University of Texas at Arlington
MavMatrix

Psychology Theses | Department of Psychology

2023

Lost in Your Halo: The Influence of Group Attractiveness and Diversity on Perceived Group Performance Ratings

Jade C. Chacon

Follow this and additional works at: https://mavmatrix.uta.edu/psychology_theses

 Part of the Psychology Commons

Recommended Citation

Chacon, Jade C., "Lost in Your Halo: The Influence of Group Attractiveness and Diversity on Perceived Group Performance Ratings" (2023). *Psychology Theses*. 159.
https://mavmatrix.uta.edu/psychology_theses/159

This Thesis is brought to you for free and open access by the Department of Psychology at MavMatrix. It has been accepted for inclusion in Psychology Theses by an authorized administrator of MavMatrix. For more information, please contact leah.mccurdy@uta.edu, erica.rousseau@uta.edu, vanessa.garrett@uta.edu.

Lost in Your Halo: The Influence of Group Attractiveness and Diversity on Perceived Group Performance Ratings

Jade C. Chacon

Masters Thesis

The University of Texas at Arlington

May 2023

Abstract

Research on attractiveness suggests that more attractive individuals are assumed to have superiority on several traits, such as happiness and extraversion. This is termed the halo effect. Additionally, within the diversity literature, although inconsistent, there is evidence supporting the importance of diversity within a team to promote better group performance and to foster creativity and innovation. This study examined, for the first time, whether the physical attractiveness halo effect applies to groups. Particularly, overall group attractiveness, group diversity, and their interactive influence on perceived group creative performance were examined. This experiment manipulated both the attractiveness and the diversity of the groups to be rated, and although the attractiveness hypotheses were not supported, findings suggest those with negative attitudes toward diverse groups perceive homogenous groups to work better together, be more cohesive, and cooperative when compared to diverse groups.

Keywords: attractiveness, group performance, halo effect, diversity

Table of Contents

Lost in Your Halo: The Influence of Group Attractiveness and Diversity on Perceived Group Performance Ratings

There is a common finding within the field of social psychology, that more attractive individuals are assumed to possess personality traits that are more desirable when compared to unattractive individuals (Dion et al., 1972). Known as the physical attractiveness heuristic, this assumption of positive traits toward more attractive individuals may impact those in the workforce who go through periodic job or performance evaluations. Additionally, in the increasingly diverse workforce we currently have, it is important to note the positive relationship between team diversity and outcomes such as creativity and group performance (Cox & Blake, 1991), as well as boardroom diversity being linked to greater, more profitable outcomes in some cases—different nationalities of CEOs are connected to different outcomes such as better solvency or profitability (Adams & Baker, 2021). Job-related diversity is correlated with innovative performance, although there is no correlation seen with demographic diversity, as shown in a meta-analysis by van Dijk et al. (2012). While these are inconsistent findings, whether or not individuals perceive that diverse groups have greater creativity and performance levels was further investigated. This predominantly exploratory study took the findings from individual level physical attractiveness halo effect research, as well as the group diversity research, and examined them at the group level for the first time.

The Physical Attractiveness Halo Effect

The physical attractiveness heuristic may also be called the halo effect, which can be unclear, as the term is also used to describe a measurement error. Regarding this measurement error, Edward L. Thorndike (1920) described a “tendency to think of the person in general as

rather good or rather inferior and to color the judgments of the qualities by this general feeling" (p. 25) and labeled it as a constant error of halo. Given that they are associated with the same name, the halo effect is at times mistaken for the attractiveness heuristic, which I will be discussing here. Not limited to attractiveness, certain traits can be seen as a halo of sorts and push this positive impression onto other personality traits. This is seen with studies like the Nisbett and Wilson (1977) study in which participants were shown either a video of a warm and friendly instructor or a cold instructor, and then rated their looks, mannerisms, and accents as either appealing or irritating. Participants also answered whether their liking for the instructor influenced their ratings. They found that participants who viewed the video of the warm instructor rated them more positively when compared to the cold instructor, meaning that one's global evaluation affects other seemingly independent evaluations as well.

Dion et al. (1972), found that physically attractive individuals are assumed by others to have more socially desirable personalities as well as to lead happier lives than unattractive individuals. They found this through having participants make first impression judgments on photographs in a booklet, then rate the photographs on 32 personality traits, on a Likert scale. Additionally, participants in the Dion et al. (1972) study were asked to rate the social desirability of the photographs, happiness, and occupational success through a series of most/least likely questions (e.g., "which stimulus person is most likely to ever be divorced?"). The results show that attractive individuals were assumed by others to hold higher level positions in their job, have happier marriages, and be happier overall. Overall, attractive individuals were judged to enjoy more fulfilling lives when compared to unattractive individuals.

There is a vast number of studies depicting this trend of attractive people presumed to have more positive traits such as sociability and popularity (Eagly et al., 1991). The expression used to refer to this tendency is called the attractiveness heuristic, or the What is Beautiful is Good effect (Dion et al., 1972). Lucker et al. (1981) had participants rate the physical attractiveness and 12 personality factors of both male and female slides, and found that sexiness, femininity/masculinity, and liking were correlated with physical attractiveness, and that this was stronger for females than for males. A more recent study by Klebl et al. (2022) found that participants assigned more positive moral traits, such as fairness and courageousness, to physically attractive individuals when compared to unattractive individuals.

Additionally, in a 1975 study by Mims et al., participants watched a prerecorded video of an unattractive male and an attractive male who put on either nice or obnoxious demeanors as they debated a topic. The second phase was similar, but with females as the confederates in the video, and participants were then asked to volunteer for a second study with either the attractive or unattractive female from the video before (in person). The attractive and nice confederates were given the highest rating on favorable personality traits, and the unattractive and obnoxious confederates were given the lowest rating. Attractive individuals were able to get more participants to voluntarily help them when compared to unattractive individuals, showing favoritism to attractive people.

The strength and generality of the attractiveness halo effect was examined in a meta-analysis by Eagly et al. (1991). They found that the core traits on which physical attractiveness had the greatest impact were sociability and popularity. Integrity and concern for others were shown to be the least impacted by attractiveness, and potency, adjustment, and intellectual

competence ranged around a moderate level. More attractive people were also shown to be perceived as less modest and vainer compared to unattractive individuals. Eagly et al. (1991) generally showed support for the *what is beautiful is good physical attractiveness heuristic*, but with a few caveats, as the heuristic is not as broadly applicable as it was originally thought to be, and depends mostly on the type of assumption the perceiver is making.

In the education realm, the attractiveness heuristic has been studied to reveal that physical attractiveness in children leads to a greater self-concept—defined as how the child feels about themselves—and higher peer acceptance, both with different- and same-sex peers (Salvia et al., 1975). In the same study, physical attractiveness was also found to be related to personal and social development. These traits in individuals all seem to compound on each other in a snowball-like effect. For instance, if the person's self-perceived attractiveness is lower, this would directly impact their self-concept. Further, this lowered self-concept would lead to a lower level of peer acceptance, as evidenced by Salvia et al. (1975).

Away from education and to a more recent and practical perspective, researchers found that the attractiveness of Chief Executive Officers (CEOs) significantly influenced their compensation. Namely, more attractive CEOs earn more compared to their less attractive counterparts, according to Li et al. (2021). Again, there is little research on the attractiveness heuristic within groups. In a similar line of thought, while not on attractiveness, a study by Naquin and Tynan (2003) found that participants credited good outcomes to the team, but poor outcomes to individuals. They call this the team halo effect. However, it was also found that participants who had more experience with and knowledge of teams did not display the same bias, as they focused more on the team itself as a whole, rather than individuals, for both

positive and negative outcomes. Based on this literature, I focused on the halo effect of attractiveness and its potential influence on a group's rated or assumed task performance.

Group Diversity and Performance

In addition to physical attractiveness, I examined diversity for its relation to perceived group performance. The literature concerning the effects of diverse work groups provides evidence of both positive and negative effects of group diversity on performance (van Dijk et al., 2012), and thus it may be of use to further research this topic. According to Cox and Blake (1991), there is a competitive advantage that organizations gain when hiring from a diverse workforce and effectively managing diverse teams. Those with a more culturally diverse workforce may increase their creativity, problem solving, flexible adaptation to change, and innovation. This enhanced creativity can then lead to greater workplace commitment and job satisfaction (Bassett-Jones, 2005).

Further, those in a diverse team can potentially have a more varied knowledge base when generating ideas. The viewpoints of these diverse groups are considered to be more dissimilar compared to those of homogeneous groups (Roberge & van Dick, 2009), and thus can lead to more innovation and creativity. Demographic diversity within groups has also been shown to enhance creativity (Paulus et al., 2015). Additionally, Nijstad and De Dreu (2002) note that in diverse groups, "cognitive stimulation may facilitate idea generation to a greater extent than in homogeneous groups" (p. 404) as homogenous groups "do not bring much unique knowledge" (p. 403) and they may have perspectives no different than anyone else in the group. In a similar line of thought but through an applied lens, a study by Quintana-Garcia et al.

(2022) found that both gender and ethnic diversity had a positive effect on creativity and innovation within US companies.

After having both homogenous and diverse groups work together to perform tasks across several time periods, Watson et al. (1993) found that while homogenous groups performed better at the start, as the sessions continued and time went on, both homogenous and diverse groups performed at approximately the same level. Similarly, and to show this inconsistency, Ayub and Jehn (2014) asked participants to imagine working with either a homogenous or diverse group, then were given a survey on group performance. Researchers found that diversity aided performance in some ways (by reducing negative attitudes) but not in others (i.e., when there was high national separation, there was more conflict and decreased performance).

This study further investigates group diversity as well as the attractiveness heuristic in relation to perceived creative performance. Will participants' expectations be that diverse or attractive groups will be better in their creative performance? Although there are decades worth of data studying the effects of different forms of diversity on actual performance, this study focused on the assumptions or expectations of the participants in their evaluations of observed group performance, and is thus largely exploratory in nature. The study aimed to understand the influence of overall group physical attractiveness as well as the diversity of a group's composition and how they relate to perceived or assumed performance and creativity ratings.

Overview and Hypotheses

In this study, participants were presented with the creative performance of a group whose overall attractiveness and diversity were experimentally manipulated. The actual performance of the groups did not differ between experimental conditions, and so any perceived differences in their performance is due to differences across experimental conditions.

Hypothesis 1: There will be a main effect of group attractiveness, such that participants will rate the creative performance of the attractive groups as higher compared to that of the unattractive groups. This prediction is an empirical extension of the individual-level attractiveness heuristic research, which has shown that more attractive individuals are assumed to have superiority on a variety of personality traits and social skills because of their physical attractiveness.

Hypothesis 2: There will be a main effect of diversity, such that participants will rate the creative performance of the diverse groups as higher compared to that of the homogenous groups. This prediction takes the concepts of diversity and group performance research, which has shown that culturally diverse teams may increase creativity and problem solving, and focuses on the assumptions of the participants in an exploratory sense.

Hypothesis 3: There will be an interaction between attractiveness and diversity such that the simple effect of diversity will be greater when the group is attractive than when the group is unattractive. This prediction is based on the reasoning for my diversity and halo effect hypotheses. Specifically, when groups are unattractive, their performance should be perceived

as lower regardless of diversity. However, when they are attractive, they will benefit from both the attractiveness heuristic as well as the lay expectations concerning group diversity.

