Management Science

The Diversity Heuristic: How Team Demographic Composition Influences Judgments of Team Creativity

Devon Proudfoot, Zachariah Berry, Edward H. Chang, Min B. Kay

The Diversity Heuristic: How Team Demographic Composition Influences Judgments of Team Creativity

Devon Proudfoot,[a,*] Zachariah Berry,[b] Edward H. Chang,[c] Min B. Kay[d]

[a] Human Resource Studies Department, School of Industrial and Labor Relations, Cornell University, Ithaca, New York 14853;
[b] Organizational Behavior Department, ILR School, Cornell University, Ithaca, New York 14853; [c] Negotiation, Organizations, and Markets Unit, Harvard Business School, Harvard University, Boston, Massachusetts 02163; [d] MBA Department of Leadership and People Management, University Canada West, Vancouver, British Columbia V6Z 0E5, Canada
*Corresponding author

Contact: devon.proudfoot@cornell.edu, https://orcid.org/0000-0002-2733-038X (DP); ztb8@cornell.edu (ZB); ehchang@hbs.edu, https://orcid.org/0000-0003-0502-2957 (EHC); min.kay@ucanwest.ca, https://orcid.org/0000-0002-0827-6437 (MBK)

Abstract. Despite mixed evidence for the relationship between demographic diversity and creativity, we propose that observers hold a lay belief that demographic diversity increases creativity and apply this lay belief in judgments about teams and their creative work. Across eight preregistered studies ($n = 5,530$), we find that observers judge teams diverse in terms of race and gender to be more creative than teams homogeneous in terms of race and gender, including in incentive-compatible predictions made about real teams competing in a creativity challenge. We also find that products attributed to demographically diverse teams are evaluated as more creative compared with identical products attributed to demographically homogenous teams. Mediation analyses provide evidence consistent with the notion that people perceive demographic diversity (i.e., social category differences) to be correlated with cognitive diversity (i.e., difference of perspectives), and this belief contributes to attributions of greater creativity to diverse teams and the ideas they generate. We can also turn off the perceived association between demographic diversity and creativity by directly manipulating people's perceptions of team cognitive diversity. Furthermore, we find evidence of a curvilinear relationship between the proportion of racial minorities or women in a group and judgments of the group's creativity. Together, our results suggest that the popular uptake of the belief that diversity boosts creativity may impact how creativity is identified in organizational contexts.

Introduction

Being able to effectively recognize *who* and *what* is creative is an important precursor to innovation (Elsbach and Kramer 2003, Mueller et al. 2014, Berg 2016). Imagine a competition in which teams of entrepreneurs pitch their ideas to investors. Which kind of team is likely to pitch a more creative idea: A team with members representing a variety of different races and ethnicities, or a team with members representing just one race/ethnicity? We propose that in this type of situation, observers will bet on the diverse team. Specifically, drawing from the notion that people hold implicit theories of creativity (Loewenstein and Mueller 2016), we predict that observers use a group of individuals' aggregated degree of demographic diversity as a signal of their collective creative ability, thus attributing more creativity to groups that appear diverse compared with groups that appear homogenous. Given popular uptake of the argument that demographic diversity boosts creativity (Rock and Grant 2016, Schiller 2018, Levine 2020), this prediction may not be surprising; however, if supported, it has important implications for decision-making in organizational contexts.

The belief that diversity benefits creativity is ubiquitous in the business world (Rock and Grant 2016, Ely and Thomas 2020). At the same time, the academic literature suggests that the relationship between demographic diversity and creativity is, in fact, complex and

dependent on many moderating factors (Mannix and Neale 2005, Galinsky et al. 2015, Ellemers and Rink 2016; also see Bell et al. (2011), Wang et al. (2019), and Byron et al. (2023) for meta-analyses). Regardless of whether demographic diversity has the potential to boost team creativity under certain conditions, in our research, we focus on examining whether observers assume that diversity enhances creativity, leading them to generally expect demographically diverse groups to outperform demographically homogenous groups on creative tasks. Furthermore, we test whether observers judge the exact same idea to be more creative when it is attributed to a demographically diverse team than when it is attributed to a demographically homogenous team, thus exploring whether homogenous teams' creative ideas may be overlooked relative to those generated by diverse teams. We test a mechanism underlying our proposed effects—the tendency to associate demographic diversity (i.e., social category differences) with cognitive diversity (i.e., difference of perspectives), given that cognitive diversity is seen as important for creativity (Runco and Jaeger 2012).

Theoretical Background
Social Context of Creativity Evaluation
A burgeoning literature illuminates how social factors shape people's judgments about others' creativity. Studies have shown that assessments of a person's creativity may be shaped by their perceived fit with creative prototypes (Elsbach and Kramer 2003) or readily observable social cues such as gender (Proudfoot et al. 2015). Our research is novel in extending the individual-level focus of previous studies—which explored how the attributes of solo creators influence evaluations of their creativity—to the group level examining how salient group-level characteristics may impact judgments about a group's collective creativity. To do so, we build on theory positing that evaluators hold lay beliefs about creativity that influence what and who they judge to be creative (Sternberg 1985, Runco and Bahleda 1986, Loewenstein and Mueller 2016). According to this perspective, a person or idea may be judged to be creative to the extent that the person or idea fits with observers' existing lay theories about what it means to be creative. For example, Loewenstein and Mueller (2016) explored observers' implicit theories of what makes ideas creative and found that, in Western contexts, observers tend to believe that creative ideas are defined more by novelty than usefulness. We propose that, in addition to holding implicit theories of creative ideas (Loewenstein and Mueller 2016) and creative individuals (Sternberg 1985, Elsbach and Kramer 2003), observers hold implicit theories about creative groups—that is, lay beliefs about the characteristics of groups that make them more or less creative.

Demographic Diversity as a Cue for Group Creativity
We argue that social perceivers may hold an implicit lay theory that demographic diversity—differences among individuals on social category dimensions such as race/ethnicity and gender (Harrison et al. 2002)—enhances group creativity. The actual relationship between demographic diversity and creativity is complex, with studies showing positive, negative, and null relationships (Milliken and Martins 1996, Williams and O'Reilly 1998, Mannix and Neale 2005, Van Knippenberg and Schippers 2007, Ellemers and Rink 2016). Yet popular uptake of the research on the effects of demographic diversity on creativity has tended to emphasize the evidence showing a positive association between group diversity and creative performance (Catalyst 2014, Phillips 2014, Eswaran 2019). Indeed, popular press business articles frequently tout the benefits of race and gender diversity for creativity and innovation in organizations (Rock and Grant 2016, Schiller 2018, Levine 2020). In our own review of the "Diversity and Inclusion" websites of 2019 s Fortune 100 companies, we found that the idea that diversity increases creativity and/or innovation was explicitly mentioned on more than half (62%) of these companies' websites (see online appendix for study details). As such, we expect that in the context of evaluating the creativity of groups, observers may use a group of individuals' aggregated degree of demographic diversity (Phillips et al. 2018) as a signal of their collective creative ability, attributing more creativity to groups that appear to be demographically diverse, compared with groups that appear to be demographically homogeneous.

As the first test of our predictions, we focus specifically on how demographic diversity (versus homogeneity) along the dimensions of race/ethnicity and gender influences judgments of group creativity. Race and gender are highly salient forms of demographic diversity, with observers spontaneously encoding visual cues related to a person's race and gender (Fiske and Neuberg 1990, Ito and Urland 2003). Race and gender are also central to lay conceptions of diversity in the United States, the context of the current studies (Bell and Hartmann 2007, Unzueta and Binning 2010). As such, race- and gender-based differences are likely implicit in observers' lay understanding of the type of demographic diversity that promotes creativity. Hence, we expect that cues of race- and gender-based diversity (versus homogeneity) may be salient and potent signals of a group's creative ability.

Mediating Role of Perceived Cognitive Diversity
Although observers may make many mediating inferences leading them to attribute more creativity to diverse teams than to homogenous teams, as a first test, we examine one assumption on the part of observers contributing to why demographically diverse groups may be judged

more creative compared with demographically homogeneous groups: observers may associate demographic diversity—differences on "surface level" demographic dimensions—with cognitive diversity—differences on more "deep level" dimensions such as personality, skills, knowledge, and perspectives (Harrison et al. 2002). Observers may assume that a group of individuals who look different from each other (e.g., a group with an Asian man, a Black man, and a White man) also think differently from each other, whereas a group of individuals who all look similar (e.g., a group of all Asian men, a group of all Black men, or a group of all White men) are assumed to all think in similar ways. Given that diversity of thought is seen as important for creative idea generation (Kanter 2010, Runco and Jaeger 2012), groups that are perceived as more cognitively diverse should also be perceived as more creative. Indeed, the belief that individuals who "look different" also "think different" is ubiquitous in popular press messaging on the benefits of diversity for creativity, and these articles often argue that creativity and innovation should follow from the unique perspectives and experiences that demographic diversity brings (Catalyst 2014, Phillips 2014, Eswaran 2019).

Academic studies also provide suggestive evidence that observers interpret demographic diversity in a group as a signal of the group's underlying cognitive diversity. For instance, Phillips et al. (2006) found that while completing a hidden profile task, participants in racially diverse groups assumed the experimenter had given individual group members more varied information compared with participants in racially homogenous groups. Although this study examined group members' own judgments, rather than those of outside observers, and specifically measured belief in how information was distributed by the experimenter rather than belief that racial diversity in a group might be a signal of underlying cognitive differences between group members (e.g., differences in perspectives, experiences, and skills), the finding of Phillips et al. (2006) provides some initial hints that the presence of racial diversity may trigger the expectation of other differences between group members. In other relevant research, Daniels et al. (2017) examined whether observers' judgments of one type of diversity can "spill over" and influence judgments of another type of diversity, finding in one study that observers judged an organization with a higher degree of racial diversity to also have a higher degree of computer programming language skill diversity. This finding is also consistent with our prediction that demographic diversity in a group may be seen as evidence of that group's cognitive diversity. We extend the research of Daniels et al. (2017) by examining whether the perceived connection between demographic diversity and cognitive diversity may influence judgments of team creativity.

Research on stereotyping and essentialism also suggests that observers have a generalized tendency to make assumptions about a person's internal attributes based on that person's demographic characteristics. Observers have been shown to use a person's social category memberships, such as their race or gender, to make inferences about their individual dispositions and personality traits (Fiske and Neuberg 1990). Research on positive stereotyping in particular suggests that observers often assume that members of different social categories (e.g., individuals of different races) possess unique skills or virtues (Cuddy et al. 2008, Czopp et al. 2015). Studies of psychological essentialism further show that such stereotypes may be traceable to a belief that surface-level differences between members of different social categories reflect deeper, immutable trait-based differences (Yzerbyt et al. 2001, Bastian and Haslam 2006, Haslam and Whelan 2008). As such, we suggest that observers' proposed tendency to attribute more creativity to demographically diverse groups than demographically homogenous groups may be traceable to the assumption that demographic diversity reflects underlying differences in perspectives, skills, and experiences, which in turn benefit group creativity.

Current Research

We predicted that observers would expect demographically diverse teams to be more creative compared with demographically homogenous teams. In addition, consistent with prior work suggesting that the attributes of individuals can shape judgments about the creativity of their ideas (Elsbach and Kramer 2003), we predicted that output attributed to a team depicted as demographically diverse would be judged as more creative compared with identical output attributed to a team depicted as demographically homogeneous. We also predicted that perceptions of team cognitive diversity would account, in part, for these proposed effects.

