Capstone Project, Summer 2020

Does the racial composition of teams affect how people evaluate their peers and distribute collective rewards among team members?

Author: Ankit Saxena

Advisor: Dr. Eugen Dimant

Project Duration: May 26, 2020 – August 7, 2020

Abstract

This study presents a variation of the dictator game experiment where participants were asked to share the collective rewards with their team members after performing a team-based task. Participants consisted of 341 males of either Black or White races from the United States. The experiment manipulated the racial composition of the teams, where participants were assigned to teams with two more members of either the same or different race as them. Participants could only see the virtual avatars of themselves and their team members, which had different skin colors based on race. After performing the task, participants were given the collective rewards for their teams and were asked to keep some mount of that for themselves and equally distribute the rest to their team members. Finally, questions on their normative and empirical expectations for this behavior were also asked to understand the motivations of their actions. Results showed no significant difference between peer evaluation scores and generosity of earnings for various racial compositions of teams. Implications of these results are discussed. Some additional analyses on the effect of other factors on the peer evaluation scores and generosity of earnings as well as the relationship between generosity and normative expectations have been discussed in the Appendix.

Keywords: Race, Team behaviors, Diversity, Generosity, Peer evaluation, Group identity

Table of contents

1. Acknowledgements

I would like to thank my Capstone Advisor, Dr. Eugen Dimant, for guiding me throughout this entire project and providing valuable feedback to my questions. His insights during this project as well as from my pervious interactions with him during the course BDS-501: Behavioral Science: Theory and Application of Experimental Methods have helped me tremendously in doing a project of a quality that I am satisfied with.

Also, I would like to thank the members of the Master of Behavioral and Decision Sciences department, specifically Dr. Chris Nave and Robin Hartley, who provided support and financial resources that made this project possible. An additional shout-out to the professors I studied under during the MBDS program in the 2019–2020 session is warranted. In spite of external difficulties like the pandemic, supported the students extremely well.

Other professors from the Master of Behavioral and Decision Science program have taught me valuable skills in their courses, which have all been useful in this project. Specifically, I would like to thank Dr. Cristina Bicchieri, Dr. Enrique Fatas, Dr. Edward Royzman, and Zarak Khan for teaching me the necessary skills to execute a successful Capstone Project.

Finally, I would like to thank Eduardo G. Araujo for being a great project partner and helping me with my project. It was great interacting him and learning from him during the course and the Capstone Project.

2. Introduction and Previous Literature

Many organizations try to build diverse teams to foster creative thinking and encourage desired team behaviors. Diversity can be in the form of different factors like race, gender, income, age, and so on. These qualities are a part of people's identities and can affect the level of in-group feelings towards that team. This study explored how diversity (or lack thereof) within a team in terms of race affected certain team-based behaviors. Specifically, this study measured two behaviors – how people evaluated the work of their peers and how they distributed collective rewards among team members. Finally, questions to measure participants' normative and empirical expectations were asked to understand if these expectations had a role in their observed behaviors.

In-group feelings based on race

Studies on homophily like Shrum et al. (1988) suggest that people tend to form in-group associations with others of similar demographic characteristics like race, which affects people's behavior. There is empirical evidence on the effect of in-group feelings induced by different mechanisms on various behaviors. For example, Gomez et al. (2000) showed that collectivist people perceived in-group members as more generous compared to out-group members. Also, Duclos and Barasch (2014) showed that people with an interdependent orientation may be encouraged to donate more towards in-group members. Finally, Blazovich (2013) showed the effect of compensation-structure and team-identity (induced by the same team color) on team performance and concluded that team identity based on team color did not affect productivity.

While these studies induced in-group feelings through various mechanisms like team color to observe behavioral changes, this study specifically focusses on race as a form of identity to induce in-group feelings. Racial identity formation has been shown to not only be prevalent but also have positive and negative externalities on each person's identity actions (Darity et al., 2006).

This study attempts to induce in-group feelings based on race by making skin color salient through virtual avatars. Virtual avatars have successfully been used in the past to induce behavioral responses by manipulating physical features. For example, Hasler et al. (2017) observed in-group

bias[1] due to race by assigning virtual body avatars to people with different skin colors. In that study, people displayed behaviors associated with interpersonal sensitivity and empathy towards others with avatars of the same skin color as their avatars', and displayed the opposite behavior towards people with avatars of a different skin color as theirs's. Another article by Yee et al. (2009) reported that physical characteristics like height and attractiveness of people's virtual avatars were significant predictors of their performance.

Building on the results of past studies on using virtual avatars to observe behavior changes, this study used virtual avatars of different skin colors to make racial differences and similarities salient among team members. An advantage of using these standardized avatars was that this approach minimized the possibility of inducing in-group feelings through other mechanisms like gender, age, etc. Appendix A shows the avatars used to represent people of different races.

Peer evaluation in teams

Peer evaluations are a standard practice for team-based projects in the corporate and academic world. Peer evaluations are used to measure the performance of team members and can often determine the compensation one receives for their work (Ohland et al., 2012). Therefore, this study used peer evaluation as one of the dependent variables of the experiment.