Exploratory Covariates

Based off the recommendation from my committee, I added in the Nakui et al. (2011) Attitudes Toward Diverse Workgroups Scale as a potential moderator. Additionally, participants rated their self-attractiveness as a potential covariate. Lastly, the Ten-Item Personality Inventory (TIPI) was included to explore personality factors as potential covariates.

Method

Participants and Design

Participants were recruited using the University of Texas at Arlington SONA subject pool, an online system for researchers to post opportunities for students to participate in experiments. Undergraduate psychology courses at the University of Texas at Arlington have an option to participate in research studies for a certain number of SONA credits. No financial compensation was given to participants, and all compensation was given in the form of said credits.

Per sample size calculations provided by G*Power (Faul et al., 2009), a sample size of 147 participants would be needed for our study in order to have enough power to detect an effect.

Of the 442 participants who participated in the study, 71 were removed during cleaning if they were tests, or if they started but did not complete the study. The remaining 371 participants' attention checks were coded, and only the participants who had at least one clear match were used, resulting in a final total of 177 participants.

The participants included 177 individuals between the ages of 17 and 57. Specifically, 128 participants were female, 43 were male, four responded other, and two preferred not to respond. Participants responses to ethnicity were coded to be either Caucasian or nonwhite, and included 132 nonwhite and 45 Caucasian participants.

Participants were randomly assigned to one of four conditions in a 2 (attractiveness: attractive/unattractive) x 2 (diversity: diverse/homogenous) between-subjects factorial design. After cleaning, there were 37 participants in the attractive homogenous condition, 51 participants in the attractive diverse condition, 45 participants in the unattractive homogenous condition, and 44 participants in the unattractive diverse condition.

Procedure

Participants who self-selected to participate in the study were directed to an informed consent document via QuestionPro, an online survey tool used by the University of Texas at Arlington for research purposes, where they agreed to continue and participate with the experiment, and they were informed that all of their private information will remain confidential.

The entirety of the research study was conducted through QuestionPro. Following the completed informed consent document, participants were then directed to the beginning of the study. Participants were told that a previous group that participated in a psychology study was told to generate ideas for a new sport, and that they would be viewing that group's interactions and rating their performance. They were then shown the faces of the group members who would appear in the idea generation session and were told to pay close attention to both the members and their ideas, as they would be asked about the details

toward the end of the study. Appendix A displays the idea generation session that was shown to participants.

The actual group performance (i.e., the ideas generated in the discussion) was kept constant for all participants across all conditions, and the experimental conditions only differed in the attractiveness of the group members' photo avatars, and the ethnic/racial diversity of the group itself. Half of participants saw attractive profile photos next to each group member's responses, and the other half saw unattractive profile photos. Within these two attractive or unattractive conditions, half were diverse and half homogenous (i.e., white). Appendix B displays the images that were used in the study.

The images used are from the Chicago FACES database, a catalog of images of the faces of both males and females, from a variety of ethnicities. This database was chosen because of the given ratings associated with each image. In particular, each photo came with an attractiveness rating as well as several different images depicting different facial expressions per face. In addition, a pilot study was run to make sure the images used would have similar ratings with a University of Texas at Arlington sample. See Table 1 for the attractiveness means and standard deviations from the Chicago FACES database and a pilot study that was conducted.[1]

[1] Data for this pilot study was gathered via QuestionPro. The sample consisted of 32 individuals affiliated with the University of Texas at Arlington, undergraduate and graduate students, and previous and current researchers. A significant difference was found between attractive and unattractive facial images, and a non-significant difference was found between the homogenous and diverse images.

Table 1

Means and Standard Deviations for Study Images

Image number	Chicago FACES *M*	Pilot *M*	Pilot *SD*
BF002	4.89	4.94	1.22
BM043	4.85	5.16	1.69
WF003	4.89	4.13	1.34
WF022	5.09	5.00	1.30
WM004	4.66	4.56	1.44
WM029	4.59	4.19	1.38
BF007	1.77	2.44	0.95
BF038	1.89	2.63	1.19
WF002	1.61	2.16	1.08
WF010	2.01	2.22	1.21
WM014	3.48	3.22	1.24
BM010	3.46	3.31	1.42

Note. Attractiveness means and standard deviations from the Chicago FACES database and the pilot study.

Once the participants read through the group's idea generation session for creating a new sport, they were directed toward a questionnaire in which they were asked a series of questions asking them to rate the performance of the group as a whole. Specifically, they rated the performance and creativity of the ideas generated (see Appendix C for the scales that were

used). After viewing the same set of profile photos from the group members, they rated the cohesiveness, likability, cooperativeness, attractiveness, and diversity of the group. Lastly, as an attention check, participants were asked to write in two to three ideas that they saw during the group's creative performance.

After they answered all of the performance rating questions in the survey, they then were directed to a demographic and attitudes section in which they were asked their age, gender identity, and ethnicity. Additionally, participants filled out the Ten-Item Personality Inventory (TIPI) as a quick measure of the Big Five Personality factors. As a potential exploratory moderator variable, participants completed the Nakui et al. (2011) Attitudes Toward Diverse Workgroups scale. Lastly, they were directed to the final questions in which they rated their own level of attractiveness, as well as the attractiveness of all of the images used throughout the study. This was the final section of the study, and participants were then directed to a debriefing and thank you message upon completion (see Appendix D).

Materials

SONA system

As previously mentioned, the University of Texas at Arlington has an online system for participants to sign up for experimental studies that researchers post. Upon completion of the study, participants were compensated through the SONA system in the form of credits to use in their class as research hour requirements. A majority of undergraduate psychology courses require participation in research, and these credit hours satisfied part of those requirements.

Chicago FACES database

The Chicago FACES database that was previously mentioned is a collection of images of faces, both male and female, from the age range of 17-65 years old, and includes a variety of ethnicities. This database was chosen specifically because of its attractiveness ratings on each image. Raters from the Chicago FACES database were asked either "now, consider the person pictured above and rate him/her with respect to other people of the same race and gender. (For example, if you indicated that the person was Asian and male, consider this person on the following traits relative to other Asian males in the United States.) - Attractive (1-7 Likert, 1 = Not at all; 7 = Extremely)" or "what is your first impression of the person pictured above? - Attractive (1-7 Likert, 1 = Not at all; 4 = Neutral; 7 = Extremely)" depending on the subset of images in the database.

Dependent variables

Perceived group performance was assessed using several questions. Participants were asked how cohesive, likable, and cooperative the group was. Additionally, they were asked to rate the group's performance of their task. All questions on the survey used Likert scales of either 1 (not at all) to 7 (very much) or 1 (poor work) to 5 (exceptional work), and the questions were presented immediately after they read through the chat session. Participants were unable to move back to previously asked questions. See Appendix C for a more detailed view of the questionnaire.

Attention and Manipulation Checks

An attention check was included in which participants were asked to write in two to three ideas that they saw during the group's performance. The purpose of this is to check if

they recall any of the ideas that were generated during the session that they read. These attention checks were coded by trained research assistants who were familiar with the idea generation session. They coded each attention check where 0 = nothing, or no familiar ideas, 1 = they were close, or kind of right, and 2 = there was at least one clear match. The attention checks were coded individually at first, and then each research assistant was paired with another, and they discussed the ratings that didn't match their partners' and came to an agreed rating. Additionally, a manipulation check was presented in which the participants rated the attractiveness of the images and the diversity of the group on 7-point Likert scales.

Results

Two independent samples t-tests were run to determine if the manipulations worked. The participants' attractiveness ratings were significantly higher for the attractive groups (M = 4.33, SD = 1.40) compared to the unattractive groups (M = 3.08, SD = 1.72), $t(175) = -5.31$, $p < .001$. The participants' diversity ratings were significantly higher for the diverse groups (M = 5.14, SD = 1.21) compared to the homogenous groups (M = 2.35, SD = 1.56), $t(175) = -13.37$, $p < .001$.

Table 2

Correlation Matrix of All Variables

Correlations

		creative rating	work well rating	competent rating	overall performance rating	cohesive rating	likable rating	cooperative rating	self attractiveness	average adws	avgdv2
creative rating	Pearson Correlation	--									
	N	177									
work well rating	Pearson Correlation	.312**	--								
	Sig. (2-tailed)	<.001									
	N	177	177								
competent rating	Pearson Correlation	.270**	.734**	--							
	Sig. (2-tailed)	<.001	<.001								
	N	177	177	177							
overall performance rating	Pearson Correlation	.594**	.473**	.461**	--						
	Sig. (2-tailed)	<.001	<.001	<.001							
	N	177	177	177	177						
cohesive rating	Pearson Correlation	.333**	.563**	.465**	.601**	--					
	Sig. (2-tailed)	<.001	<.001	<.001	<.001						
	N	177	177	177	177	177					
likable rating	Pearson Correlation	.316**	.454**	.437**	.451**	.505**	--				
	Sig. (2-tailed)	<.001	<.001	<.001	<.001	<.001					
	N	177	177	177	177	177	177				
cooperative rating	Pearson Correlation	.287**	.596**	.451**	.546**	.676**	.602**	--			
	Sig. (2-tailed)	<.001	<.001	<.001	<.001	<.001	<.001				
	N	177	177	177	177	177	177	177			
self attractiveness	Pearson Correlation	.203**	.214**	.210**	.173*	.142	.115	.175*	--		
	Sig. (2-tailed)	.007	.004	.005	.021	.060	.126	.019			
	N	177	177	177	177	177	177	177	177		
average adws	Pearson Correlation	.223**	.039	.203**	.216**	.064	.202**	.139	.082	--	
	Sig. (2-tailed)	.003	.604	.007	.004	.397	.007	.066	.279		
	N	177	177	177	177	177	177	177	177	177	
avgdv2	Pearson Correlation	.537**	.837**	.770**	.691**	.794**	.725**	.804**	.237**	.187*	--
	Sig. (2-tailed)	<.001	<.001	<.001	<.001	<.001	<.001	<.001	.001	.013	
	N	177	177	177	177	177	177	177	177	177	177