We tested our predictions across eight preregistered studies and three supplemental studies. Study 1 showed that when asked to consider how best to organize employees into teams to produce creative ideas, observers judged demographically diverse team configurations to be more effective for creativity than demographically homogenous team configurations (also see Study S1). Study 2 showed that in a real creativity competition in which observers were incentivized for accuracy, observers expected racially diverse teams to outperform racially homogenous teams on a creative task and provided mediation evidence for perceived cognitive diversity as a mechanism underlying this effect. Studies 3–4b showed that products attributed to demographically diverse teams were judged more creative compared with identical products attributed to

Table 1. Participant Demographics for Study 1 to Study 6

	Study 1	Study 2	Study 3	Study 4a	Study 4b	Study 5a	Study 5b	Study 6
Gender								
Woman	68%	51%	51%	48%	53%	51%	52%	43%
Man	31%	49%	49%	50%	45%	49%	47%	57%
Other identity	1%			2%	1%		1%	
Age								
M_{age}	23.32	39.73	40.51	42.29	41.23	45.96	46.23	41.64
SD_{age}	6.93	11.95	12.88	13.39	12.94	16.70	16.67	19.54
Race/ethnicity								
White	40%	78%	77%	80%	73%	72%	75%	73%
Asian	47%	12%	10%	8%	10%	5%	5%	10%
Black	5%	8%	9%	8%	10%	13%	11%	10%
Hispanic	11%	6%	7%	5%	7%	10%	10%	7%
Other	3%	1%	1%	2%	2%	2%	3%	2%

demographically homogenous teams and offered moderation evidence consistent with our mechanistic account (also see Studies S2 and S3). Studies 5a and 5b explored whether there is a threshold above which observers judge a group to have enough diversity to produce creative ideas. We found an inverted U-shaped relationship between the proportion of racial minorities (or women) in a group and belief that the group had enough diversity to produce creative ideas. Finally, Study 6 explored our predicted effect in the context of selection decisions, showing that when a team's task requires creativity, evaluators show a stronger tendency to select new members to add to the team who increase the team's racial diversity.

We report all variables, manipulations, and measures included in each study. For all studies, study design, sample size, predictions, exclusion criteria, and analysis plans were preregistered at AsPredicted.org. Data, materials, code, and preregistrations for all studies are available at https://osf.io/eymn7/?view_only=cc0a2274dc3b4a25b7f38c38bf99b1e3. For all studies, analyses were conducted using SPSS. All studies received ethics approval from Cornell University's institutional review board. Complete details for all supplemental studies are reported in the online appendix.

Study 1

As an initial test of our prediction that observers expect diverse teams to outperform homogenous teams on creative tasks, we asked study participants to consider how best to arrange a group of employees into teams. Specifically, we compared participants' evaluation of the creativity of teams configured to be homogenous in terms of both race and gender to participants' evaluation of the creativity of teams configured to be diverse in terms of race and gender. We also included two additional team configuration conditions: a condition in which teams were configured to be diverse in terms of race but not gender and a condition in which teams

were configured to be diverse in terms of gender but not race to explore potential additive effects of different dimensions of demographic diversity on perceived creativity. We predicted that observers would judge teams configured to be diverse in terms of both race and gender to be more creative compared with teams configured to be homogenous in terms of race and gender. We did not have specific a priori predictions for how participants would judge the creativity of teams that were diverse in terms of race only or gender only.
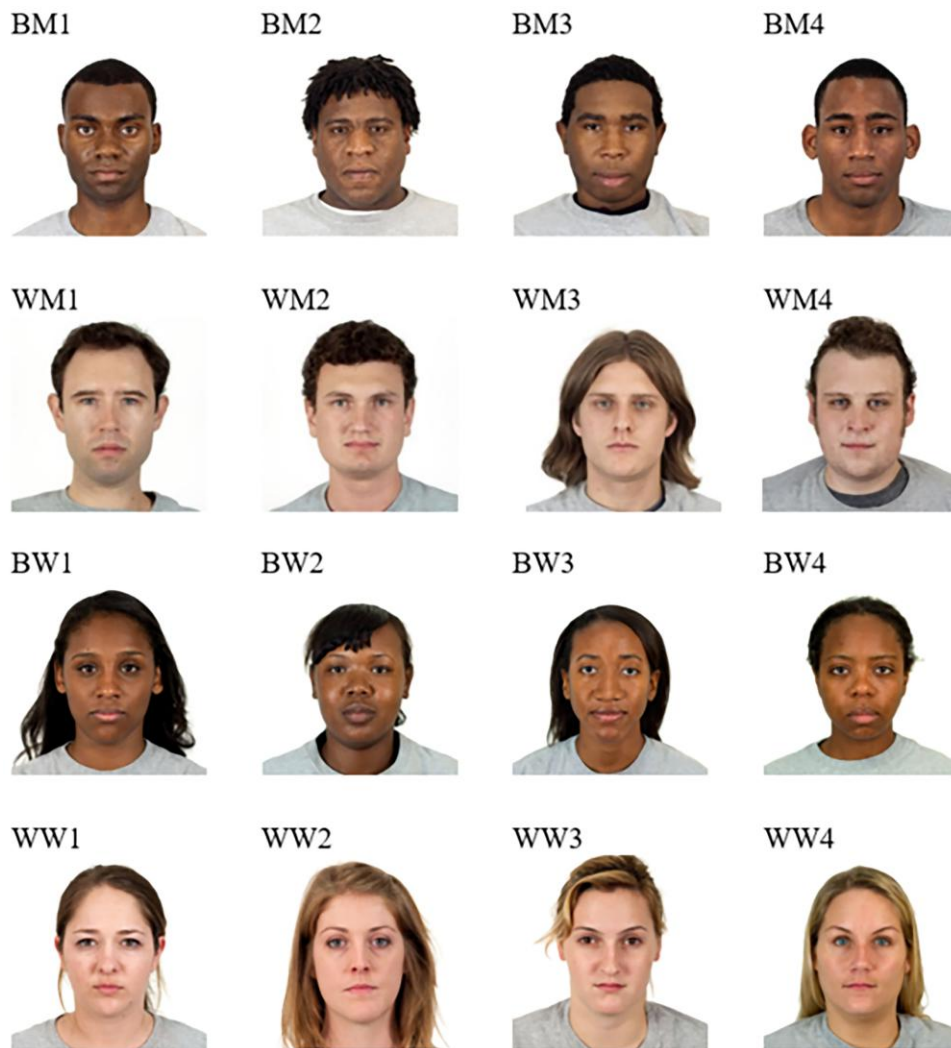
Participants

We recruited participants through the behavioral laboratory of a business school at a large private university in the Northeastern United States. The pool consisted of students, university staff, and community members. Participants completed the survey in exchange for course credit or the chance to win a $20 Amazon gift card. Sample size was determined by how many respondents completed the survey within two weeks after the survey was sent out. We recruited 517 participants. No participants were excluded from analyses. This study was preregistered at Aspredicted.org (#93579). Participant demographics for Study 1 to Study 6 are presented in Table 1.

Materials and Procedure

First, participants read the following passage:

> "In this survey, we are interested in how people evaluate managers' decisions. A manager has 16 employees. Half are women and half are men. Half are Black and half are White. The manager must organize the employees into four teams, with four employees in each team. Each team will work together to generate creative ideas. The manager's goal is to create teams that will be effective at generating creative ideas together. On the next page, you will see how the manager organized the employees into teams. You will be asked to consider whether you think the way the manager sorted employees into teams will be effective for generating creative ideas."

Figure 1. (Color online) Face Stimuli (Study 1)

Participants were then randomly assigned to one of four conditions across which we varied how the employees were grouped together into teams. All participants then saw the exact same 16 faces (4 Black male faces, 4 White male faces, 4 Black female faces, and 4 White female faces). Faces were headshots taken from the Chicago Face Database (Ma et al. 2015). Using existing ratings of the faces provided by the database, we chose faces that had a neutral expression and had been rated as between the ages of 25 and 35 and as relatively unambiguous in terms of their sex and race. See Figure 1 for face stimuli used.

Participants in the first condition saw the faces organized into teams that were homogenous in terms of both race and gender (i.e., a team of four Black men, a team of four White men, a team of four Black women, and a team of four White women). Participants in the second condition saw the faces organized into teams that were diverse in terms of gender but not race (i.e., two teams each composed on two Black men and two Black women, two teams each composed of two White men and two White women). Participants in the third condition saw the faces organized into teams that were diverse in terms of race but not gender (i.e., two teams each composed of two Black men and two White men, two teams composed of two Black women and two White women). Participants in the fourth condition saw the faces organized teams that were diverse in terms of both race and gender (i.e., four teams each composed of one Black man, one White man, one Black woman, and one White woman). After participants viewed the team configuration, they were asked, "How effective do you think this configuration is for generating creative ideas?" rated on a seven-point scale (1 = not at all effective for generating creative ideas, 7 = extremely effective for generating creative ideas).

Results and Discussion

There was a main effect of condition on judgments of how effective the team configuration was for creativity ($F(3, 513) = 158.08$, $p < 0.001$; $\eta^2 = 0.480$; Figure 2). As predicted, teams that were configured to be diverse in terms of both race and gender were judged more effective for generating creative ideas (mean (M) = 5.38, standard deviation (SD) = 1.36) compared with teams configured to be homogenous in terms of both race and gender ($M = 2.24$, $SD = 1.21$; $t(513) = -20.75$, $p < 0.001$, $d = 2.58$). Teams configured to be diverse in terms of both race and gender were also judged more effective for generating creative ideas compared with teams diverse in terms of gender only ($M = 2.94$, $SD = 1.13$; $t(513) = -16.07$, $p < 0.001$, $d = 2.01$) and teams diverse in terms of race only ($M = 3.43$, $SD = 1.17$; $t(513) = -12.87$, $p < 0.001$, $d = 1.60$), indicating additive effects of introducing multiple types of diversity on judgments of team creativity.

Teams configured to be diverse in terms of gender only and teams configured to be diverse in terms of race only were judged more effective for generating creative ideas compared with teams configured to be homogenous in terms of race and gender (gender diverse versus homogenous: $t(513) = -4.61$, $p < 0.001$, $d = 0.57$; racially diverse versus homogenous: $t(513) = -7.89$, $p < 0.001$, $d = 0.98$). Interestingly, teams configured to be diverse in terms of race only were judged more effective for creative idea generation compared with teams configured to be diverse in terms of gender only ($t(513) = -3.25$, $p = 0.007$, $d = 0.41$). We replicated Study 1's pattern of effects in a separate sample that more closely approximated the general U.S. population, recruited from CloudResearch ($n = 400$, see Study S1 in the online appendix).

Study 2

Study 1 provided evidence that observers perceive demographically diverse teams to be more effective for creative idea generation than demographically homogenous teams

in a hypothetical scenario. Study 2 tested whether our predicted effect extended to incentive-compatible judgments made about real teams. We assembled teams who competed in a team-based creativity challenge. The teams included three racially homogenous teams (an all-Asian team, and all-Black team, an all-White team) and a racially diverse team made up of an Asian member, a Black member, and a White member. Study participants then read about the creativity challenge and viewed one of the four teams who competed in the challenge. Participants were given a $0.25 bonus and had the option to bet any amount of this bonus on this team winning the competition. We predicted that participants given the option to bet on the racially diverse team would bet more of their bonus on the team winning the creativity competition compared with participants given the option to bet on the homogenous teams. We also tested whether the belief that racial diversity gives rise to cognitive diversity may be a mechanism underlying this effect.

Method

Participants. We requested 800 participants from CloudResearch (Litman et al. 2017). We did not know exactly what effect size to expect, thus we aimed to maximize power by recruiting 800 participants (400 participants per condition), which provided adequate power to detect a small- to medium-sized effect. U.S. residents with a >95% approval rating on CloudResearch were invited to participate. A total of 801 participants completed the survey, in exchange for $0.40 each, with the possibility of an additional $0.25 bonus. No participants were excluded from analyses. This study was preregistered at Aspredicted. org (#64817).

