* 18, 6 (2015), 817–832.

[33] Tero Karras, Samuli Laine, and Timo Aila. 2019. A style-based generator architecture for generative adversarial networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition.* 4401–4410.

[34] Imran Khan, Ismail Khan, and Ismail Senturk. 2019. Board diversity and quality of CSR disclosure: evidence from Pakistan. *Corporate Governance: The International Journal of Business in Society* (2019).

[35] Akiva Kleinerman, Ariel Rosenfeld, and Sarit Kraus. 2018. Providing explanations for recommendations in reciprocal environments. In *Proceedings of the 12th ACM conference on recommender systems*. 22–30.

[36] Bart P Knijnenburg, Martijn C Willemsen, Zeno Gantner, Hakan Soncu, and Chris Newell. 2012. Explaining the user experience of recommender systems. *User Modeling and User-Adapted Interaction* 22, 4 (2012), 441–504.

[37] Ruud Koopmans and Susanne Veit. 2014. Ethnic diversity, trust, and the mediating role of positive and negative interethnic contact: A priming experiment. *Social science research* 47 (2014), 91–107.

[38] Ziva Kunda, Dale T Miller, and Theresa Claire. 1990. Combining social concepts: The role of causal reasoning. *Cognitive Science* 14, 4 (1990), 551–577.

[39] Fu Wen Kuo and Shu Ching Yang. 2017. The moderating effect of the need to belong and classroom composition on belongingness seeking of minority adolescents. *Journal of adolescence* 61 (2017), 22–30.

[40] Karim R Lakhani, Anne-Laure Fayard, Natalia Levina, and Stephanie Healy Pokrywa. 2012. OpenIDEO. *Harvard Business School Technology & Operations Mgt. Unit Case* 612-066 (2012).

[41] Bruce Lansky. 2011. *25,000+ Baby Names*. Hachette UK.

[42] Mark R Leary, Kristine M Kelly, Catherine A Cottrell, and Lisa S Schreindorfer. 2013. Construct validity of the need to belong scale: Mapping the nomological network. *Journal of personality assessment* 95, 6 (2013), 610–624.

[43] Tracy Levett-Jones and Judith Lathlean. 2009. 'Don't rock the boat': Nursing students' experiences of conformity and compliance. *Nurse education today* 29, 3 (2009), 342–349.

[44] Ioanna Lykourentzou, Shannon Wang, Robert E Kraut, and Steven P Dow. 2016. Team dating: A self-organized team formation strategy for collaborative crowdsourcing. In *Proceedings of the 2016 CHI conference extended abstracts on human factors in computing systems*. 1243–1249.

[45] Elizabeth Mannix and Margaret A Neale. 2005. What differences make a difference? The promise and reality of diverse teams in organizations. *Psychological science in the public interest* 6, 2 (2005), 31–55.

[46] Tobias Mirsch, Christiane Lehrer, and Reinhard Jung. 2017. Digital nudging: Altering user behavior in digital environments. *Proceedings der 13. Internationalen Tagung Wirtschaftsinformatik (WI 2017)* (2017), 634–648.

[47] Richard L Moreland. 2010. Are dyads really groups? *Small Group Research* 41, 2 (2010), 251–267.

[48] Robert R Morris, Mira Dontcheva, and Elizabeth M Gerber. 2012. Priming for better performance in microtask crowdsourcing environments. *IEEE Internet Computing* 16, 5 (2012), 13–19.

[49] Eddy S Ng and Greg J Sears. 2020. Walking the talk on diversity: CEO beliefs, moral values, and the implementation of workplace diversity practices. *Journal of Business Ethics* 164, 3 (2020), 437–450.

[50] Peter Nijkamp and Jacques Poot. 2015. Cultural diversity: a matter of measurement. In *The economics of cultural diversity*. Edward Elgar Publishing.

[51] Gabriele Paolacci, Jesse Chandler, and Panagiotis G Ipeirotis. 2010. Running experiments on amazon mechanical turk. *Judgment and Decision making* 5, 5 (2010), 411–419.

[52] Cheong Ha Park, KyoungHee Son, Joon Hyub Lee, and Seok-Hyung Bae. 2013. Crowd vs. crowd: large-scale cooperative design through open team competition. In *Proceedings of the 2013 conference on Computer supported cooperative work*. 1275–1284.

[53] Eyal Peer, Laura Brandimarte, Sonam Samat, and Alessandro Acquisti. 2017. Beyond the Turk: Alternative platforms for crowdsourcing behavioral research. *Journal of Experimental Social Psychology* 70 (2017), 153–163.

[54] C Carl Pegels, Yong I Song, and Baik Yang. 2000. Management heterogeneity, competitive interaction groups, and firm performance. *Strategic Management Journal* 21, 9 (2000), 911–923.