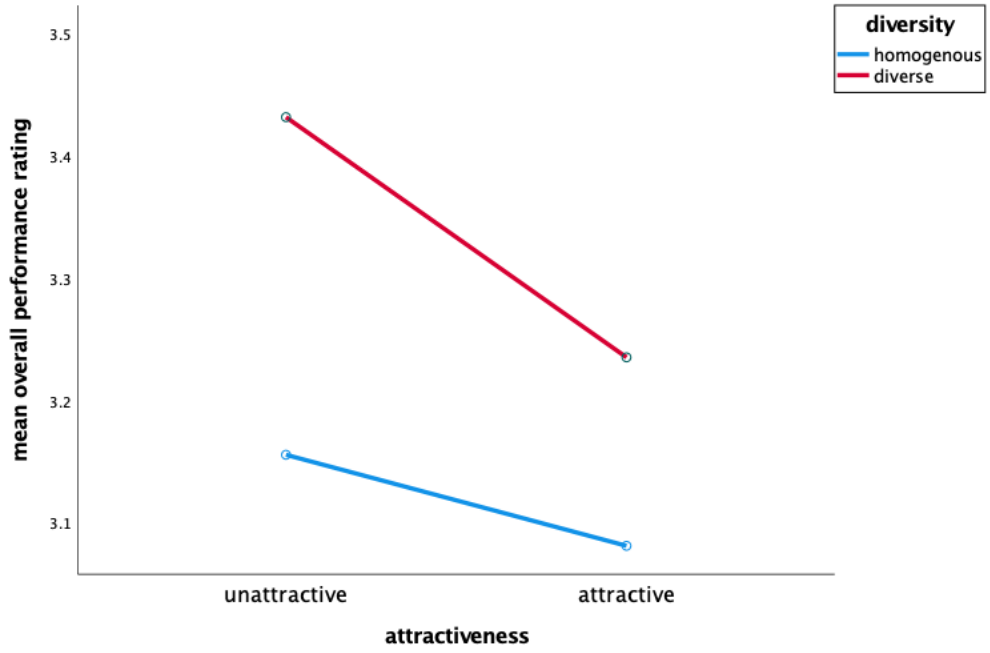
**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

Table 2 presents a correlation matrix for key study dependent variables. To test the hypotheses that there were main effects of attractiveness, diversity, and an interaction between the two, a 2 x 2 factorial analysis of variance (ANOVA) was conducted. The effects of the two independent variables, attractiveness (attractive/unattractive) and diversity (diverse/homogenous), were tested on the dependent variable of perceived overall group performance ratings. Differences in overall group performance by attractiveness were not significant, $F(1, 173) = .91$, $p = .342$. Thus, Hypothesis 1 was not supported. Differences in overall group performance by diversity were also not significant, $F(1, 173) = 2.29$, $p = .132$. Hypothesis 2 was not supported. There was not a statistically significant interaction between the effects of attractiveness and diversity, $F(1, 173) = .18$, $p = .668$. Hypothesis 3 was also not

supported. See Figure 1 for a visual representation of the means. Although diverse groups were generally rated higher in perceived overall performance when compared to homogenous groups, nothing was significant.

Figure 1
Means for Overall Performance Ratings



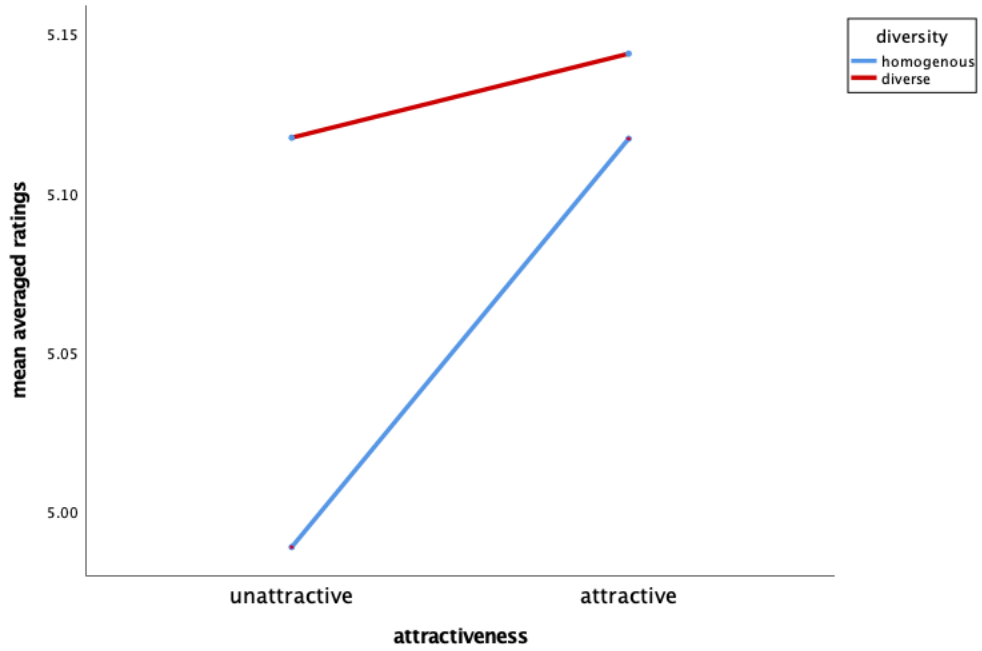
Exploratory Analyses

After obtaining high reliability (α = .86) for all of the dependent variables (apart from the overall performance item), the five dependent variables were averaged together into one variable to further analyze. A 2 (attractiveness: attractive/unattractive) x 2 (diversity: diverse/homogenous) factorial ANOVA was conducted with the new average dependent variable. Although visually it seems like there is an effect (see Figure 2), differences in the averaged ratings by attractiveness were not significant, $F(1, 173) = .24$, $p = .624$. Further, neither the averaged rating by diversity, $F(1, 173) = .24$, $p = .623$, nor the interaction between

attractiveness and diversity, $F(1, 173) = .11$, $p = .747$, were significant. As the averaged rating wasn't significant, I explored various individual dependent variables below.

Figure 2

Means for Averaged Ratings



Attractiveness, Diversity, and Ethnicity with Dependent Variables

To explore any effect of the participants' ethnicity as well as attractiveness and diversity of the group, a three-way between-subjects analysis of variance (ANOVA) was conducted to analyze the effects of attractiveness (attractive/unattractive), diversity (diverse/homogenous) and ethnicity (Caucasian/nonwhite[2]) on the group's perceived creativity, how well they worked together, their competence, overall performance, cohesiveness, likability, and cooperativeness.

[2] Ethnicity was dichotomized here into nonwhite and white for ease of comparison, as there were quite a few ethnicity options on the demographic question (Hispanic or Latino, American Indian or Alaska Native, Asian, Black or African American, Native Hawaiian or Other Pacific Islander, Caucasian or White, Multiracial, Other, or Prefer not to respond). Of the 177 participants, 45 were Caucasian or White, and 132 were Nonwhite.

There was a significant three-way interaction between attractiveness, diversity, and ethnicity on how well the group worked together, $F(1, 169) = 6.03$, $p = .015$ (see Figures 3 and 4). For White participants, unattractive homogenous groups are thought to work best together, and attractive homogenous groups were rated lowest. For Nonwhite participants, attractive homogenous groups are thought to work best together, and unattractive homogenous groups were rated lowest. Simple main effects analysis showed that work well ratings for the unattractive homogenous group were significantly higher for White participants than for Nonwhite participants, $F(1, 169) = 5.82$, $p = .017$. There were no other significant simple main effects.

Figure 3

Means for Work Well Rating for Whites

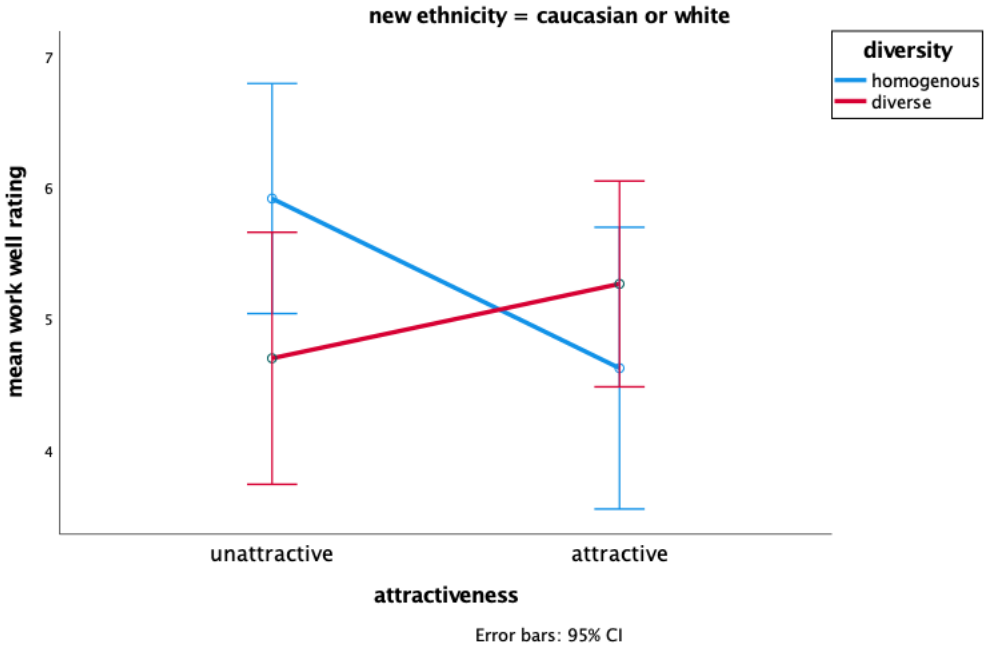
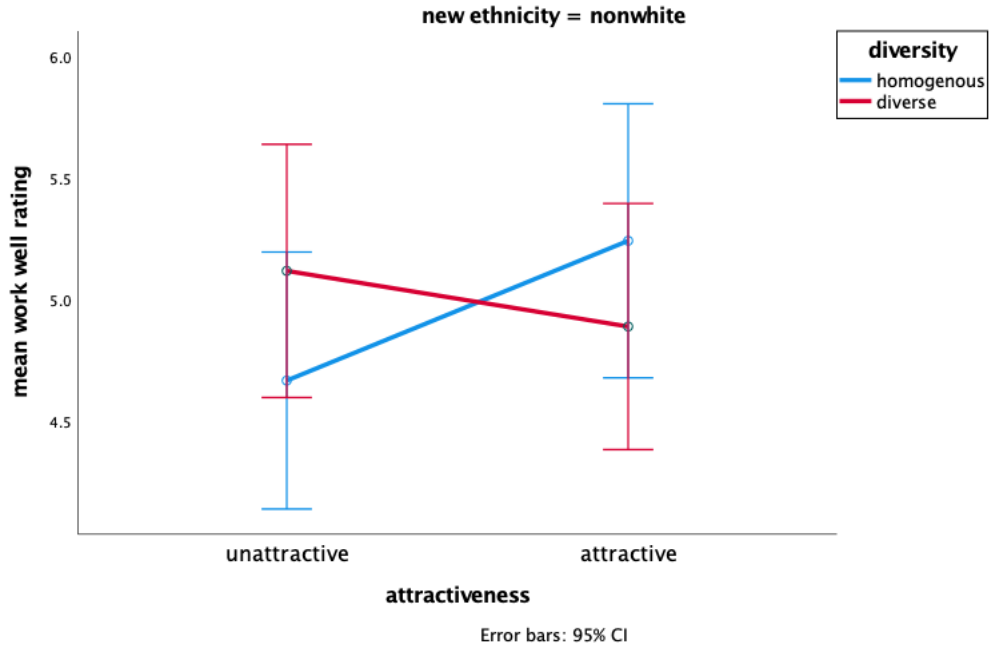


Figure 4

Means for Work Well Rating for Nonwhites



There was also a significant three-way interaction between attractiveness, diversity, and ethnicity on competence, $F(1, 169) = 3.94$, $p = .049$ (see Figures 5 and 6). For White participants, unattractive homogenous groups were thought to be more competent, and attractive homogenous groups were rated lowest. For Nonwhite participants, unattractive diverse groups were thought to be the most competent, and unattractive homogenous groups were rated lowest. Simple main effects analysis showed that competence ratings for the unattractive homogenous group were significantly higher for White participants than for Nonwhite participants, $F(1, 169) = 6.77$, $p = .010$. Additionally, within the competence ratings from Nonwhite participants, the unattractive homogenous group was rated significantly higher than the unattractive diverse group, $F(1, 169) = 4.85$, $p = .029$. There were no other significant simple main effects.