Materials and Procedure

Creation of Team Stimuli. Prior to conducting the main study, we recruited male U.S. residents via Prolific

Figure 2. Effect of Team Configuration Condition on Perceived Effectiveness for Creativity (Study 1)
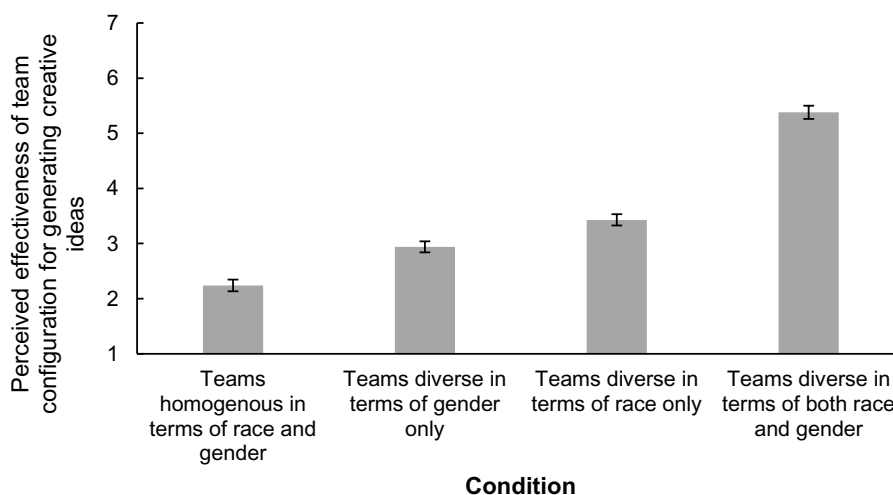
Figure 3. (Color online) Team Photos (Study 2)

Academic (Peer et al. 2017) to participate in a team-based creativity challenge. Specifically, using the demographics filters provided by Prolific, we recruited four three-person teams: one consisting of three East Asian members, one consisting of three Black members, one consisting of three White members, and one diverse team consisting of one East Asian member, one Black member, and one White member. All team members confirmed their race and gender prior to completing the creativity challenge and were asked to select a headshot to represent themselves in the study from options presented by the researchers. Each team then completed the alternative uses task (Guilford 1967) via an online chat platform. Specifically, each team was given five minutes to generate different uses for a newspaper. The four teams (represented by team members' selected headshots) are shown in Figure 3.

Next, two research assistants for whom the demographic composition of the teams was masked independently read through the four teams' chat transcripts. Following the measure of creative performance of Silvia et al. (2009), the research assistants were instructed to consider a set of ideas to be highly creative to the extent that they perceived them as original and useful and believed that only a few teams would come up with them. The research assistants then used a scale of one to seven to score each team's set of ideas, with one indicating "not creative at all" and seven indicating "extremely creative." The two research assistants' ratings were highly reliable (intraclass correlation coefficient (ICC) = 0.962) and thus were averaged to create a creativity score for each team.

Main Study. Participants read that the purpose of the study was to examine how accurate people were at guessing how creative different teams will be. Participants then read the following information: "A few weeks ago, we conducted an online study where we recruited participants to take part in a team-based creativity challenge. Each team was made up of three male U.S. residents. Each participant was asked to select a photo to represent themselves for the creativity challenge. Each team entered a chatroom together and had five minutes to complete the creativity challenge. Each of the teams chatted for five minutes about different ways to use a newspaper. We then scored each team's uses in terms of how creative they were."

Participants then read that they would see one of the four teams that competed in the creativity challenge and would have a chance to bet on whether the team won the creativity challenge. Participants read that they would be allocated a bonus of $0.25 to bet and that they could bet any amount of this bonus on the team. Participants further read that if the team actually won the creativity challenge, whatever amount they bet would be doubled; however, if the team did not win the creativity challenge, whatever amount they bet would be forfeited. Participants were then given examples of the outcomes associated with betting different amounts and completed two comprehension questions about the betting scheme which they had to answer correctly before proceeding.

Participants were then randomly assigned to view a photo of either the diverse team from the creativity challenge or one of the three homogenous teams from the challenge. Participants read, "We would like you to bet on whether this team came up with the most creative set of ideas in the creativity competition—that is, the more original and useful ideas that few teams would come up with. How many cents of your 25-cent bonus would you like to bet that this team won the creativity competition? Below, type in any number ranging from 0 to 25 to indicate how many cents of your 25-cent bonus you would like to bet." After participants

Figure 4. Mediational Model (Study 2)
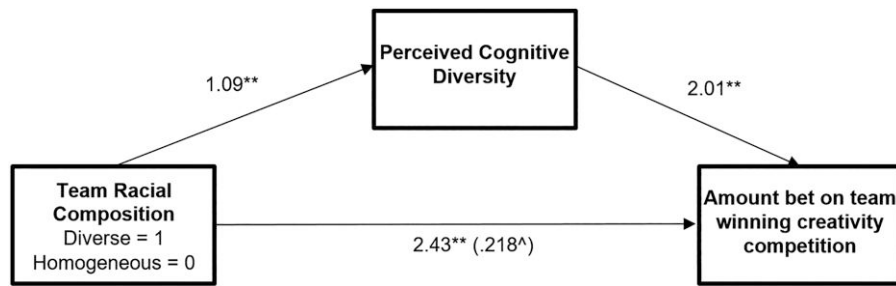
Notes. This figure shows the indirect effect of team composition condition (diverse versus homogeneous) on amount bet on team winning creativity competition through perceived cognitive diversity of the team. Regression coefficients are unstandardized. $**p < 0.001, \hat{}p = 0.742$.

made their bet, they were shown a photo of the team again and were asked to rate the team on perceived cognitive diversity using a scale adapted from the measure of Van der Vegt and Janssen (2003): "Please rate the extent to which the members of this team: (i) differ in their way of thinking, (ii) differ in their knowledge and skills, (iii) differ in how they view the world, and (iv) differ in their beliefs about what is right and what is wrong." These items were rated on a seven-point scale (1 = Very small extent, 7 = Very large extent) and were averaged to form a perceived cognitive diversity composite ($\alpha = 0.88$).

Results and Discussion
As predicted, participants who saw a diverse team bet more of their bonus (measured in cents) on the team winning the creativity challenge ($M = 14.27$, $SD = 8.69$) compared with participants who saw a homogenous team ($M = 11.85$, $SD = 9.10$; $t(799) = 3.86$, $p < 0.001$, $d = 0.27$).[1] Using the PROCESS macro for SPSS of Hayes (2017), there was a significant indirect effect from team composition (diverse versus homogenous) to amount bet through perceptions of team cognitive diversity ($b = 2.18$, standard error (SE) = 0.31, 95% confidence interval [1.58, 2.80]), suggesting that the belief that the racially diverse team was more cognitively diverse compared with the racially homogenous teams may help explain participants' tendency to bet more of their bonus on the diverse team winning the creativity competition than the homogenous teams (Figure 4).

Comparing bets by homogenous team race, we found significant effects for comparisons involving the diverse team versus the all-Black team and the diverse

team versus the all-White team and a marginally significant effect for the diverse team versus the all-Asian team (Table 2).

Study 3
Study 2 showed that in an incentive-compatible paradigm in which participants judged real teams, observers predicted that a racially diverse team would be more creative than racially homogeneous teams. We also provided mediation evidence that perceived cognitive diversity may be one mechanism underlying this effect. In Study 3, we tested whether participants attribute more creativity to diverse teams than homogenous teams even when the team's output is held constant. We predicted that a product attributed to a racially diverse team would be judged more creative compared with an identical product attributed to a racially homogenous team. We used a larger sample of team photo stimuli than were used in Studies 1 and 2 and rotated the individual face photos across teams to rule out the possibility that any effects we observed were due to the specific faces that composed the diverse and homogenous teams.

In addition, although Study 2 showed our predicted effect in a fully between-participants design in which participants only viewed one team in isolation, we were interested in directly examining whether showing participants one team in isolation versus presenting multiple teams together moderated our predicted effect in the context of product evaluation, given that a team's degree of demographic diversity would likely be more salient in a joint evaluation context compared with a separate evaluation context. Thus, Study 3 varied both

Table 2. Bets (in Cents) by Homogenous Team Race (Study 2)

Homogenous team race	$M_{diverse\ team}$	$M_{homogeneous\ team}$	df	t	p	Cohen's d
All-Asian team	14.27	12.81	517	1.66	0.098	0.17
All-Black team	14.27	12.19	524	2.37	0.018	0.23
All-White team	14.27	10.64	530	4.27	<0.001	0.42

Figure 5. (Color online) Face Stimuli (Study 3)
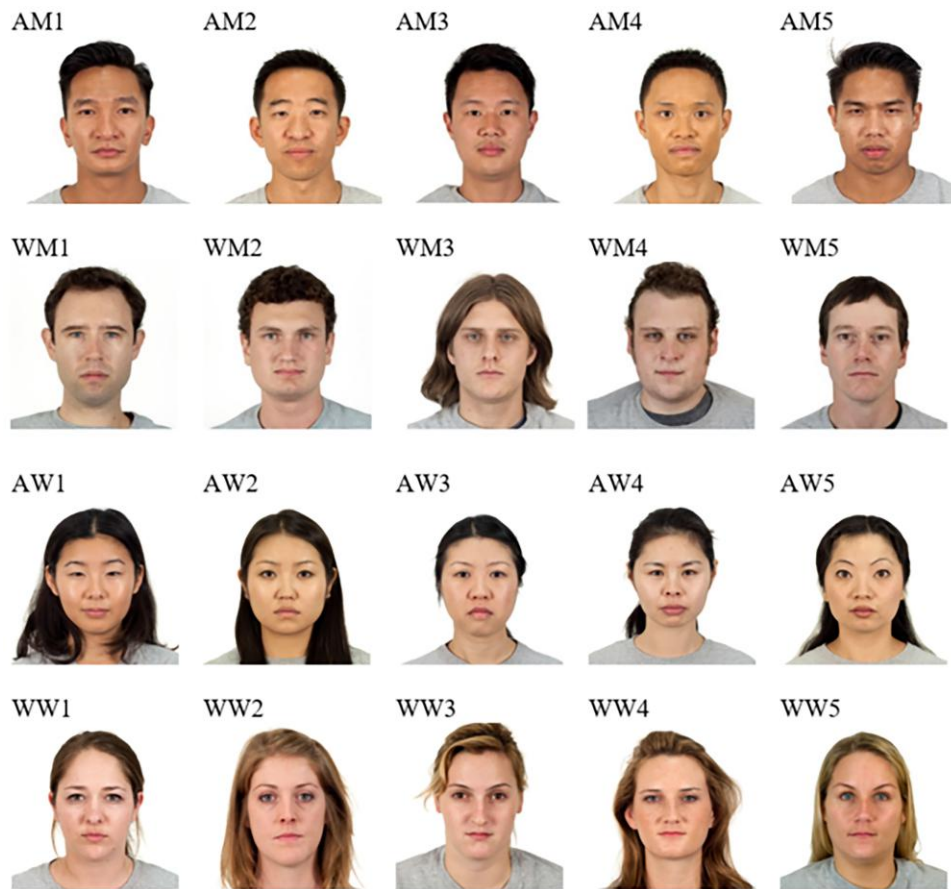
whether participants evaluated a diverse team or a homogenous team and whether this team was presented jointly with other teams prior to evaluating that team's product.

Method

Participants. We requested 800 participants from CloudResearch. We did not know exactly what effect size to expect, thus we aimed to maximize power by recruiting 800 participants (200 participants per condition), which provided adequate power to detect a small-to-medium-sized effect. U.S. residents with a >95% approval rating on CloudResearch were invited to participate. A total of 798 participants completed the survey, in exchange for $0.40 each. No participants were excluded from analyses. This study was preregistered at Aspredicted.org (#96608).