[55] John A Pelesko. 2007. *Self assembly: the science of things that put themselves together*. CRC Press.

[56] Victoria C Plaut, Flannery G Garnett, Laura E Buffardi, and Jeffrey Sanchez-Burks. 2011. "What about me?" Perceptions of exclusion and Whites' reactions to multiculturalism. *Journal of personality and social psychology* 101, 2 (2011), 337.

[57] Arun Rai. 2020. Explainable AI: From black box to glass box. *Journal of the Academy of Marketing Science* 48, 1 (2020), 137–141.

[58] Giulio Rossetti, Luca Pappalardo, Riivo Kikas, Dino Pedreschi, Fosca Giannotti, and Marlon Dumas. 2016. Homophilic network decomposition: a community-centric analysis of online social services. *Social Network Analysis and Mining* 6, 1 (2016), 1–18.

[59] Niloufar Salehi, Andrew McCabe, Melissa Valentine, and Michael Bernstein. 2017. Huddler: Convening stable and familiar crowd teams despite unpredictable availability. In *Proceedings of the 2017 ACM conference on computer supported cooperative work and social computing*. 1700–1713.

[60] Silke L Schneider and Anthony F Heath. 2020. Ethnic and cultural diversity in Europe: validating measures of ethnic and cultural background. *Journal of Ethnic and Migration Studies* 46, 3 (2020), 533–552.

[61] Feng Shi, Misha Teplitskiy, Eamon Duede, and James A Evans. 2019. The wisdom of polarized crowds. *Nature human behaviour* 3, 4 (2019), 329–336.

[62] Antonio Solanas, Rejina M Selvam, José Navarro, and David Leiva. 2012. Some common indices of group diversity: Upper boundaries. *Psychological Reports* 111, 3 (2012), 777–796.

[63] Dominic Spohr. 2017. Fake news and ideological polarization: Filter bubbles and selective exposure on social media. *Business Information Review* 34, 3 (2017), 150–160.

[64] 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 (2006), 389–406.

[65] Flannery G Stevens, Victoria C Plaut, and Jeffrey Sanchez-Burks. 2008. Unlocking the benefits of diversity: All-inclusive multiculturalism and positive organizational change. *The journal of applied behavioral science* 44, 1 (2008), 116–133.

[66] Sherry E Sullivan, Monica L Forret, Shawn M Carraher, and Lisa A Mainiero. 2009. Using the kaleidoscope career model to examine generational differences in work attitudes. *Career Development International* (2009).

[67] Cass R Sunstein. 2015. Nudging and choice architecture: Ethical considerations. *Yale Journal on Regulation, Forthcoming* (2015).

[68] Cass R Sunstein. 2017. A prison of our own design: divided democracy in the age of social media. *Democratic Audit UK* (2017).

[69] Nava Tintarev and Judith Masthoff. 2007. A survey of explanations in recommender systems. In *2007 IEEE 23rd international conference on data engineering workshop*. IEEE, 801–810.

[70] John C Turner and Katherine J Reynolds. 2011. Self-categorization theory. *Handbook of theories in social psychology* 2, 1 (2011), 399–417.

[71] Jean M Twenge, Stacy M Campbell, Brian J Hoffman, and Charles E Lance. 2010. Generational differences in work values: Leisure and extrinsic values increasing, social and intrinsic values decreasing. *Journal of management* 36, 5 (2010), 1117–1142.

[72] Unsplash. 2021. *Unsplash*. https://unsplash.com/

[73] Félice van Nunspeet, Naomi Ellemers, and Belle Derks. 2015. Reducing implicit bias: How moral motivation helps people refrain from making "automatic" prejudiced associations. *Translational Issues in Psychological Science* 1, 4 (2015), 382.

[74] H Jack Walker, Hubert S Feild, Jeremy B Bernerth, and J Bret Becton. 2012. Diversity cues on recruitment websites: Investigating the effects on job seekers' information processing. *Journal of Applied Psychology* 97, 1 (2012), 214.

[75] H Jack Walker, Hubert S Feild, William F Giles, Achilles A Armenakis, and Jeremy B Bernerth. 2009. Displaying employee testimonials on recruitment web sites: effects of communication media, employee race, and job seeker race on organizational attraction and information credibility. *Journal of Applied Psychology* 94, 5 (2009), 1354.

[76] Markus Weinmann, Christoph Schneider, and Jan Vom Brocke. 2016. Digital nudging. *Business & Information Systems Engineering* 58, 6 (2016), 433–436.

[77] Yang Yang and Alison M Konrad. 2011. Understanding diversity management practices: Implications of institutional theory and resource-based theory. *Group & Organization Management* 36, 1 (2011), 6–38.

[78] Khobaib Zaamout. 2020. A Team Composition Approach For Social Crowdsourcing Communities. (2020).