Figure 5

Means for Competence Rating for Whites

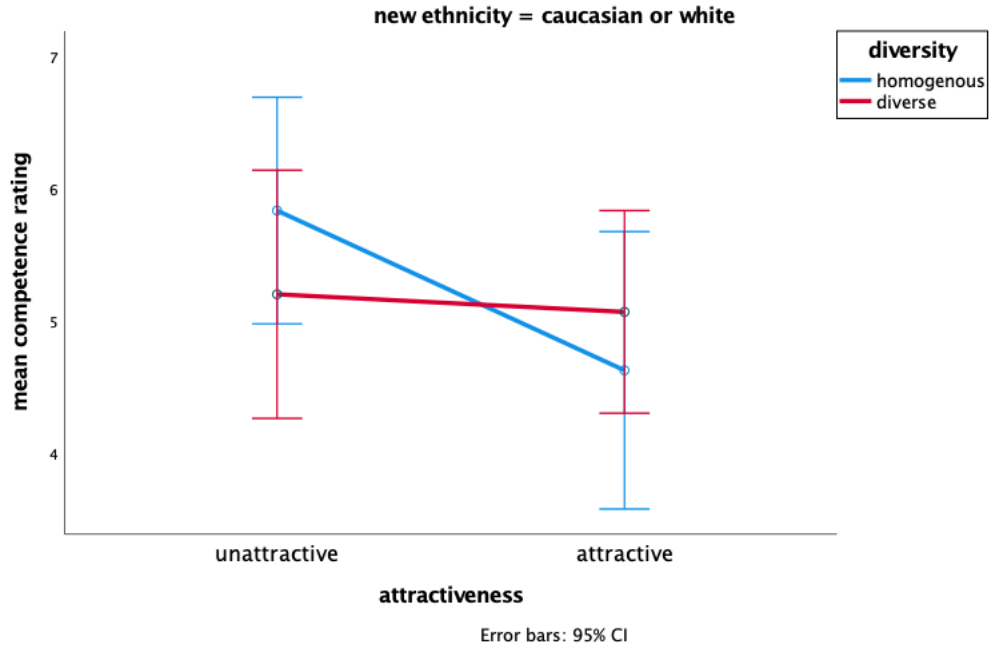
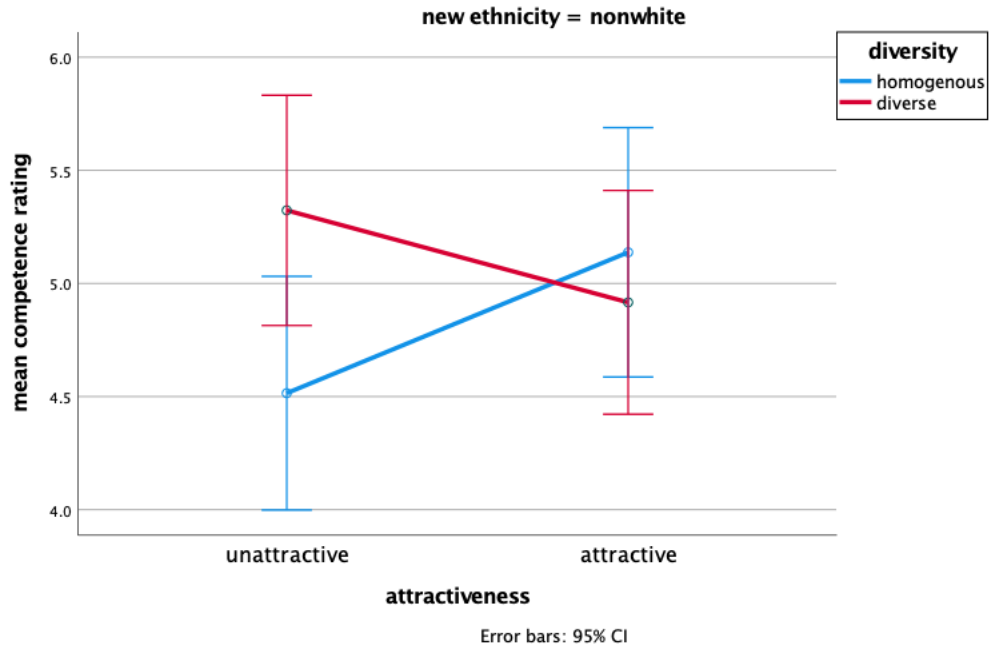


Figure 6

Means for Competence Rating for Nonwhites



Lastly, there was a significant three-way interaction between attractiveness, diversity, and ethnicity on the group's overall performance rating, $F(1, 169) = 4.05$, $p = .046$ (see Figures 7 and 8). For White participants, unattractive homogenous groups were thought to perform the best overall, and attractive homogenous groups were rated lowest. For Nonwhite participants, both of the diverse groups had a greater overall performance rating than both of the homogenous groups.[3] Simple main effects analysis showed that overall performance ratings for the unattractive homogenous group were significantly higher for White participants than for Nonwhite participants, $F(1, 169) = 6.69$, $p = .011$. Additionally, within the overall performance ratings from Nonwhite participants, the unattractive homogenous group was rated significantly higher than the unattractive diverse group, $F(1, 169) = 5.47$, $p = .021$. There were no other significant simple main effects.

[3] The same pattern remains when controlling for self-attractiveness, ADWS, and both self-attractiveness and ADWS.

Figure 7

Means for Overall Performance Rating for Whites

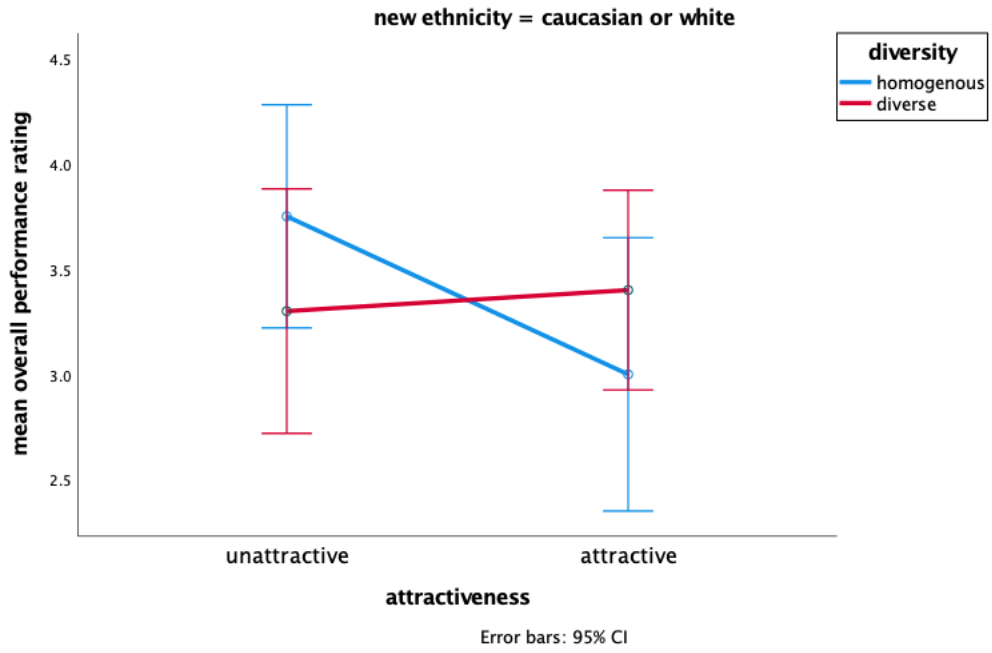
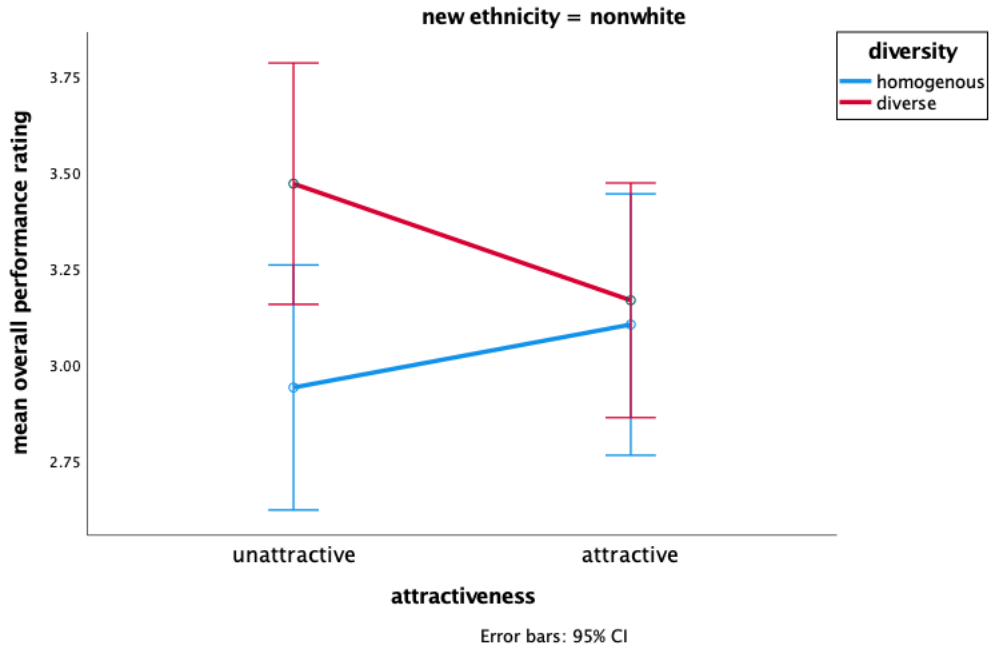


Figure 8

Means for Overall Performance Rating for Nonwhites



ADWS as a Moderator and Self-Attractiveness as a Covariate

To examine ADWS as a potential moderator variable, a median split of the ADWS was performed, and a three-way analysis of covariance (ANCOVA) was conducted while controlling for participants' self-attractiveness ratings. The independent variables were attractiveness, diversity, ADWS (negative/positive). The dependent variables included the participants rating on the group's creativity, how well they worked together, their competence, overall performance, cohesiveness, likability, and cooperativeness.