Materials and Procedure. All participants read about five teams who ostensibly competed in a creativity competition in which each team submitted an idea for a company logo. Participants were then randomly assigned to either a *joint evaluation* condition or a *separate evaluation* condition. Participants in the *joint evaluation* condition saw photos of five teams. We varied the

teams in terms of both the race and gender of team members. Specifically, all participants saw an all-Asian male team, an all-White male team, an all-Asian female team and an all-White female team. Participants also saw a diverse team consisting of one Asian man, one White man, one Asian woman, and one White woman. Team photos were created using headshots taken from the Chicago Face Database (Ma et al. 2015). From this database, we selected 20 faces (5 White women, 5 White men, 5 Asian women, 5 Asian men). All faces had a neutral expression and had previously been rated as between the ages of 25 and 35 and as relatively unambiguous in terms of their sex and race. All participants saw the exact same 20 faces. Across participants, we varied how the faces were organized into the four homogenous teams and one diverse team. As our stimuli consisted of five faces that were the same in terms of both race and gender, each participant was shown four members of each race/gender group together in a homogenous team and the remaining member of that group in the diverse team. The specific member of each group shown in the diverse team was rotated across participants. Figure 5 shows face stimuli used in Study 3.

After participants viewed the five teams together, they were randomly assigned to either evaluate the

Figure 6. Logo Stimuli (Study 3)

logo submitted by the diverse team or the logo submitted by one of the four homogenous teams. Before participants viewed the team's logo, they read some background information about the focal team, specifically, where the focal team was based, how long the team had been working together, how many clients they had, how many projects they had completed prior to evaluating their logo. This information was held constant across conditions. We then showed participants the team's logo which was identical across conditions. Figure 6 shows logo stimuli used. Participants were asked to evaluate whether they thought the team's logo was likely the most creative in the competition (Yes/No).

Participants in the *separate evaluation* condition followed the same procedure and completed the same dependent measure as participants in the *joint evaluation* condition except that they only saw a photo of the focal team they evaluated. That is, while participants in the *separate evaluation* condition were told that five

teams competed in the challenge, they did not see photos of the other four teams prior to evaluating the focal team.

Results and Discussion

We tested whether there was an interaction between team composition condition (diverse versus homogeneous) and evaluation type condition (joint versus separate) on logo evaluation. This analysis revealed a main effect of team composition ($b = -0.258$, $SE = 0.07$, Wald $\chi^2(1) = 12.25$, $p < 0.001$), whereby, overall, participants who saw a diverse team were more likely to judge the logo to be the most creative in the competition compared with participants who saw a homogeneous team. This main effect was qualified by a significant interaction ($b = 0.261$, $SE = 0.07$, Wald $\chi^2(1) = 12.56$, $p < 0.001$). In the *joint evaluation* condition, there was a significant effect of team composition. A larger proportion of participants judged the logo as likely to be the most creative in the competition when the logo was attributed to a diverse team (73.4%) compared with when an identical logo was attributed to a homogeneous team (49.5%; $\chi^2(1, n = 399) = 24.23$, $p < 0.001$, $w = 0.25$). In the *separate evaluation* condition, the effect of team composition was nonsignificant ($\chi^2(1, n = 399) = 0.001$, $p = 0.974$, $w = 0.00$); 57.8% of participants judged the logo as likely to be most creative when attributed to a diverse team, whereas 58.0% of participants judged the logo as likely to be most creative when attributed to a homogenous team. Table 3 shows effects by homogenous team race within evaluation type condition (joint versus separate).

Thus, Study 3's results showed that in the context of product evaluation, racially diverse teams' output was judged as more creative compared with identical output by racially homogenous teams, but this effect was specific to an evaluative context in which participants viewed a series of teams varying on demographic make-up before evaluating the focal team's product.

Table 3. Comparison of Teams by Homogenous Team Race (Study 3)

Team comparison	% choosing homogenous team's logo as most creative versus % choosing diverse team's logo as most creative	n	χ^2	p	w
Joint evaluation condition					
Asian male teams versus diverse teams	44.9 versus 73.4	256	14.82	<0.001	0.24
White male teams versus diverse teams	37.2 versus 73.4	250	21.19	<0.001	0.29
Asian female teams versus diverse teams	62.5 versus 73.4	255	2.28	0.131	0.09
White female teams versus diverse teams	51.9 versus 73.4	259	9.01	0.003	0.19
Separate evaluation condition					
Asian male teams versus diverse teams	66.7 versus 57.8	243	1.31	0.252	0.07
White male teams versus diverse teams	52.6 versus 57.8	249	0.480	0.488	0.04
Asian female teams versus diverse teams	54.9 versus 57.8	243	0.139	0.709	0.02
White female teams versus diverse teams	58.3 versus 57.8	240	0.004	0.948	0.00

We did not find our predicted effect when the focal team was judged in isolation, a result we discuss further in the General Discussion.

Study 4a

In Study 2, we provided evidence of our proposed mechanism (that perceivers view demographic diversity as linked to cognitive diversity) via mediation. Study 4a's goal was to provide evidence of mechanism through moderation. We argue that one reason why observers attribute greater creativity to products created by demographically diverse teams as opposed to demographically homogeneous teams is because observers assume that demographically diverse teams are more cognitively diverse compared with demographically homogenous teams, and cognitive diversity is seen as important for creativity. If this account is correct, we would expect that participants' tendency to attribute more creativity to a racially diverse team's output than to a racially homogenous team's output would be attenuated if we described the racially diverse team as lacking in cognitive diversity and described the racially homogeneous team as possessing cognitive diversity.

Thus, in Study 4a, observers viewed both a racially diverse team and a racially homogeneous team and a product idea ostensibly created by each team. All participants viewed two products. We randomized which team (diverse or homogeneous) was assigned which product. In addition, observers were randomly assigned to one of two experimental conditions: a *baseline* condition in which observers simply chose which team's output they thought was more creative, and a *cognitive diversity manipulation* condition, in which observers read information indicating that the racially diverse team was low on cognitive diversity while the racially homogenous team was high on cognitive diversity before choosing which team's output they thought was more creative. We predicted that, in the *baseline* condition, observers would be more likely to select the diverse team's product as more creative, consistent with the results from our previous studies. However, we predicted that this effect would be diminished in the *cognitive diversity manipulation* condition, in which we manipulated the cognitive diversity of the teams such that the racially diverse team was portrayed as having less cognitive diversity than the racially homogenous team.

Method

Participants. We requested 1,092 participants from CloudResearch. This sample size was determined by a power analysis using the effect size found in a pilot study ($n = 402$, see Study S2 in the online appendix). U.S. residents with a >95% approval rating were invited to participate. A total of 1,095 participants completed the survey, in exchange for $0.40 each. No participants were excluded from analyses. This study was preregistered at Aspredicted.org (#96341).

Materials and Procedure. Participants read that they would evaluate the creativity of product ideas generated by two teams of consultants. Participants were then randomly assigned to either a *baseline* condition or a *cognitive diversity manipulation* condition. In the *baseline* condition, participants saw an all-White team and a racially diverse team, presented side by side. Teams were constructed using headshots taken from the Chicago Face Database (Ma et al. 2015). From this database, we selected 20 faces (5 East Asian male faces, 5 Black male faces, 5 Hispanic male faces, and 5 White male faces). All faces had a neutral expression and had previously been rated as between the ages of 25 and 35 and as relatively unambiguous in terms of their sex and race. Figure 7 shows face stimuli used in Study 4a.

From this set of face stimuli, we constructed five pairs of teams, each consisting of an all-White four-person team and a racially diverse four-person team with one Asian member, one Black member, one Hispanic member, and one White member. Across participants, we varied which faces from our stimuli set were shown in the all-White team versus the diverse team. As there were five White male faces in the stimuli set, each participant was shown four of the White male faces together in a homogenous team and the remaining White male face in a diverse team with a randomly selected Asian male face, Black male face, and Hispanic male face. The specific White male face shown in the diverse team was rotated across participants.

In the *cognitive diversity manipulation* condition, participants were randomly assigned to one of the same White team-diverse team pairs as were shown in the *baseline* condition, with the addition of background information provided about each team. Specifically, participants in the *cognitive diversity manipulation* condition read the following about the all-White team: "Each of Team 1's members are trained in a different specialty and each grew up in different circumstances. As such, Team 1's members bring very different knowledge areas, sets of skills, and perspectives to the table." Participants in the *cognitive diversity manipulation* condition also read the following about the diverse team: "Each of Team 2's members are trained in the same specialty and each grew up in similar circumstances. As such, Team 2's members bring very similar knowledge areas, sets of skills, and perspectives to the table." Participants in both conditions then saw the product ideas ostensibly generated by the two teams. We used the two product ideas shown in Figure 8 and randomly varied which product was assigned to which team. Participants were asked to choose which product idea they thought was more creative.

Figure 7. (Color online) Face Stimuli (Study 4a)
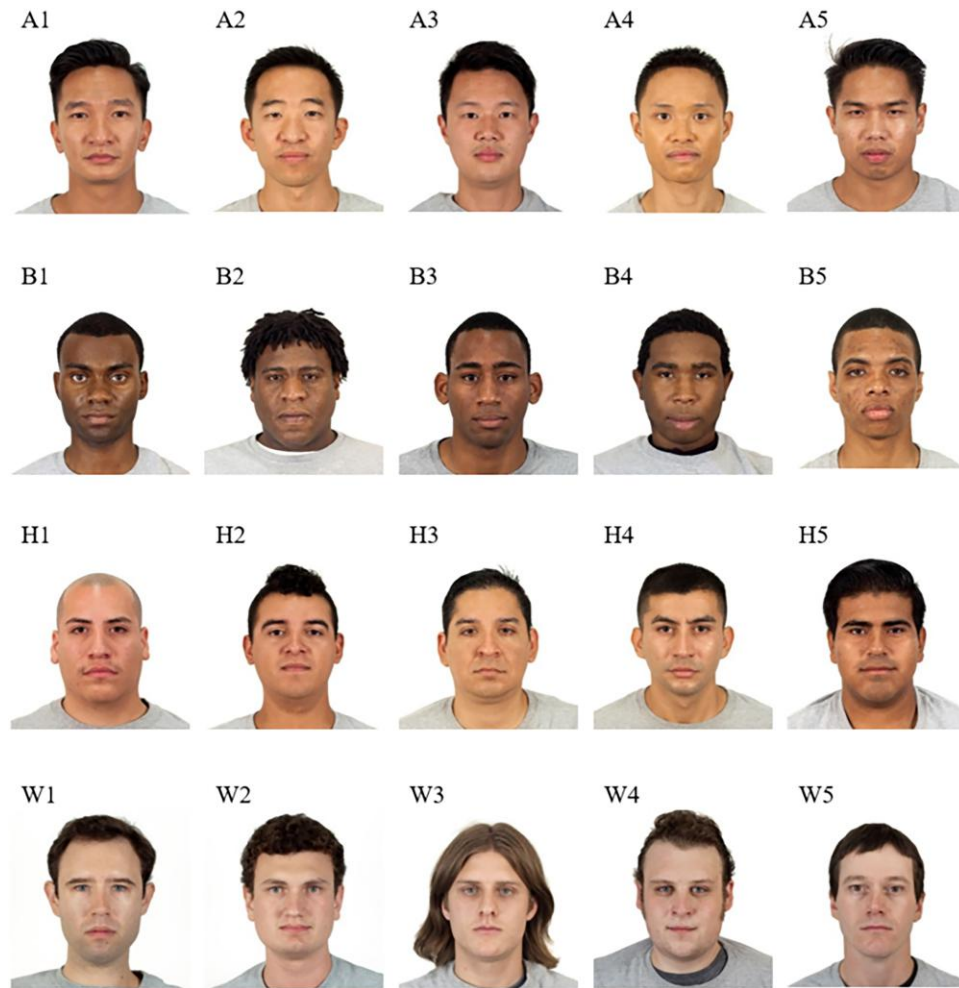
Results and Discussion

Product assignment (stirrer assigned to diverse team versus strainer assigned to diverse team) did not moderate the effect of condition (*baseline* versus *cognitive diversity manipulation*) on product choice ($b = -0.090$, $SE = 0.062$, Wald $\chi^2(1) = 2.08$, $p = 0.150$). Consistent with predictions, as shown in Figure 9, describing the racially diverse team as low in cognitive diversity and the racially homogeneous team as high in cognitive diversity significantly shifted participants' judgments regarding which team's product was more creative ($\chi^2(1, n = 1,095) = 24.83$, $p < 0.001$, $w = 0.15$).