While there was no significant three-way interaction between attractiveness, diversity, and ADWS on any of the dependent variables, there was a significant interaction between diversity and ADWS on how well the group worked together, while controlling for their self-attractiveness, $F(1, 168) = 5.45$, $p = .021$, partial $\eta^2 = .03$. Those who have a negative attitude toward diverse workgroups perceived the homogenous group ($M = 5.22$, $SD = 1.44$) to work better together than the diverse group ($M = 4.66$, $SD = 1.38$), and that those who have a positive attitude toward diverse workgroups perceived the diverse group ($M = 5.51$, $SD = 1.49$) to work better together than the homogenous group ($M = 4.91$, $SD = 1.77$). Post hoc comparisons using the Bonferroni correction indicated that within the work well rating, those with more positive ADWS rated the diverse groups significantly different than the homogenous groups. Specifically, those with more positive ADWS rated the diverse groups higher ($M = 5.33$, $SD = .189$) in working well together when compared with the homogenous groups (M = 4.78, $SD = .184$).

There was also a significant interaction between diversity and ADWS on cohesiveness, while controlling for their self-attractiveness ratings, $F(1, 168) = 4.99$, $p = .027$, partial $\eta^2 = .03$.

Those with negative attitudes toward diverse workgroups perceived the homogenous group (*M* = 4.81, *SD* = 1.31) to be more cohesive than the diverse group (*M* = 4.46, *SD* = 1.39), and that those with positive attitudes toward diverse workgroups perceived the diverse group (*M* = 5.08, *SD* = 1.18) to be more cohesive than the homogenous group (*M* = 4.42, *SD* = 1.73). Post hoc comparisons using the Bonferroni correction indicated that within the cohesiveness rating, those with more positive ADWS rated the diverse groups significantly different than the homogenous groups. Specifically, those with more positive ADWS rated the diverse groups higher (*M* = 4.93, *SD* = .162) in cohesiveness when compared with the homogenous groups (M = 4.41, *SD* = .158).

A significant interaction between diversity and ADWS on cooperativeness, while controlling for their self-attractiveness was also found, $F(1, 168) = 4.94$, $p = .028$, partial $\eta^2 = .03$. Those with a negative attitude toward diverse workgroups perceived the homogenous group (*M* = 5.30, *SD* = 1.02) to be more cooperative than the diverse group (*M* = 4.93, *SD* = 1.43), and that those with a positive attitude toward diverse workgroups perceived the diverse group (*M* = 5.87, *SD* = 1.17) to be more cooperative than the homogenous group (*M* = 5.31, *SD* = 1.43). Post hoc comparisons using the Bonferroni correction indicated that within the cooperativeness rating, those with more positive ADWS rated the diverse groups significantly different than the homogenous groups. Specifically, those with more positive ADWS rated the diverse groups higher (*M* = 5.50, *SD* = .156) in cooperativeness when compared with the homogenous groups (M = 5.13, *SD* = .152). Although the group's creativity and likability were trending toward significance, there were no statistically significant interactions between

diversity and ADWS competence or overall performance while controlling for their self-attractiveness ratings.

Discussion

While decades worth of research has been done on the physical attractiveness halo effect as well as group diversity, the novelty of this study lies in examining the halo effect at the group level rather than the individual level as well as focusing on the assumed relationship between diversity and performance. Hypothesis one, that there would be a main effect of group attractiveness, such that participants would rate the creative performance of the attractive groups as higher compared to that of the unattractive groups, was not supported. Hypothesis two, that there would be a main effect of diversity, such that participants would rate the creative performance of the diverse groups as higher compared to that of the homogenous groups was not supported. Lastly, hypothesis three, that there would be an interaction between attractiveness and diversity such that the simple effect of diversity would be greater when the group was attractive than when the group was unattractive was also not supported. While this was initially disheartening, as I stated earlier, there are inconsistencies within the research. Specifically, there are both positive and negative effects observed of group diversity on performance (van Dijk et al., 2012), and task conflict was shown to be positively correlated with cultural diversity (Stahl et al., 2010), which may have led to the participants not rating perceived creative performance of the diverse groups as higher as I originally predicted.

Although there were no significant results within the initial analyses regarding group attractiveness, further exploratory analyses revealed that on its own, and after controlling for self-attractiveness, ADWS, or both, there are significant differences within how participants

rated how well the group worked together. Particularly, White participants rated the unattractive homogenous group highest in all exploratory analyses within the work well rating. When controlling for the self-attractiveness, ADWS, and both, White participants rated the unattractive homogenous group highest in perceived ability to work well together. In only the first exploratory analyses mentioned (the three-way ANOVA with attractiveness x diversity x ethnicity) did other dependent variables have any significance. Specifically, there was a significant difference in how participants rated the perceived competence and overall performance of the group. White participants rated the unattractive homogenous group highest in competence, and Nonwhite participants rated the unattractive diverse group highest in competence. Interestingly, for Nonwhite participants, both diverse groups had a greater perceived overall performance rating than both homogenous groups, although again, White participants rated the unattractive homogenous group highest in perceived overall performance. Moreover, there is a pattern throughout in which nonwhite participants rated the attractive homogenous group highest in the work well condition. Perhaps because the mean age of participants is 21, they grew up watching television shows that didn't portray as many nonwhite characters as the media does today. It comes as no surprise that media and mass communication influence one's expectations about others (Bandura, 2002), and historically, the media has catered mostly toward White audiences (Tukachinski, 2015). Although there has been a recent surge in the recognition of ethnic minorities in the entertainment industry (Borum Chatoo, 2018), there is still room for growth societally to aid in more promising attitudes towards groups. Caution must also be used in interpreting the findings from the white participants alone, as the sample size (and thus number of participants per cell) is very low.

Additionally, those with a negative attitude toward diverse groups tend to perceive homogenous groups as performing better. Specifically, it was found that those who have a negative attitude toward diverse groups perceive homogenous groups to work better together, be more cohesive, and cooperative with one another when compared to diverse groups. Similarly, those who have a positive attitude toward diverse groups perceive diverse groups to work better together, be more cohesive, and cooperative with one another when compared to homogenous groups. Perhaps the Similarity Attraction Theory, which essentially states that people are more attracted to those who are similar to themselves (Byrne, 1971) could help explain this finding, as participants may have distinguished themselves from others and were more inclined to perceive their ingroup to perform better.

The majority of attractiveness halo effect research was conducted in the 1970s and 1980s, and this study is a more modern approach to that research. Thinking in terms of the college aged population at that time, it's likely that a majority of the studies had a sample of white males. Dion et al.'s (1972) What is Beautiful is Good study described participants only by their gender and number. With this population in mind, those who were included in the study most likely didn't have the most positive attitudes toward diverse groups, as they were likely white males in the United States in the 1970's who were living in a society that didn't require them to. The current study mirrored these sentiments and themes that were present in the research of that time through its outcomes of greater perceived performance of homogenous groups when participants had negative attitudes toward diverse groups. Explicitly, when broken down into negative and positive attitudes toward diverse groups, and when comparing those

with negative attitudes to the previous research's sample, it is not surprising that the results are similar for comparable groups of people.

Limitations

In terms of limitations, it was noted after the study began that the photos chosen and used from the Chicago FACES database, particularly the unattractive photos, are all overweight individuals. This may have influenced the results of the unattractive groups, as participants may have had a weight bias that was not controlled for. If this study were replicated, it would be beneficial to either add in photos of unattractive people that aren't overweight, or to measure participants attitudes toward weight to control for potential weight bias.

Regarding the photos used, they all looked the same—the subjects were all wearing the same gray t-shirts, and all photos had the same white background. Although they were told that these were participants from a previous research study, this may have been a cue to participants of our deception. In the future, it may be appropriate to use different databases to gather a variety of different images to use.

Additionally, the study was completely online, and with that comes the inability to control the environment in which the participants are taking the study. After cleaning the data, I found quite a bit of dropouts, and with the spring semester of data collection being exceptionally unfortunate, it's easy to infer that an online study such as this one struggled to keep the participants' attention. For better quality data, it may be useful to bring participants into the lab to take the study to better control their environment.

Implications and Future Directions

It is important to note that our variables are a few facets of many that relate to perceived group performance, and there are other directions for future studies. Specifically, there may be other factors that relate to the evaluation of group performance in an online study versus a lab study in which one can better control the environment. Additionally, as these are only a few measures of creative performance in an academic research setting, and it is important to note that there are still other areas of research to consider, such as differing the group diversity—not just ethnic diversity, but perhaps those in different socio-economic status groups, disability status, or age.

Results from this study may lead to a greater understanding of the creative performance of diverse or homogenous groups, which has a very applied value to it. Within the realm of group composition, perception of a group's ability to work well together, or its cohesiveness or cooperativeness could lead to earnings boosts or increased wages for team members. Additionally, the contribution of this research could be useful in areas such as creativity, talent acquisition, and organizational management. This research may also be helpful for those in leadership positions as they complete performance evaluations for groups on their team. The results from this study could be used as an additional piece of training for any rater who is required to evaluate groups, as it would increase their knowledge of how and when bias occurs and may lead to prevention of that bias. Also, in a talent acquisition setting, it may be useful to have a blind interview when making hiring decisions in order to reduce bias.

Moreover, there may be a need to prevent leaders from accessing the individual identities of group members when evaluating performance, as to not promote a bias

preference for either homogenous or diverse groups. This could be beneficial to minimize the disadvantage of high performing groups that may happen to be either diverse or homogenous. White evaluators should be aware of this bias to rate homogenous groups on overall performance, competence, and how well they work together better than diverse groups. Although ethnic diversity was examined here, it may be of interest in looking further into having teams with mixed levels of attractiveness, or diversity of attractiveness to examine whether this affects the attractiveness halo effect in some way. Additionally, as suggested by the findings of van Dijk et al. (2012), it may be of interest to focus on different dimensions of group diversity for future research, such as skill level or job-related diversity.

References

Adams, M., & Baker, P. L. (2021). Does boardroom nationality affect the performance of UK insurers? *The British Accounting Review*, *53*(5), 100923. https://doi.org/10.1016/j.bar.2020.100923

Ayub, N., & Jehn, K. (2014). When diversity helps performance. *International Journal of Conflict Management*, *25*(2), 189–212. https://doi.org/10.1108/ijcma-04-2013-0023

Bandura, A. (2001). Social Cognitive Theory of Mass Communication. *Media Psychology*, *3*(3), 265–299. https://doi.org/10.1207/s1532785xmep0303_03

Bassett-Jones, N. (2005). The paradox of Diversity Management, creativity and Innovation. *Creativity and Innovation Management*, *14*(2), 169–175. https://doi.org/10.1111/j.1467-8691.00337.x

Borum Chattoo, C. (2018). Oscars so white: Gender, racial, and ethnic diversity and social issues in U.S. documentary films (2008–2017). *Mass Communication and Society*, *21*(3), 368–394. https://doi.org/10.1080/15205436.2017.1409356

Byrne, D. E. (1971). *The Attraction Paradigm*. Academic Press.

Chat simulator - the easy way to fake a chat conversation. (n.d.). Retrieved February 25, 2022, from https://chat-simulator.com/simulator

Cox, T. H., & Blake, S. (1991). Managing cultural diversity: Implications for organizational competitiveness. *Academy of Management Perspectives*, *5*(3), 45–56. https://doi.org/10.5465/ame.1991.4274465

Dion, K., Berscheid, E., & Walster, E. (1972). What is Beautiful is Good. *Journal of Personality and Social Psychology*, *24*(3), 285–290.