In the *baseline* condition, participants were more likely to choose the diverse team's product as more creative, with 58.0% of participants choosing the diverse team's product as more creative and 42.0% choosing the homogeneous team's product as more creative ($\chi^2(1, n = 553) = 14.32$, $p < 0.001$, $w = 0.16$). This pattern reversed in the *cognitive diversity manipulation* condition, with 43.0% of participants in this condition choosing the diverse team's product as more creative and

57.0% choosing the homogeneous team's product as more creative ($\chi^2(1, n = 542) = 10.66$, $p = 0.001$, $w = 0.14$). These results support our mechanistic account—that the tendency to judge demographically diverse teams as more creative compared with demographically homogenous teams may be driven in part by a tendency to associate demographic diversity with cognitive diversity.

Study 4b

Study 4b aimed to conceptually replicate Study 4a's results using a different manipulation of cognitive diversity. Instead of directly manipulating team cognitive diversity by explicitly describing the diverse team as cognitively homogenous and the homogenous team as cognitively diverse, we provided personal details about each team member which conveyed that the racially homogenous team was cognitively diverse, and the racially diverse team was cognitively homogenous. Consistent with Study 4a, we predicted that, at baseline, observers would be more likely to select the diverse team's product as more creative and that this

**Figure 8.** (Color online) Product Stimuli (Study 4a)

effect would be diminished when we manipulated the cognitive diversity of the teams such that the racially diverse team was portrayed as having less cognitive diversity than the racially homogenous team.

Method
Participants. We requested 1,092 participants from CloudResearch, maintaining the same sample size goal as Study 4a. U.S. residents with a >95% approval rating were invited to participate. A total of 1,108 participants completed the survey, in exchange for $0.40 each. No participants were excluded from analyses. This study was preregistered at Aspredicted.org (#93969).

Materials and Procedure. Materials and procedure were identical to Study 4a, with the exception of the cognitive

Figure 9. Effect of Team Composition and Condition on Product Choice (Study 4a)
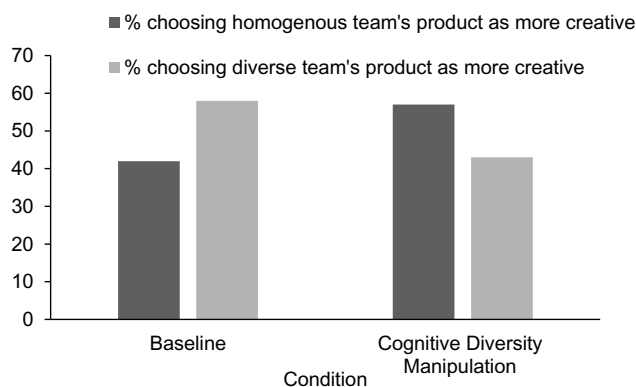
diversity manipulation. Participants in the *cognitive diversity manipulation* condition were provided with information about each team member's hometown, their college major, current role, and their Myers Briggs personality type. Each member of the all-White team was shown to have a different hometown, a different college major, a different current role, and a different Myers Briggs personality type. Conversely, each member of the racially diverse team was shown to have the same hometown, the same college major, the same current role, and the same Myers Briggs personality type.[2] See Figure 10 for an example of the stimuli used.

Results and Discussion
Product assignment (stirrer assigned to diverse team versus strainer assigned to diverse team) did not moderate the effect of condition (*baseline* versus *cognitive diversity manipulation*) on product choice ($b = 0.066$, $SE = 0.064$, Wald $\chi^2(1) = 1.09$, $p = 0.297$). Consistent with predictions, portraying the racially diverse team as lacking cognitive diversity and the racially homogeneous team as high in cognitive diversity significantly shifted participants' judgments regarding which team's product was more creative ($\chi^2(1, n = 1,108) = 15.25$, $p < 0.001$, $w = 0.12$). In the *baseline* condition, participants were more likely to choose the diverse team's product as more creative, with 56.8% of participants choosing the diverse team's product as more creative and 43.2% choosing the homogeneous team's product as more creative ($\chi^2(1, n = 555) = 10.14$, $p = 0.001$, $w = 0.14$). This pattern reversed in the *cognitive diversity manipulation* condition, with 45.0% of participants in this condition choosing the diverse

Figure 10. (Color online) Example Stimuli (Study 4b)

Notes. Racially homogenous team with high cognitive diversity. Racially diverse team lacking cognitive diversity.

team's product as more creative and 55.0% choosing the homogeneous team's product as more creative ($\chi^2(1, n = 553) = 5.47$, $p = 0.019$, $w = 0.10$). Along with the results of Study 4a, these results are consistent with the notion that observers' tendency to judge output by a racially diverse team as more creative compared with identical output by a racially homogenous team may be driven by an assumption that a racially diverse team is more cognitively diverse than a racially homogenous team.

Studies 5a and 5b

The majority of our studies thus far have compared teams that were completely homogeneous (for instance, all members of the same race/ethnicity) to teams that were completely diverse (e.g., each member of a different race/ethnicity). Given that teams are often somewhere in the middle between these two extremes (e.g., having a majority of members of one race or gender and a minority of members of another race or gender), in Studies 5a and 5b, we explored the how much diversity a group needed to be seen as creative. To do so, we adapted the diversity threshold paradigm of Danbold and Unzueta (2020). We randomly assigned observers to view one of 21 organizations. We varied the demographic make-up of the organization's workforce in 5% increments. In Study 5a, the organizations ranged from 100% racial minority employees/0% White employees to 100% White employees/0% racial minority employees.

In Study 5b, the organizations ranged from 100% women employees/0% men employees to 100% men employees/0% women employees. We asked observers to assess whether the organization had sufficient diversity for creativity. In both studies we tested whether there was an inverted U-shaped relationship between degree of diversity and belief that the organization was sufficiently diverse to produce creative ideas.

Study 5a

Study 5a examined how the degree of racial diversity in an organization's workforce impacted judgments of that organization's creativity. We predicted a curvilinear relationship between proportion of racial minority employees in an organization and judgments that the organization had sufficient diversity to be creative.

Method

Participants. We recruited a nationally representative sample of 400 U.S. residents using Lucid theorem's quota survey sampling service (Coppock and McClellan 2019). We did not know exactly what effect size to expect; thus, we aimed to maximize power by recruiting 400 participants, which provided adequate power to detect a small- to medium-sized effect. A total of 405 participants completed the survey.[3] This study was preregistered at Aspredicted.org (#94347).

Figure 11. Example Stimuli (Study 5a)

Racial minority employees = 45%
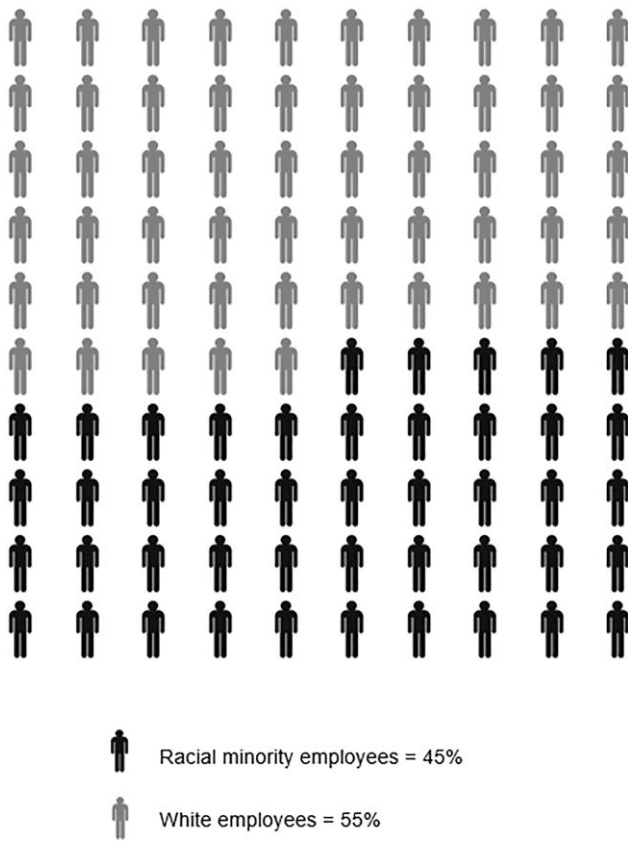
White employees = 55%

Materials and Procedure. Following the method used by Danbold and Unzueta (2020), participants were randomly assigned to evaluate 1 of 21 organizations. The organizations differed in intervals of 5% from 0% racial minority employees/100% White employees to 100% racial minority employees/0% White employees. Figure 11 shows the stimuli used in one of the conditions.

Participants rated the organization on three items using a seven-point scale (1 = strongly agree to 7 = strongly disagree): (a) "There is enough variety in terms of race at this company for the company to produce highly creative ideas." (b) "The percentage of racial minority employees and white employees in this company's workforce is ideal for creativity." (c) "This company is unlikely to produce innovative ideas for new products" (reverse coded). These items were averaged to form a composite measure reflecting participants' agreement that there was sufficient racial diversity in the organization's workforce for creativity ($\alpha = 0.72$).

Results and Discussion
We used a two-lines test (Simonsohn 2018) to examine whether there was a U-shaped relationship between the percentage of racial minority employees at the organization and participants' agreement that the organization had sufficient diversity for creativity. As predicted,

we found evidence of a U-shaped effect. Results are visualized in Figure 12. The relationship between percentage of racial minority employees and agreement that the organization had sufficient racial diversity for creativity was positive and significant at low to moderate percentages of racial minority employees ($b = 0.03$, $z = 4.73$, $p < 0.001$) and was negative and significant at moderate to high percentages of racial minority employees ($b = -0.03$, $z = -3.33$, $p = 0.001$).

We also examined whether mean ratings within each condition differed from the scale midpoint (4) of our composite measure. Results are summarized in Table 4. These analyses indicated that participants judged organizations with between 45% and 65% racial minority employees to have sufficient racial diversity (i.e., above the midpoint on the scale) for creativity.

Study 5b
Study 5b examined how the degree of gender diversity in an organization's workforce impacted judgments of that organization's creativity. We predicted a curvilinear relationship between proportion of women employees in an organization and judgments that the organization had sufficient diversity to be creative.

Method
Participants. Maintaining the same sample size goal as Study 5a, we recruited a nationally representative sample of 400 U.S. residents using the Lucid theorem. A total of 406 participants completed the survey. This study was preregistered at Aspredicted.org (#94405).

Materials and Procedure. Participants were randomly assigned to evaluate one of 21 organizations. Materials were similar to those of Study 5a except that the organizations differed in intervals of 5% from 0% women employees/100% men employees to 100% women employees/0% men employees. Participants rated the organization on three items using a seven-point scale (1 = strongly agree to 7 = strongly disagree): (a) "There is enough variety in terms of gender at this company for the company to produce highly creative ideas." (b) "The percentage of men and women in this company's workforce is ideal for creativity." (c) "This company is unlikely to produce innovative ideas for new products" (reverse coded). These items were averaged to form a composite measure reflecting participants' agreement that there was sufficient gender diversity at the organization for creativity ($\alpha = 0.77$).

Results and Discussion
Results are visualized in Figure 13. As predicted, a two-lines test showed that the relationship between percentage of women employees and agreement that the organization had sufficient gender diversity for creativity

Figure 12. Effect of Condition on Agreement with Sufficient Racial Diversity for Creativity Statements (Study 5a)
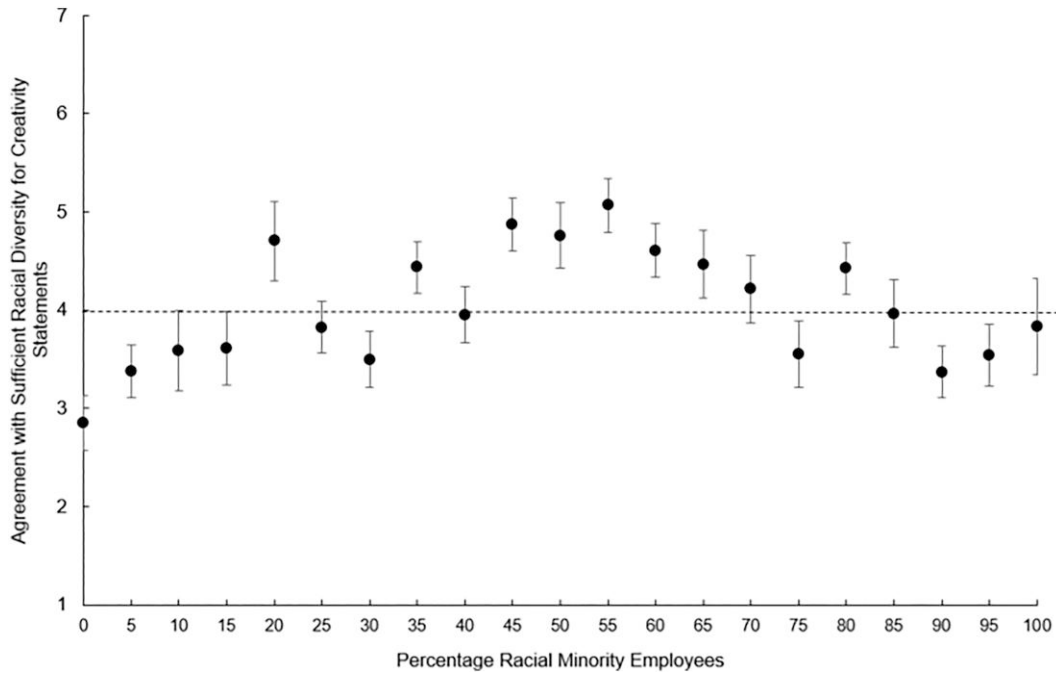
Note. Standard errors bars are shown.

was positive and significant at low to moderate percentages of women employees ($b = 0.06$, $z = 7.21$, $p < 0.001$) and was negative and significant at moderate to high percentages of women employees ($b = -0.05$, $z = -5.42$, $p < 0.001$), providing evidence of a U-shaped effect.

We also examined whether mean ratings within each condition differed from the scale midpoint (4) of our composite measure. Results are summarized in Table 4. The pattern of results suggests that participants judged organizations with between 45% and 55% women employees

Table 4. Results (Studies 5a and 5b)

Study 5a						Study 5b					
Percentage racial minority employees	*M*	*SD*	*t*	*df*	*p*	Percentage women employees	*M*	*SD*	*t*	*df*	*p*
100%	3.83	1.71	−0.338	11	0.742	100%	2.31	0.95	−6.68	13	<0.001*
95%	3.54	1.28	−1.44	15	0.171	95%	3.32	1.46	−2.29	23	0.032*
90%	3.37	1.10	−2.42	17	0.027*	90%	2.98	1.05	−4.22	18	<0.001*
85%	3.97	1.56	−0.096	19	0.925	85%	3.51	1.76	−1.15	16	0.269
80%	4.43	1.32	1.61	24	0.120	80%	3.68	1.40	−0.984	18	0.338
75%	3.56	1.65	−1.32	23	0.200	75%	3.58	1.25	−1.61	22	0.122
70%	4.22	1.54	0.630	19	0.536	70%	3.86	1.55	−0.447	22	0.659
65%	4.47	1.09	1.35	9	0.209	65%	4.02	1.21	0.062	19	0.952
60%	4.61	1.13	2.23	16	0.041*	60%	4.76	1.73	1.69	14	0.113
55%	5.07	1.50	3.90	29	<0.001*	55%	5.17	1.15	5.55	29	<0.001*
50%	4.76	1.53	2.29	20	0.033*	50%	5.45	0.91	6.56	16	<0.001*
45%	4.87	1.24	3.24	20	0.004*	45%	4.89	1.47	2.42	15	0.029*
40%	3.95	1.35	−0.157	21	0.876	40%	4.04	1.43	0.117	15	0.909
35%	4.44	1.05	1.66	15	0.117	35%	3.95	1.33	−0.169	19	0.868
30%	3.50	1.29	−1.74	19	0.098	30%	3.48	1.49	−1.62	20	0.122
25%	3.83	1.32	−0.657	24	0.518	25%	3.54	1.37	−1.45	18	0.164
20%	4.70	1.21	1.75	8	0.118	20%	3.52	1.34	−1.61	19	0.123
15%	3.62	1.62	−1.03	19	0.318	15%	2.71	1.01	−6.24	23	<0.001*
10%	3.59	1.73	−0.998	17	0.332	10%	2.78	1.65	−2.88	14	0.012*
5%	3.38	0.99	−2.33	13	0.037	5%	2.22	1.40	−4.93	14	<0.001*
0%	2.85	1.45	−4.11	26	<0.001*	0%	2.39	1.22	−5.77	18	<0.001*

*$p < 0.05$ threshold for statistical significance.

Figure 13. Effect of Condition on Agreement with Sufficient Gender Diversity for Creativity Statements (Study 5b)
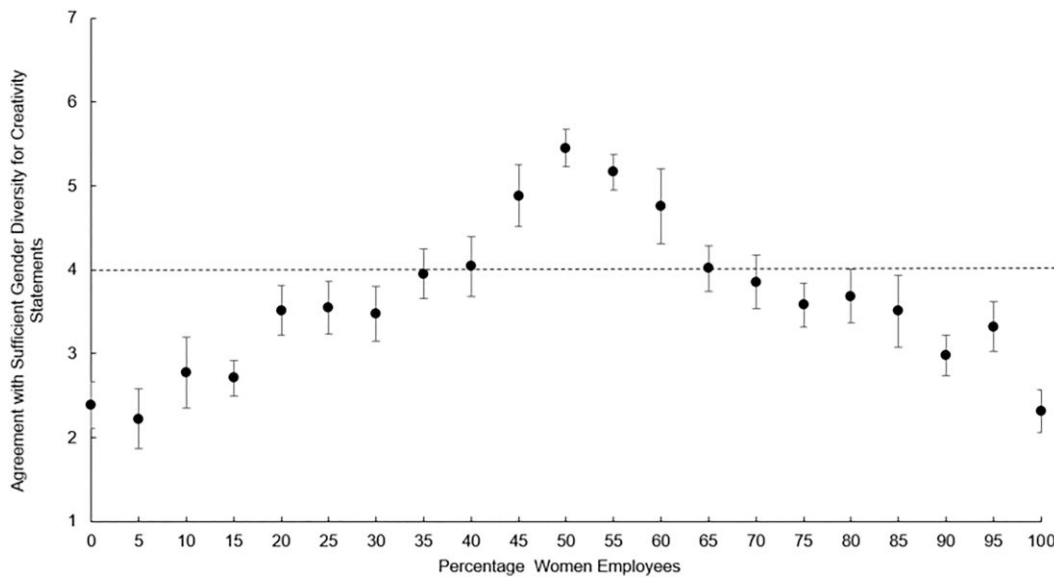
Note. Standard errors bars are shown.

to have sufficient gender diversity (i.e., above the mid-point on the scale) for creativity.

Together, the results of Study 5a and Study 5b provide evidence of a curvilinear relationship between degree of demographic diversity along the lines of race and gender and perceptions of that group's ability to generate creative ideas. We find that observers perceive there to be an optimal amount of demographic heterogeneity in groups for creativity—for racial diversity, 45%–65% racial minorities, and for gender diversity, 45%–55% women.

Study 6

Study 6 explored an additional implication of our prediction that observers associate diversity with creativity. If we are correct that observers hold a lay belief that demographically diverse teams are more creative compared with demographically homogenous teams, we would expect this belief to impact evaluators' decisions about how to assemble teams for creative tasks. Thus, in Study 6, rather than ask observers to judge the creativity of homogenous and diverse teams, we manipulated whether a team's task involved creativity and measured whether observers chose to add a member to the team who increased the team's racial diversity. We predicted that observers would be particularly likely to add new members to teams who increased the team's racial diversity when the team's task required creativity.

Method
Participants. We requested 400 participants from Cloud-Research, maintaining the same sample size goal as Study 3. U.S. residents with a >95% approval rating were invited to participate. A total of 400 participants

completed the survey, in exchange for $0.40 each. No participants were excluded from analyses. This study was preregistered at Aspredicted.org (#78093).

Materials and Procedure. Participants read that the purpose of the study was to understand how evaluators selected new members to join teams. Participants then read, "Imagine that you are a manager at a mid-sized company. Your role includes assembling teams of employees to come up with ideas for how to increase the company's market share." Participants were then randomly assigned to one of two task conditions (creative versus noncreative). Participants in the *creative task* condition read, "While some teams at the company have been assembled to specifically generate practical, non-creative ideas for how to increase market share, your job is to assemble a team that will specifically generate extremely creative, innovative ideas for how to increase market share." Participants in the *noncreative task* condition read, "While some teams at the company have been assembled to specifically generate creative, innovative ideas for how to increase market share, your job is to assemble a team that will specifically generate extremely practical, non-creative ideas for how to increase market share." Participants in both conditions were then shown a photo of a team with two male members. Team photos were constructed using headshots from the Face Research Laboratory London Set (DeBruine and Jones 2017). From this database, we selected two East Asian male faces, two Black male faces, and two White male faces, based on the self-reported race and gender of the photographed individuals. All faces selected had a neutral expression and were self-reported to be between the ages of 20 and

Figure 14. (Color online) Photo Stimuli for Team Pairs (Study 6)

30. Using these headshots, we create six two-person teams (Figure 14). Three of the teams were composed of members of the same race (White-White, Asian-Asian, Black-Black), and three of the teams were composed of members of different races (Asian-Black, Asian-White, Black-White).

Participants were randomly assigned to view one of the six teams and were asked to select a candidate to add to the team. Participants were shown three head-shots of candidates to choose between: one Asian male, one Black male, and one White male (Figure 15). These headshots were also from the Face Research Laboratory London Set. Candidate race and gender were determined by self-report. All faces selected had a neutral expression and were self-reported to be between the ages of 20 and 30. Our dependent measure was whether participants chose to diversify the team: If participants chose to add a candidate to the team who was a different race than the existing team members, their response was coded as one. If participants chose to add a candidate to the team who was the same race as at least one existing team member, their response was coded as zero.

Results and Discussion

As predicted, our manipulation of the team's task (creative versus noncreative) shifted participants' tendency to diversify the team $(\chi^2(1, n = 400) = 9.25, p = 0.002, w = 0.15$; Figure 16).