Eagly, A. H., Ashmore, R. D., Makhijani, M. G., & Longo, L. C. (1991). What is beautiful is good, but . . .: A meta-analytic review of research on the physical attractiveness stereotype. *Psychological Bulletin*, *110*(1), 109–128. https://doi.org/10.1037/0033-2909.110.1.109

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*, 1149-1160.

Klebl, C., Rhee, J. J., Greenaway, K. H., Luo, Y., & Bastian, B. (2021). Beauty goes down to the core: Attractiveness biases moral character attributions. *Journal of Nonverbal Behavior*, *46*(1), 83–97. https://doi.org/10.1007/s10919-021-00388-w

Lucker, G. W., Beane, W. E., & Helmreich, R. L. (1981). The strength of the halo effect in physical attractiveness research. *The Journal of Psychology*, *107*(1), 69–75. https://doi.org/10.1080/00223980.1981.9915206

Mims, P. R., Hartnett, J. J., & Nay, W. R. (1975). Interpersonal attraction and help volunteering as a function of physical attractiveness. *The Journal of Psychology*, *89*(1), 125–131. https://doi.org/10.1080/00223980.1975.9923913

Nakui, T., Paulus, P., & Van Der Zee, K. (2011). The role of attitudes in reactions toward diversity in workgroups. *Journal of Applied Social Psychology*, *41*(10), 2327–2351. https://doi.org/10.1111/j.1559-1816.2011.00818.x

Naquin, C. E., & Tynan, R. O. (2003). The team halo effect: Why teams are not blamed for their failures. *Journal of Applied Psychology*, *88*(2), 332–340. https://doi-org.ezproxy.uta.edu/10.1037/0021-9010.88.2.332

Nijstad, B. A., & De Dreu, C. K. W. (2002). Creativity and Group Innovation. *International Association for Applied Psychology*.

Nisbett, R. E., & Wilson, T. D. (1977). The halo effect: Evidence for unconscious alteration of judgments. *Journal of Personality and Social Psychology*, *35*(4), 250–256. https://doi.org/10.1037/0022-3514.35.4.250

Paulus, P. B., van der Zee, K. I., & Kenworthy, J. (2016). Cultural diversity and team creativity. *The Palgrave Handbook of Creativity and Culture Research*, 57–76. https://doi.org/10.1057/978-1-137-46344-9_4

Quintana-García, C., Marchante-Lara, M., & Benavides-Chicón, C. G. (2022). Boosting innovation through gender and ethnic diversity in management teams. *Journal of Organizational Change Management*, *35*(8), 54–67. https://doi.org/10.1108/jocm-05-2021-0137

Roberge, M.-É., & van Dick, R. (2009). Recognizing the benefits of diversity: When and how does Diversity Increase Group Performance? *Human Resource Management Review*, *20*(4), 295–308. https://doi.org/10.1016/j.hrmr.2009.09.002

Salvia, J., Sheare, J. B., & Algozzine, B. (1975). Facial attractiveness and personal-social development. *Journal of Abnormal Child Psychology*, *3*(3), 171–178. https://doi.org/10.1007/bf00916748

Stahl, G. K., Maznevski, M. L., Voigt, A., & Jonsen, K. (2009). Unraveling the effects of cultural diversity in teams: A meta-analysis of research on multicultural work groups. *Journal of International Business Studies*, *41*(4), 690–709. https://doi.org/10.1057/jibs.2009.85

Thorndike, E. L. (1920). A constant error in psychological ratings. *Journal of Applied Psychology*, *4*(1), 25–29. https://doi.org/10.1037/h0071663

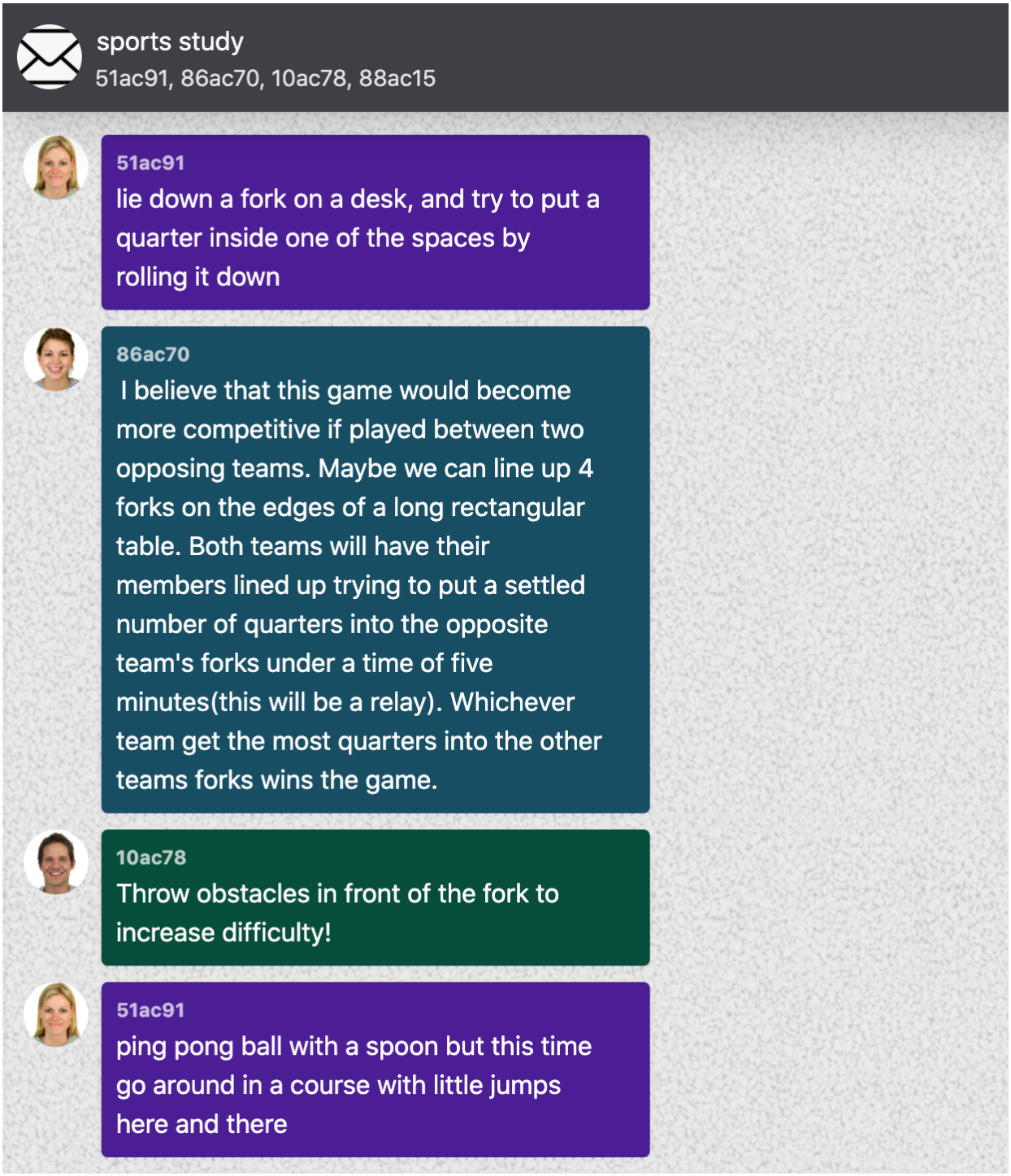
Tukachinsky, R. (2015). Where we have been and where we can go from here: Looking to the future in research on media, race, and ethnicity. *Journal of Social Issues*, *71*(1), 186–199. https://doi.org/10.1111/josi.12104

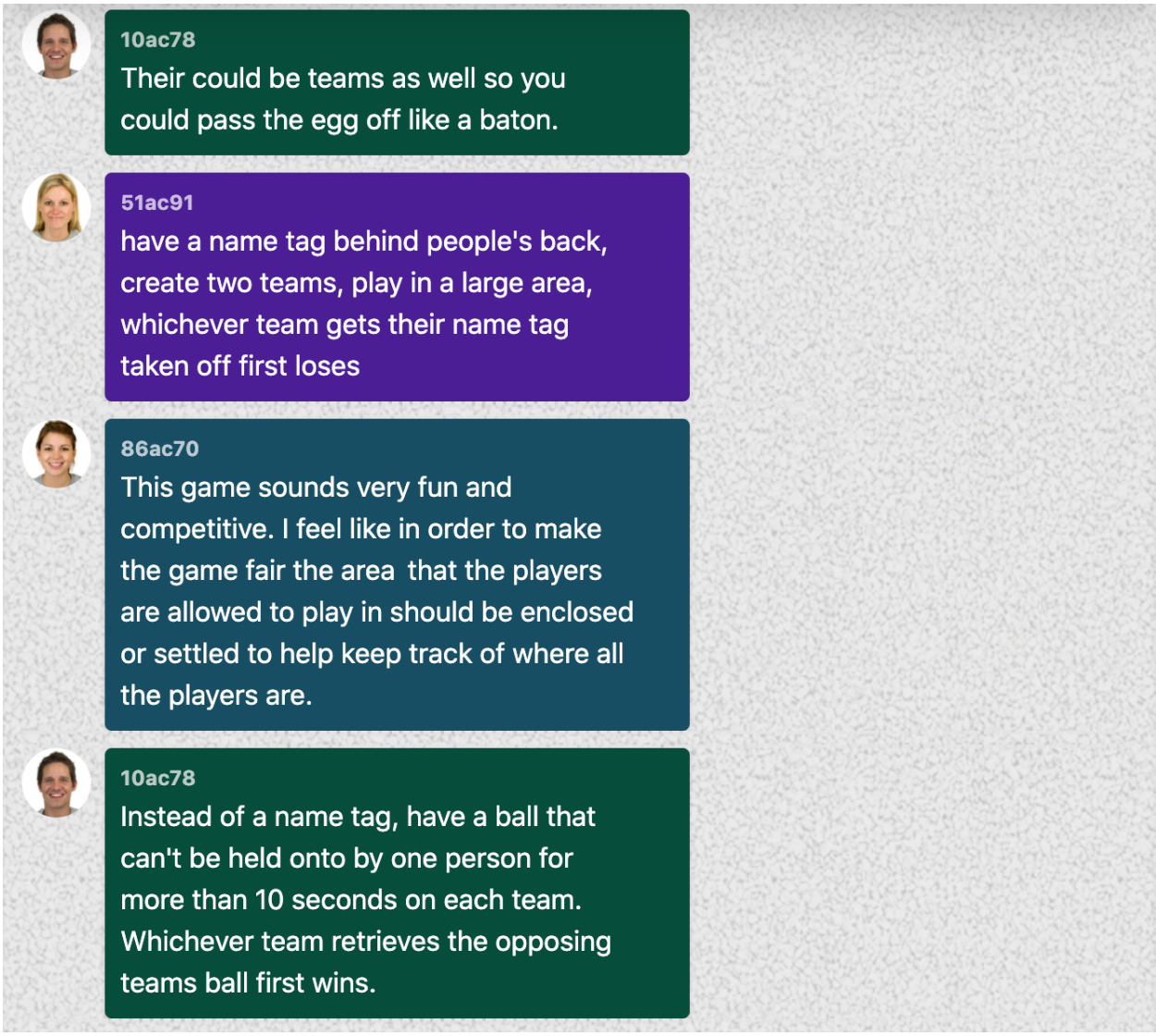
van Dijk, H., van Engen, M. L., & van Knippenberg, D. (2012). Defying conventional wisdom: A meta-analytical examination of the differences between demographic and job-related diversity relationships with performance. *Organizational Behavior and Human Decision Processes*, *119*(1), 38–53. https://doi.org/10.1016/j.obhdp.2012.06.003