In the *creative task* condition, as expected, 76.5% of participants chose to diversify the team, more than would be expected by chance (i.e., 50%; $\chi^2(1, n = 200) = 56.18, p < 0.001, w = 0.53$). By contrast, in the *noncreative*

Figure 15. (Color online) Candidate Stimuli (Study 6)

Figure 16. Effect of Team Task Condition on Choice to Diversify the Team (Study 6)
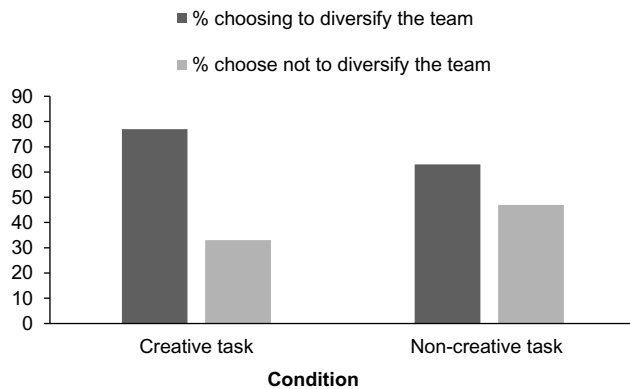
task condition, whereas a larger proportion of participants (62.5%) chose to diversify the team than would be expected by chance (50%; $\chi^2(1, n = 200) = 12.50$, $p < 0.001$, $w = 0.25$), this proportion was significantly smaller than in the creative task condition ($\chi^2(1, n = 400) = 9.25$, $p = 0.002$). This pattern of results is consistent with our prediction that observers should be particularly likely to assemble a racially diverse team when the team's task required creativity.

In additional analyses, we examined the effect of condition separately for participants who saw a team that started out as a homogenous pair (T1 T2 T3) and for participants who saw a team that started out as a diverse pair (T4 T5 T6). For participants who saw a team that started out as a homogenous pair (T1 T2 T3), our manipulation of the team's task shifted participants' tendency to diversify the team ($\chi^2(1, n = 198) = 5.72$, $p = 0.017$, $w = 17$). In the *creative task* condition, 85.1% of participants chose to diversify the team when it started out as homogenous, significantly more than would be expected by chance (i.e., 66.7%; $\chi^2(1, n = 101) = 15.53$, $p < 0.001$, $w = 0.39$). By contrast, in the *noncreative task* condition, 71.1% chose to diversify the team when it started out as homogenous, which was not significantly different than what would be expected by chance (66.7%; $\chi^2(1, n = 97) = 0.871$, $p = 0.351$, $w = 0.09$).

For participants who saw teams that started out as a diverse pair (T4 T5 T6), our manipulation of the team's task shifted participants' tendency to diversify the team, though the effect was marginally significant ($\chi^2(1, n = 202) = 3.75$, $p = 0.053$, $w = 0.14$). In the *creative task* condition, 67.7% of participants chose to diversify the team when it started out as diverse, significantly more than would be expected by chance (i.e., 33.3%; $\chi^2(1, n = 99) = 52.55$, $p < 0.001$, $w = 0.73$). By contrast, in the *noncreative task* condition, 54.4% chose to diversify the team, which also significantly more than would be expected by chance (33.3%; $\chi^2(1, n = 103) = 20.51$, $p < 0.001$, $w = 0.45$), but a marginally smaller proportion than in the creative task condition. Furthermore, in both the *creative task*

condition and the *noncreative task* condition, we found that a larger proportion of participants chose to diversify the homogenous team than chose to diversify the already diverse team (creative task condition: $\chi^2(1, n = 200) = 8.49$, $p = 0.004$, $w = 0.21$; noncreative task condition: $\chi^2(1, n = 200) = 5.99$, $p = 0.014$, $w = 0.17$).

Overall, these results provide evidence that evaluators show a stronger preference for racially diversifying teams when the team's task is to produce creative ideas relative to when the team's task is to produce noncreative ideas, consistent with our prediction that observers believe that demographic diversity boosts creativity. This effect appears to be relatively robust to whether the team started out as homogenous or diverse.

General Discussion

Eight preregistered studies support our prediction that observers hold a lay belief that diversity increases creativity and apply this lay belief in judgments about teams and their creative work. Specifically, our studies show that observers predict that demographically diverse teams—that is, groups diverse in terms of race and/or gender—will outperform demographically homogenous teams on creative tasks (Studies 1 and 2, Study S1), including in an incentive-compatible study in which participants were paid for accuracy and teams were evaluated in isolation, rather than jointly (Study 2). We also show that identical output is judged more creative when attributed to demographically diverse teams compared with demographically homogenous teams (Studies 3–4b). We identify one mechanism underlying these effects: Observers believe that demographic diversity is associated with cognitive diversity, with cognitive diversity contributing to perceptions of creativity. Our mechanistic account is supported by evidence from mediation (Study 2) and moderation (Study 4a and 4b, Study S2). Furthermore, we find evidence of a curvilinear relationship between the proportion of racial minorities or women in a group and judgments of the group's creativity (Studies 5a and 5b). We also find that, although evaluators choose to diversify teams for both creative and noncreative tasks, evaluators more often choose candidates who add demographic diversity to teams when the team's task requires creativity (Study 6).

Contributions and Implications
Although our results may not be surprising—after all, the argument that diversity boosts creativity is a core element of the "business case for diversity" (Kochan et al. 2003, Mannix and Neale 2005)—we believe they are important to document, as the aggregation of decades of research on group creativity does not show a robust positive association between demographic diversity and group creative performance (Bell et al. 2011, Wang et al. 2019, Byron et al. 2023). Thus, our

findings not only suggest that evaluators tend to apply a heuristic—a *diversity heuristic* one might say—when assessing team creativity, but in the context of the literature on the actual relationship between demographic diversity and team creativity, our results suggest that this heuristic can lead evaluators astray, impeding accurate assessment of teams and their work. Our results suggest that the tendency to attribute greater creativity to demographically diverse teams compared with demographically homogenous teams may influence organizational decisions regarding which teams are assigned to creative projects, which teams' creative ideas are ultimately selected and invested in, and how creative teams are assembled.

Furthermore, we find that people believe that the inclusion of a variety of different races in groups and teams facilitates creativity. This result points to a potential silver lining for diverse teams, given past research showing that diverse teams may be subject to negative evaluations by outsiders on a variety of outcomes (Van Dijk et al. 2012, Lount et al. 2015). We show that diverse teams may benefit along the dimension of creativity evaluations, even in an incentive-compatible study where participants are paid for accuracy in predicting the output of real teams. Yet the pattern of results we find points to implications for members of minority populations and women that are more complex. We demonstrate that the tendency to associate diversity with creativity may benefit non-White individuals and women, but only if they are members of diverse groups. Our studies show that creativity judgments of racial minority and women individuals depend on group composition—minority individuals and women working in groups with others who share their social category membership were less likely to be seen as creative compared with similar individuals working in groups with socially dissimilar others. In addition, we find that the assumption that cognitive diversity underlies demographic diversity explains people's beliefs about why diversity facilitates team creativity. This mechanism suggests that the promotion of this particular lay theory of creativity may unintentionally reinforce stereotyping and essentialist thinking—the notion that those who "look different" also "think different" (Fiske and Neuberg 1990, Bastian and Haslam 2006). Our work thus adds to a growing number of studies illuminating how diversity-related beliefs that seem favorable for minorities may have unintended negative consequences (Gündemir and Galinsky 2018, Wilton et al. 2019).

As demographic diversity is increasing the U.S. labor force and as workers are increasingly collaborating in small groups on projects rather than working alone (Kozlowski and Bell 2003, Wuchty et al. 2007, Toossi 2015), our research also provides a timely extension of previous work on the role of social cues in creativity evaluation. Although previous studies have focused on how attributes of individuals influence judgments of their personal creativity (Elsbach and Kramer 2003, Proudfoot et al. 2015), our work identifies a key group-level attribute—a collection of individuals' aggregated level of demographic diversity—influencing judgments about a group's collective creativity. Our research is also the first to consider observers' implicit theories of creative groups, thus contributing to a growing literature exploring people's lay theories about where creative ideas come from and the factors that are important for creativity (Sternberg 1985, Loewenstein and Mueller 2016).

Limitations and Future Directions

An important limitation of our research pertains to our test of mechanism. We focused on perceived cognitive diversity of the team as one factor driving observers' tendency to attribute more creativity to demographically diverse teams than to demographically homogenous teams. However, it is possible that observers are making many mediating inferences that could be contributing to our effect which we did not measure. One possibility, for instance, is that observers assume more close-minded people self-select into homogenous groups than into diverse groups, thus explaining our effects. We tested this possibility in a supplemental study, finding that our effects were not moderated by whether observers were told the members self-selected into the team or were randomly assigned to the team (see Study S3 in the online appendix). It is also possible that observers are making inferences about different group processes taking place in diverse groups and homogenous groups, which could also contribute to our effects. For instance, observers might tend to believe that cognitive diversity in demographically diverse teams facilitates creative performance because cognitive diversity produces greater task conflict. More research is needed to fully explore the inferences observers are making—including inferences about group process—that might contribute to the tendency to attribute more creativity to diverse teams than to homogenous teams.

In addition, our research specifically examined how demographic differences along the dimensions of race and gender informed judgments of group creativity. We did so because race and gender are highly salient forms of demographic difference and central to lay conceptions of diversity (Unzueta and Binning 2010). Future research is needed to explore whether a group's degree of perceived diversity on other demographic dimensions (e.g., age) and on combinations of demographic dimensions might influence a group's perceived creativity, as well as which other types of demographic diversity not examined here are understood as particularly strong indicators of group creativity.

Another limitation of our studies is that they relied exclusively on experiments conducted with laboratory

and online samples. Although our results hold both in hypothetical scenarios and in incentive-compatible judgments of real teams, future tests of these theories in the field would undoubtedly be useful. Our findings may also be specific to Western cultural contexts where the belief that diversity is beneficial for group performance is particularly pervasive (Ely and Thomas 2020). As people's lay theories of creativity have been shown to differ across cultures (Loewenstein and Mueller 2016), future studies are needed to examine the generalizability of our findings to non-U.S. settings.

It is also important to note potential boundary conditions to our effects. First, our studies provided little extra information about the teams being evaluated beyond their level of demographic diversity. Although this approach allowed us to precisely isolate the effect of demographic diversity on judgments of group creativity, it also raises the possibility of a boundary condition—in situations in which much more is known about the teams being evaluated (such as in Study 4b's manipulation of cognitive diversity), observers likely have more direct evidence of their cognitive diversity, which would make demographic diversity a less potent cue for their creativity. Second, we find weaker effects for comparisons involving some of the homogenous teams we included in our studies. Specifically, we did not find a statistically significant effect when we compared judgments of the diverse team with the team of Asian men (Study 2) or when we compared the diverse team with the team of Asian women (Study 3). Future studies are needed to more directly explore whether certain types of demographically homogenous groups may be seen as similarly creative to diverse groups.

Finally, although we find our effects across study designs, including between subjects and within subjects, incentivized and not, and in joint and separate evaluation, Study 3 provides evidence that our effects may be more robust in joint evaluation as compared with separate evaluation. Specifically, while we find our predicted effects in separate evaluation in Study 2 and Studies 5a and 5b, in which participants are asked to make predictions about team creative performance, Study 3 showed that when asked to judge products, our predicted effect only emerged when participants evaluated the focal team in the context of other teams; we did not find the same effect when participants only saw one team in isolation. In many creativity evaluation contexts, people engage in joint evaluation (e.g., when judging which idea to endorse or move forward with at work; when deciding which projects to back or fund on Kickstarter or in venture capital; in tournament-style competitions like the Oscars) as opposed to separate evaluation, but the evidence we present suggests that our effects may be more likely to occur in contexts in which teams are evaluated in the context of other teams.

Concluding Remarks

Demographic diversity is often celebrated as an important precursor to creativity and innovation, despite mixed evidence supporting this claim. Our studies show that the lay belief that diversity enhances creativity leads people to attribute more creativity to demographically diverse groups compared with demographically homogenous groups, even when both groups generate the exact same idea. Our findings document a potential barrier to identifying creativity in organizational contexts.

Endnotes

[1] Our research assistants' blind scoring of the teams' sets of ideas revealed that the diverse team's set of ideas were, in fact, the most creative. Hence, in this instance, observers who predicted that the diverse team would be most creative were accurate.

[2] We planned to measure the perceived cognitive diversity of each team after participants chose which team's product they thought was more creative using Study 2's measure to confirm that our subtle manipulation of cognitive diversity was effective. However, because of a programming error, only participants in the *cognitive diversity manipulation* condition rated the cognitive diversity of the teams. Analysis of these ratings showed that our manipulation was successful in this condition. The racially homogenous team (which was described as having higher variety of experiences and perspectives) was rated as more cognitively diverse ($M = 5.02$, $SD = 1.19$) compared with the racially diverse team (which was described as having no variety of experiences and perspectives; $M = 3.20$, $SD = 1.44$; $t(535) = 21.98$, $p < 0.001$, $d = 0.95$). This result is in contrast to Study 2, which showed that, at baseline, the racially homogeneous teams were rated as less cognitively diverse compared with the racially diverse team.

[3] The data sets for Study 5a and Study 5b, both recruited from the Lucid theorem, included extra responses that the Lucid theorem classified as "speeders." The Lucid theorem did not count these responses toward the nationally representative samples recruited. In the main text of the manuscript, we report results from the nationally representative samples. For transparency, in the online appendix, we report the results of Study 5a and Study 5b with the full samples (no responses excluded) and with the first 400 responses collected for each sample. In both studies, results are consistent across samples using different exclusion criteria.

References

Bastian B, Haslam N (2006) Psychological essentialism and stereotype endorsement. *J. Experiment. Soc. Psych.* 42:228–235.

Bell JM, Hartmann D (2007) Diversity in everyday discourse: The cultural ambiguities and consequences of "happy talk". *Amer. Sociol. Rev.* 72:895–914.

Bell ST, Villado AJ, Lukasik MA, Belau L, Briggs AL (2011) Getting specific about demographic diversity variable and team performance relationships: A meta-analysis. *J. Management* 37:709–743.

Berg JM (2016) Balancing on the creative highwire: Forecasting the success of novel ideas in organizations. *Admin. Sci. Quart.* 61:433–468.

Byron K, Keem S, Darden T, Shalley CE, Zhou J (2023) Building blocks of idea generation and implementation in teams: A meta-analysis of team design and team creativity and innovation. *Personality Psych.* 76:49–278.

Catalyst (2014) *Diversity Matters* (Catalyst, New York).

Coppock A, McClellan OA (2019) Validating the demographic, political, psychological, and experimental results obtained from a new source of online survey respondents. *Research and Politics* 6:1–14.

Cuddy AJ, Fiske ST, Glick P (2008) Warmth and competence as universal dimensions of social perception: The stereotype content model and the BIAS map. *Adv. Experiment. Soc. Psych.* 40:61–149.

Czopp AM, Kay AC, Cheryan S (2015) Positive stereotypes are pervasive and powerful. *Perspectives Psych. Sci.* 10:451–463.

Danbold F, Unzueta MM (2020) Drawing the diversity line: Numerical thresholds of diversity vary by group status. *J. Personality Soc. Psych.* 118:283.

Daniels DP, Neale MA, Greer LL (2017) Spillover bias in diversity judgment. *Organ. Behav. Human Decision Processes* 139:92–105.

DeBruine LM, Jones BC (2017) Face research laboratory London set. Retrieved October 26, 2021, https://doi.org/10.6084/m9.figshare.5047666.v3.

Ellemers N, Rink F (2016) Diversity in work groups. *Current Opinion Psych.* 11:49–53.

Elsbach KD, Kramer RM (2003) Assessing creativity in Hollywood pitch meetings: Evidence for a dual-process model of creativity judgments. *Acad. Management J.* 46:283–301.

Ely RJ, Thomas DA (2020) Getting serious about diversity: Enough already with the business case. *Harvard Bus. Rev.* 98:114–122.

Eswaran V (2019) The business case for diversity in the workplace is now overwhelming. *World Economic Forum.* Retrieved https://www.weforum.org/agenda/2019/04/business-case-for-diversity-in-the-workplace/.

Fiske ST, Neuberg SL (1990) A continuum of impression formation, from category-based to individuating processes: Influences of information and motivation on attention and interpretation. *Advances in Experimental Social Psychology* (Academic Press, New York), 1–74.

Galinsky AD, Todd AR, Homan AC, Phillips KW, Apfelbaum EP, Sasaki SJ, Richeson JA, et al. (2015) Maximizing the gains and minimizing the pains of diversity: A policy perspective. *Perspectives Psych. Sci.* 10:742–748.

Guilford JP (1967) *The Nature of Human Intelligence* (McGraw-Hill, New York).

Gündemir S, Galinsky AD (2018) Multicolored blindfolds: How organizational multiculturalism can conceal racial discrimination and delegitimize racial discrimination claims. *Soc. Psych. Personality Sci.* 9:825–834.

Harrison DA, Price KH, Gavin JH, Florey AT (2002) Time, teams, and task performance: Changing effects of surface-and deep-level diversity on group functioning. *Acad. Management J.* 45:1029–1045.

Haslam N, Whelan J (2008) Human natures: Psychological essentialism in thinking about differences between people. *Soc. Personality Psych. Compass* 2:1297–1312.

Hayes AF (2017) *Introduction to Mediation, Moderation, and Conditional Process Analysis: A Regression-Based Approach* (Guilford Publications).

Ito TA, Urland GR (2003) Race and gender on the brain: Electrocortical measures of attention to the race and gender of multiply categorizable individuals. *J. Personality Social Psych.* 85:616.

Kanter RM (2010) Think outside the building. *Harvard Bus. Rev.* 88:44.

Kochan T, Bezrukova K, Ely R, Jackson S, Joshi A, Jehn K, Leonard J, et al. (2003) The effects of diversity on business performance: Report of the diversity research network. *Human Resources Management* 42:3–21.

Kozlowski SW, Bell BS (2003) *Work Groups and Teams in Organizations.* Borman WC, Ilgen DR, Klimoski RJ, eds. *Handbook of Psychology: Industrial and Organizational Psychology*, vol. 12 (Wiley, New York), 333–375.

Levine SR (2020) Diversity confirmed to boost innovation and financial results. Retrieved https://www.forbes.com/sites/forbesinsights/2020/01/15/diversity-confirmed-to-boost-innovation-and-financial-results/#7c620165c4a6.

Litman L, Robinson J, Abberbock T (2017) TurkPrime.com: A versatile crowdsourcing data acquisition platform for the behavioral sciences. *Behav. Res. Methods* 49(2):433–442.

Loewenstein J, Mueller J (2016) Implicit theories of creative ideas: How culture guides creativity assessments. *Acad. Management Discovery* 2:320–348.

Lount RB Jr, Sheldon OJ, Rink F, Phillips KW (2015) Biased perceptions of racially diverse teams and their consequences for resource support. *Organ. Sci.* 26:1351–1364.

Ma DS, Correll J, Wittenbrink B (2015) The Chicago face database: A free stimulus set of faces and norming data. *Behav. Res. Methods* 47:1122–1135.

Mannix E, Neale MA (2005) What differences make a difference? The promise and reality of diverse teams in organizations. *Psych. Sci. Public Interest* 6:31–55.

Milliken FJ, Martins LL (1996) Searching for common threads: Understanding the multiple effects of diversity in organizational groups. *Acad. Management Rev.* 21:402–433.

Mueller JS, Wakslak CJ, Krishnan V (2014) Construing creativity: The how and why of recognizing creative ideas. *J. Experiment. Soc. Psych.* 51:81–87.

Peer E, Brandimarte L, Samat S, Acquisti A (2017) Beyond the Turk: Alternative platforms for crowdsourcing behavioral research. *J. Experiment. Soc. Psych.* 70:153–163.

Phillips KW (2014) How diversity makes us smarter. *Sci. Amer.* 311:43–47.

Phillips KW, Northcraft GB, Neale MA (2006) Surface-level diversity and decision-making in groups: When does deep-level similarity help? *Group Processing Intergroup Relations* 9:467–482.

Phillips LT, Slepian ML, Hughes BL (2018) Perceiving groups: The people perception of diversity and hierarchy. *J. Personality Soc. Psych.* 114(5):766.

Proudfoot D, Kay AC, Koval CZ (2015) A gender bias in the attribution of creativity: Archival and experimental evidence for the perceived association between masculinity and creative thinking. *Psych. Sci.* 26:1751–1761.

Rock D, Grant H (2016) Why diverse teams are smarter. Retrieved https://hbr.org/2016/11/why-diverse-teams-are-smarter.

Runco MA, Bahleda MD (1986) Implicit theories of artistic, scientific, and everyday creativity. *J. Creative Behav.* 20:93–98.

Runco MA, Jaeger GJ (2012) The standard definition of creativity. *Creative Res. J.* 24:92–96.

Schiller B (2018) Want a more innovative company? Simple: Hire a more diverse workforce. Retrieved https://www.fastcompany.com/40515712/want-a-more-innovative-company-simple-hire-a-more-diverse-workforce.

Silvia PJ, Martin C, Nusbaum EC (2009) A snapshot of creativity: Evaluating a quick and simple method for assessing divergent thinking. *Thinking Skills Creativity* 4:79–85.

Simonsohn U (2018) Two lines: A valid alternative to the invalid testing of U-shaped relationships with quadratic regressions. *Adv. Methods Practice Psych. Sci.* 1(4):538–555.

Sternberg RJ (1985) Implicit theories of intelligence, creativity, and wisdom. *J. Personality Soc. Psych.* 49:607.

Toossi M (2015) Labor force projections to 2024: The labor force is growing, but slowly. Retrieved https://www.bls.gov/opub/mlr/2015/article/labor-force-projections-to-2024.htm.

Unzueta MM, Binning KR (2010) Which racial groups are associated with diversity? *Cultural Diversity Ethnic Minority Psych.* 16(3):443.

Van der Vegt GS, Janssen O (2003) Joint impact of interdependence and group diversity on innovation. *J. Management* 29:729–751.

Van Dijk H, Van Engen ML, Van Knippenberg D (2012) Defying conventional wisdom: A meta-analytical examination of the differences between demographic and job-related diversity relationships with performance. *Organ. Behav. Human Decision Processes* 119:38–53.

Van Knippenberg D, Schippers MC (2007) Work group diversity. *Annu. Rev. Psych.* 58:515–541.

Wang J, Cheng GHL, Chen T, Leung K (2019) Team creativity/innovation in culturally diverse teams: A meta-analysis. *J. Organ. Behav.* 40:693–708.

Williams KY, O'Reilly CA III (1998) Work group diversity. *Res. Organ. Behav.* 20:77–140.

Wilton LS, Apfelbaum EP, Good JJ (2019) Valuing differences and reinforcing them: Multiculturalism increases race essentialism. *Soc. Psych. Personality Sci.* 10(5):681–689.

Wuchty S, Jones BF, Uzzi B (2007) The increasing dominance of teams in production of knowledge. *Science* 316:1036–1039.

Yzerbyt V, Corneille O, Estrada C (2001) The interplay of subjective essentialism and entitativity in the formation of stereotypes. *Personality Soc. Psych. Rev.* 5:141–155.