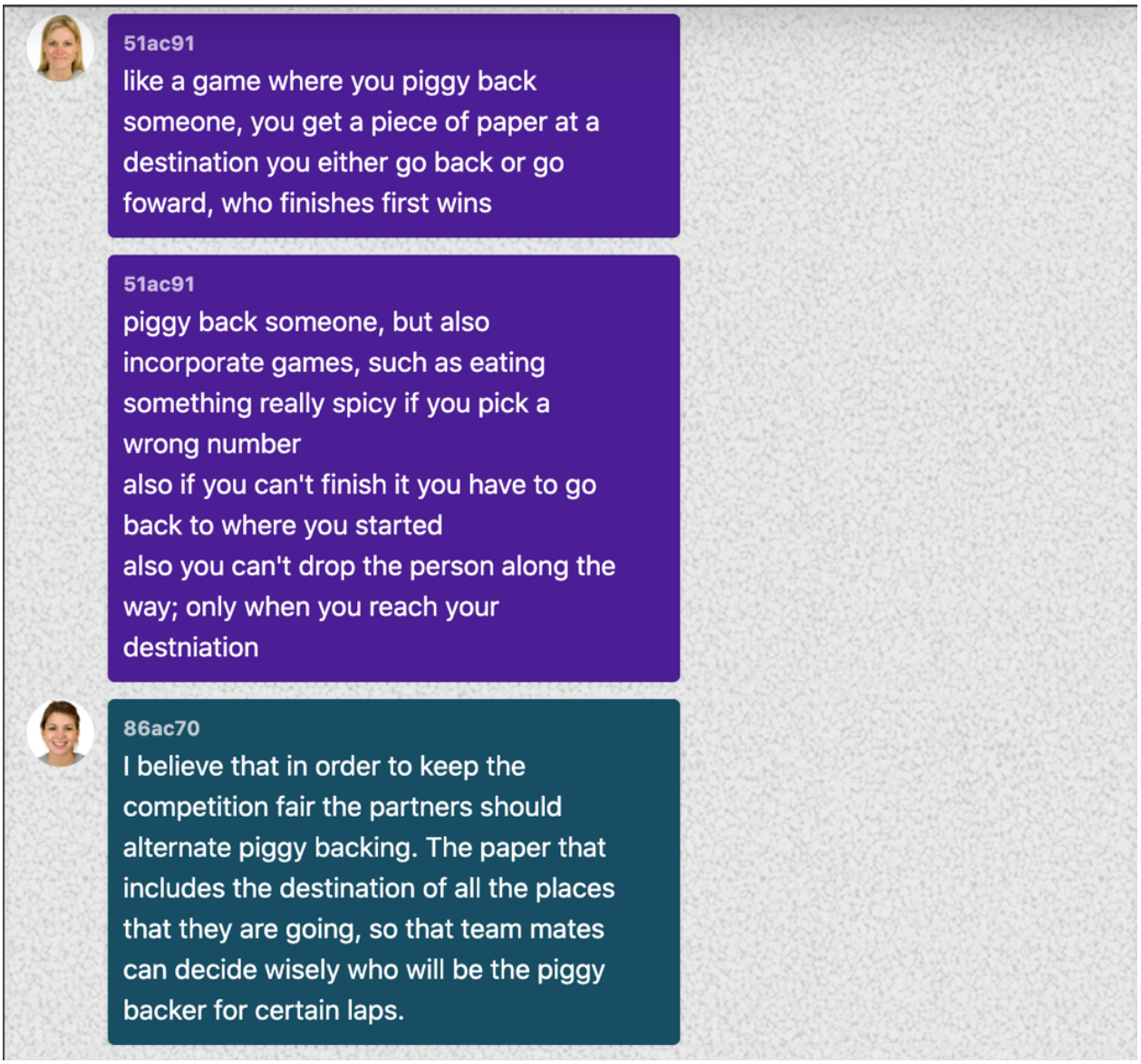
Watson, W. E., Kumar, K., & Michaelsen, L. K. (1993). Cultural diversity's impact on interaction process and performance: Comparing homogeneous and diverse task groups. *Academy of Management Journal*, *36*(3), 590–602. https://doi.org/10.5465/256593

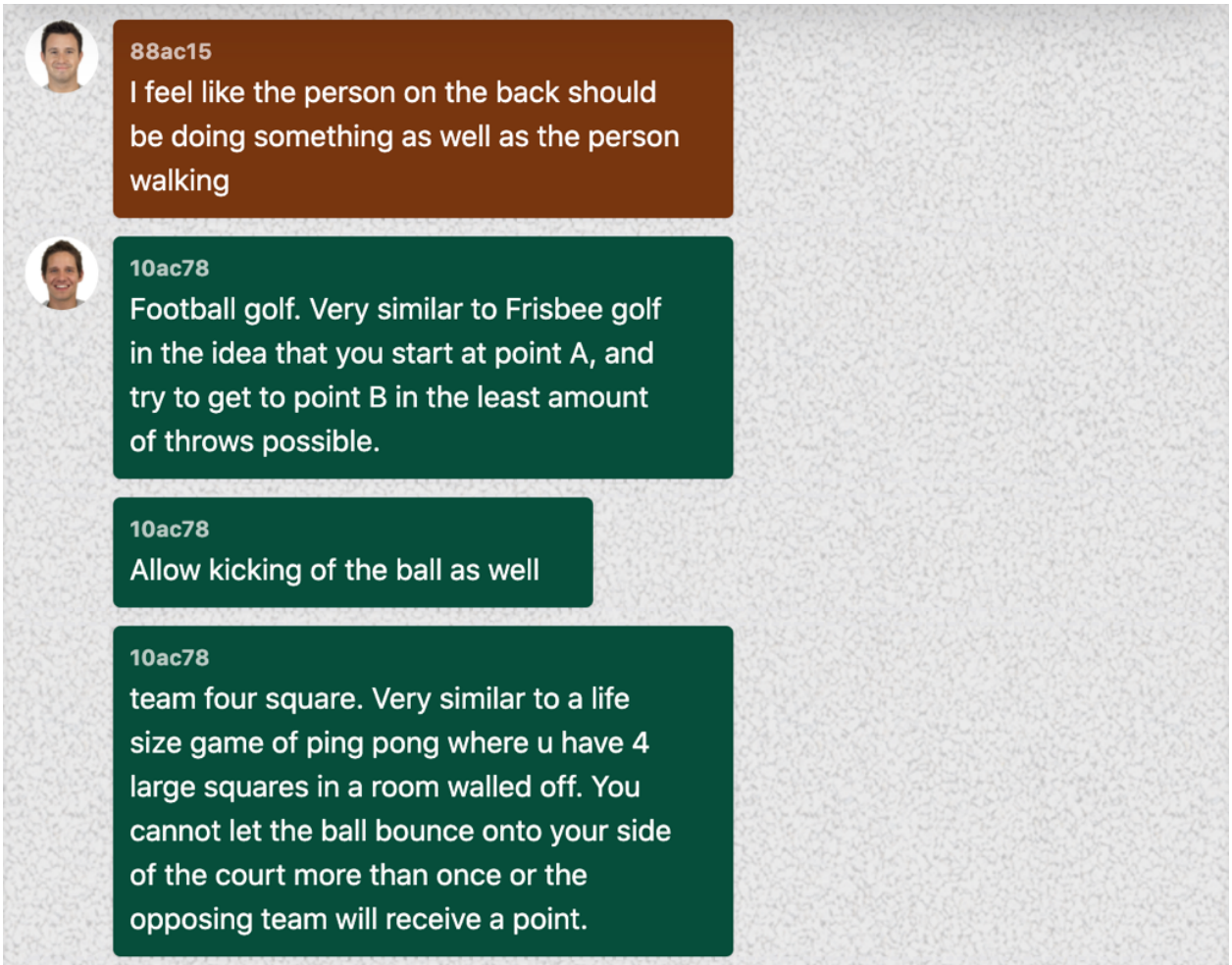
Appendix A

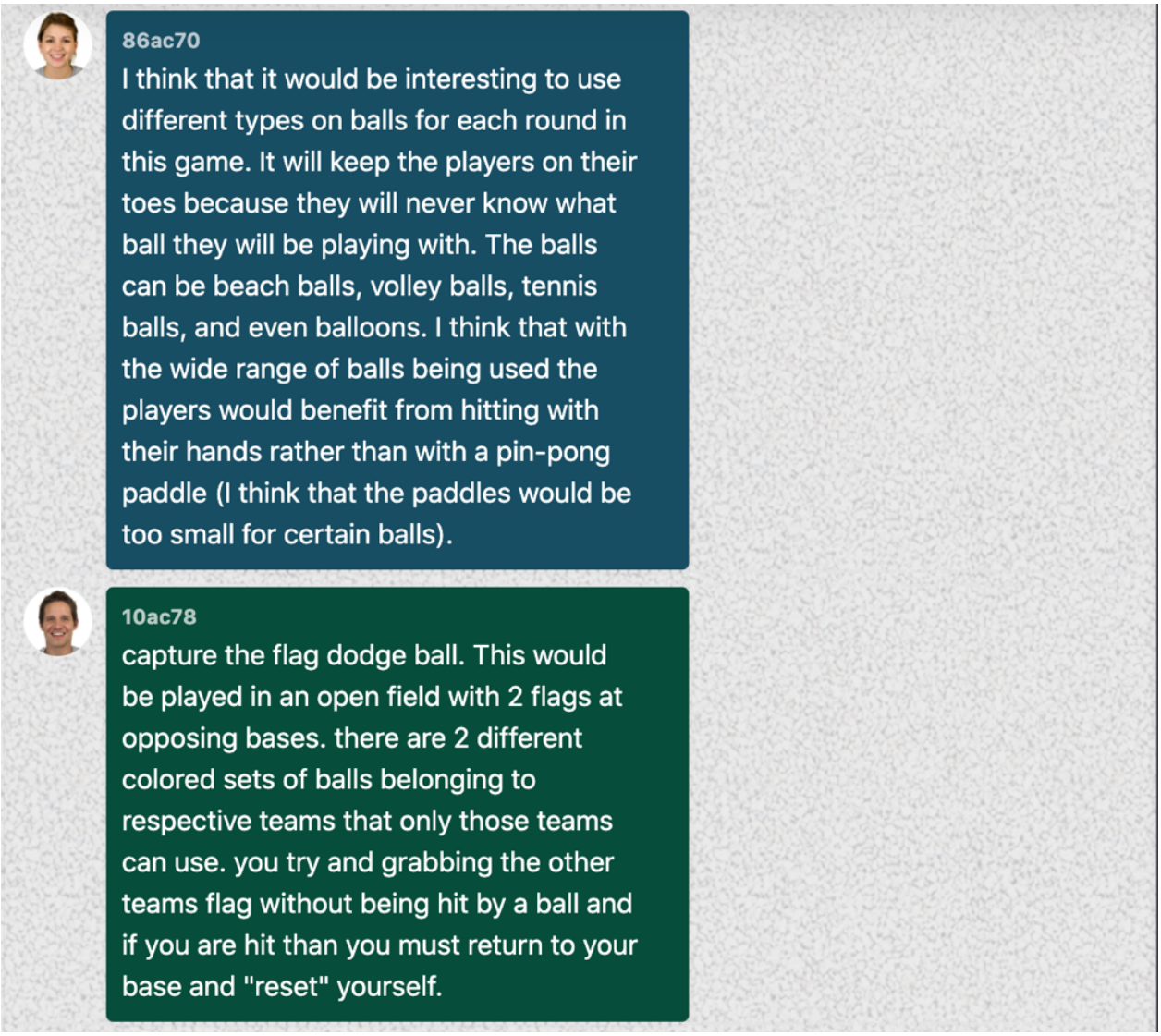
Idea generation session for the attractive homogenous group (note that only the images changed across conditions; the text remained the same).

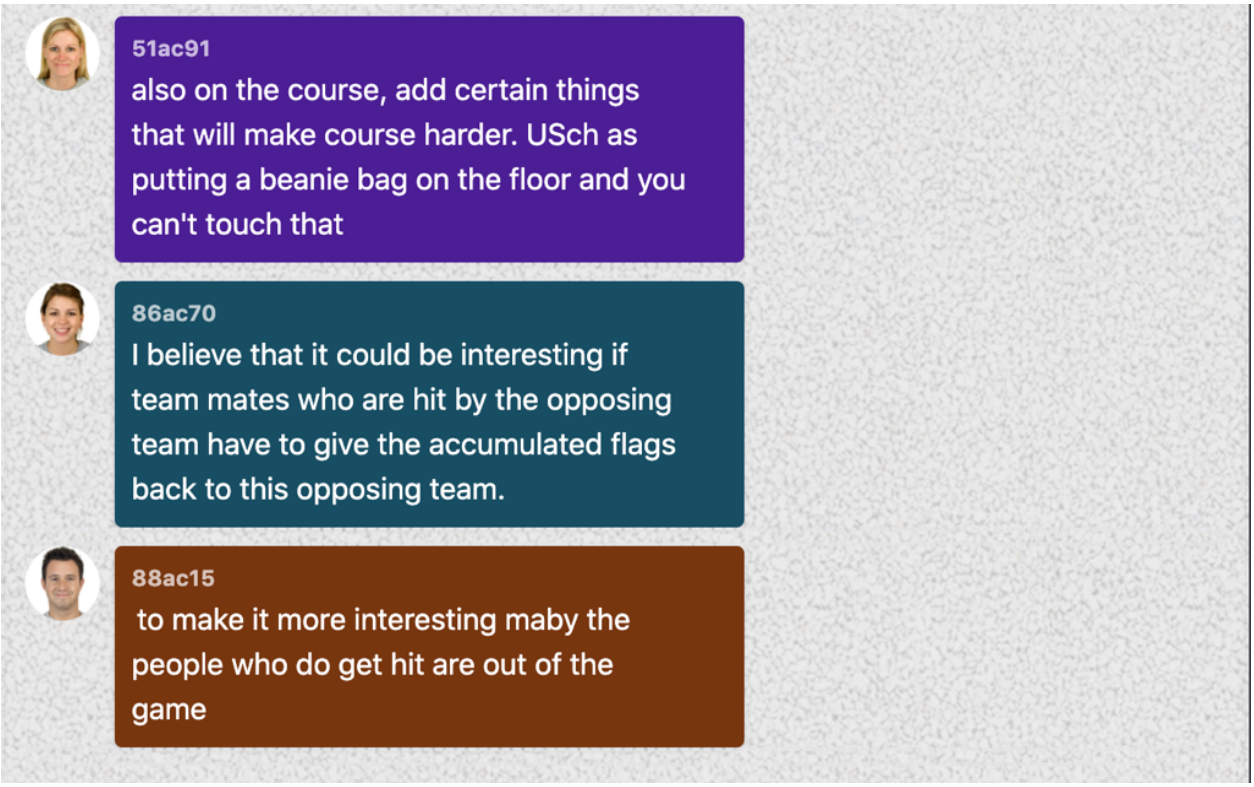



10ac78
Their could be teams as well so you could pass the egg off like a baton.
51ac91
have a name tag behind people's back, create two teams, play in a large area, whichever team gets their name tag taken off first loses
86ac70
This game sounds very fun and competitive. I feel like in order to make the game fair the area that the players are allowed to play in should be enclosed or settled to help keep track of where all the players are.
10ac78
Instead of a name tag, have a ball that can't be held onto by one person for more than 10 seconds on each team. Whichever team retrieves the opposing teams ball first wins.


51ac91
like a game where you piggy back someone, you get a piece of paper at a destination you either go back or go foward, who finishes first wins
51ac91
piggy back someone, but also incorporate games, such as eating something really spicy if you pick a wrong number
also if you can't finish it you have to go back to where you started
also you can't drop the person along the way; only when you reach your destniation
86ac70
I believe that in order to keep the competition fair the partners should alternate piggy backing. The paper that includes the destination of all the places that they are going, so that team mates can decide wisely who will be the piggy backer for certain laps.


88ac15
I feel like the person on the back should be doing something as well as the person walking
10ac78
Football golf. Very similar to Frisbee golf in the idea that you start at point A, and try to get to point B in the least amount of throws possible.
10ac78
Allow kicking of the ball as well
10ac78
team four square. Very similar to a life size game of ping pong where u have 4 large squares in a room walled off. You cannot let the ball bounce onto your side of the court more than once or the opposing team will receive a point.


86ac70
I think that it would be interesting to use different types on balls for each round in this game. It will keep the players on their toes because they will never know what ball they will be playing with. The balls can be beach balls, volley balls, tennis balls, and even balloons. I think that with the wide range of balls being used the players would benefit from hitting with their hands rather than with a pin-pong paddle (I think that the paddles would be too small for certain balls).
10ac78
capture the flag dodge ball. This would be played in an open field with 2 flags at opposing bases. there are 2 different colored sets of balls belonging to respective teams that only those teams can use. you try and grabbing the other teams flag without being hit by a ball and if you are hit than you must return to your base and "reset" yourself.


51ac91
also on the course, add certain things that will make course harder. USch as putting a beanie bag on the floor and you can't touch that
86ac70
I believe that it could be interesting if team mates who are hit by the opposing team have to give the accumulated flags back to this opposing team.
88ac15
to make it more interesting maby the people who do get hit are out of the game

Appendix B

Attractive and homogenous group:

Attractive and diverse group:

Unattractive and homogenous group:

Unattractive and diverse group:




Appendix C

On a scale of 1 (not at all) to 7 (very much):

How creative are the group's ideas?

How well did this group work together?

How competent is this group on their task?

On a scale of 1 (poor work) to 5 (exceptional work):

Overall, how would you evaluate this group's performance on their task of creating a new sport

1—poor work

2—needs improvement

3—meets expectations

4—exceeds expectations

5—exceptional work

On a scale of 1 (not at all) to 7 (very much):

How cohesive is this group?

How likable is this group?

How cooperative is this group?

On a scale of 1 (not at all) to 7 (extremely):

Considering this group as a whole, how attractive are they?

On a scale of 1 (not at all) to 7 (very much):

Considering this group as a whole, how diverse are they?

On a scale from 1 (not at all) to 7 (extremely):

Rate what you perceive to be your own level of attractiveness

Consider the person pictured and rate his/her attractiveness, where 1= not at all attractive, and 7 = extremely attractive.

Write in two to three ideas that you saw during the group's creative performance.

Demographic questions:

What is your age?

What is your gender identity?

Male

Female

Other

Prefer not to respond

What is your ethnicity?

Hispanic or Latino

American Indian or Alaska Native

Asian

Black or African American

Native Hawaiian or Other Pacific Islander

Caucasian or White

Multiracial

Other

Prefer not to respond

Ten-Item Personality Inventory (TIPI)

Here are a number of personality traits that may or may not apply to you. Please write a number next to each statement to indicate the extent to which you agree or disagree with that statement. You should rate the extent to which the pair of traits applies to you, even if one characteristic applies more strongly than the other.

Disagree strongly 1 2 3 4 5 6 7 *agree strongly*

Extraverted, enthusiastic

Critical, quarrelsome

Dependable, self-disciplined

Anxious, easily upset

Open to new experiences, complex

Reserved, quiet

Sympathetic, warm

Disorganized, careless

Calm, emotionally stable

Conventional, uncreative

Attitudes Toward Diverse Workgroups Scale (ADWS)

In today's work environment, people are often confronted with groups that vary in their diversity in terms of age, gender, race or ethnicity, expertise, background, and country of origin. We would like to know your feelings about working in groups that vary in their degree of diversity. A diverse group is one in which people differ from one another in one or more significant ways. In the statements below, please indicate to what extent you agree or disagree with them by marking the appropriate space on the answer sheet.

Strongly disagree 1 2 3 4 5 *Strongly agree*

1. I don't enjoy working with people who come from different countries. [A]
2. Working in diverse groups can increase one's understanding of those who are different from me. [P]
3. Being a leader of a diverse group should enhance a person's leadership ability. [P]
4. I prefer to socialize with people from my own ethnic group. [A]

5. For complicated problems, diverse groups will be able to solve the problem more easily. [P]
6. Groups whose members are diverse will be more creative. [P]
7. In general, I prefer socializing with people like myself. [A]
8. Workgroups with members from different cultural backgrounds are likely to be effective. [P]
9. Differences in political ideology within groups can stimulate one's thinking. [P]
10. The experiences of group members who come from different countries can be helpful in groups that are trying to generate novel ideas. [P]
11. I prefer working with people who are very similar to me. [A]
12. It is easier to be motivated when working with people who are like me. [A]
13. I find interacting with people from different backgrounds very stimulating. [P]
14. The experience of working with diverse group members will prepare me to be a more effective employee in an organization. [P]
15. Diverse groups can provide useful feedback on one's ideas. [P]
16. Solutions of complex problems require groups with diverse experiences or backgrounds. [P]
17. Conversations in diverse groups tend to be somewhat uncomfortable. [A]

Note. A = affective; P = productive.

Appendix D

Thank you for your participation in this experiment. You were told that the study related to group creativity and performance ratings, and that you were to rate these previous groups on certain performance measures. The ideas used are really from a prior group, but the images you saw for the have been manipulated to be either attractive and homogenous, attractive and diverse, unattractive and homogeneous, and unattractive and diverse. We did this because we are examining whether overall group physical attractiveness and diversity lead to better performance ratings on creative tasks. Now that you know the true purpose of this research study, you may decide whether you want to have your data removed from the study or not.

We urge you not to discuss this study with anyone who is currently participating or who might participate at a future point in time.

Your participation is greatly appreciated by the researchers involved. If you have any questions about this study, please contact Jade Chacon at jade.chacon@mavs.uta.edu or Jared Kenworthy at kenworthy@uta.edu.