Participants were assigned to groups of three members and were asked to perform a task. Likewise, they were informed that the other two team members were doing the same task. After that, participants were asked to rate the work of their team members, and this evaluation was used as one of the dependent variables of this study. Furthermore, May and Gueldenzoph (2006) found some evidence that people gave higher ratings to team members who had a similar social style – based on the Social Styles Theory (Merrill and Reid, 1999; TRACOM, 1991) – than to those who had a different social style. Based on these findings, the first hypothesis of this study is as follows:

H1: People give a higher peer evaluation score to team members of the same race.

[1] According to Taylor and Doria (1981), in-group bias refers to a pattern of favoring members of one's in-group over out-group members

Generosity of collective earnings

Generosity is considered to be a desirable pro-social behavior in teams of all kinds. Brown (2011) studied the benefits of generosity in group settings and discovered that in-group helping behavior like generosity supported more egalitarian group relations, as indexed by reduced social dominance orientation (Pratto et al., 1994).

Generosity can be shaped by a number of factors. For example, Cramer et al. (1986) tested the effect of mood, similarity, and equity on peoples' generosity in sports settings and concluded that positive mood was correlated with peoples' generosity, which was consistent with the Equity Theory (Adams and Freedman, 1976). Furthermore, Duclos and Barasch (2014) showed that people with an interdependent orientation may be encouraged to donate more towards in-group team members. These results led to the second hypothesis of this study:

H2: People are more generous with the collective rewards of their team when working with peers of similar race as them.

This study uses a modified dictator game to measure participants' generosity. The dictator game has been extensively and effectively used to measure behaviors like generosity in economics studies and therefore is the method adopted by this study to measure generosity (Bolton et al., 1998).

Behaviors influenced by social norms

Social norms and expectations can play a key role in determining one's behavior. Bicchieri and Xiao (2009) explain that there are two different expectations that shape human behavior – empirical expectations (what we expect others to do) and normative expectations (believe others think we ought to do). To understand if these expectations played a role in the participants' observed behavior like generosity, questions were asked in the survey to measure empirical and normative expectations.

3. Experimental Design

Online experiment details

To test the hypotheses, an online experiment was conducted with a between-subject design (Appendix B). Participants were first given a basic overview of the experiment. They were told that they would first complete a task and then evaluate each other's work. Participants then had to answer basic demographic questions to obtain their gender, race, and age. Their answer to race was used subsequently to assign them randomly to various teams of varying racial compositions. However, in reality, they were merely assigned to hypothetical team members with predetermined avatars[2].

Next, participants were briefed on the nature of their task in the experiment, where they were told that each team member would come up with two possible names of a company that would be described to them. After that, the participants evaluated each other's responses real-time on a scale of 1 (worst) to 5 (best). Again, they only evaluated hypothetical responses predetermined by the experimenters[3]. Then, they were shown the voting results, which was predetermined by the experimenters and used as a control variable[4].

The participants were then provided with the entire team's earnings and then asked to keep part of it and distribute the rest equally among the team members, which was structured as a dictator game. A oneness scale was used to measure the level of inclusion the participants felt during the experiment (Appendix C). To further understand the underlying behavioral mechanisms that may have dictated their behavior, questions on empirical and normative expectations were asked (Appendix D). At the end, a debriefing form was provided to educate the participants on the exact nature of the study.

[2] This deception was necessary to ensure authentic responses and was approved by the Institutional Review Board. Furthermore, before submission of the responses, participants were provided with a detailed description of the exact purpose of the study and the deception techniques used in the study. They were then asked if they wanted to proceed with the submission or not. Refer to Appendix B to see the entire survey.

[3] This deception was necessary to ensure authentic responses and was approved by the Institutional Review Board.

[4] This deception was necessary to ensure authentic responses and was approved by the Institutional Review Board

Independent variable (Racial composition)

The independent variable was the racial composition of the team. Participants were assigned to one of the following treatments[5]:

1. Treatment I (White-In-group): A White participant was assigned to two other White team members.
2. Treatment II (White-Out-group): A White participant was assigned to two other Black team members.
3. Treatment III (Black-In-group): A Black participant was assigned to two other Black team members.
4. Treatment IV (Black-Out-group): A Black participant was assigned to two other White team members.
5. Control group: No race prime

To make these race primes salient, virtual avatars were used with different skin colors and were displayed to the participant. They were again shown subsequently to strengthen the prime.

Dependent variables (Evaluation score; Generosity)

The first dependent variable was the average of the evaluation scores given to the output of the participants' team members[6]. This was calculated using the average of the submitted values on five-point Likert scales (Joshi et al., 2015). This was done to measure the peer-to-peer evaluation scores. This variable will be referred to as 'evaluation score' in the following sections.

The second dependent variable was the amount of money kept (and hence distributed) by the participant among team members[7]. A total of USD 6 was provided and keeping an amount of USD 2 would mean an equal distribution of USD 2 among all team members including themselves. This variable will be referred to as 'generosity score' in the following sections.

[5] Pre-registered on AsPredicted.org
[6] Pre-registered on AsPredicted.org
[7] Pre-registered on AsPredicted.org

Control variable

The predetermined results of the peer-to-peer evaluation that were randomly displayed in two ways – one, where the participant received the highest points from peers, and two, where the participant received the lowest points from peers. This was done to ensure the points displayed in this step did not affect the final results in any way.

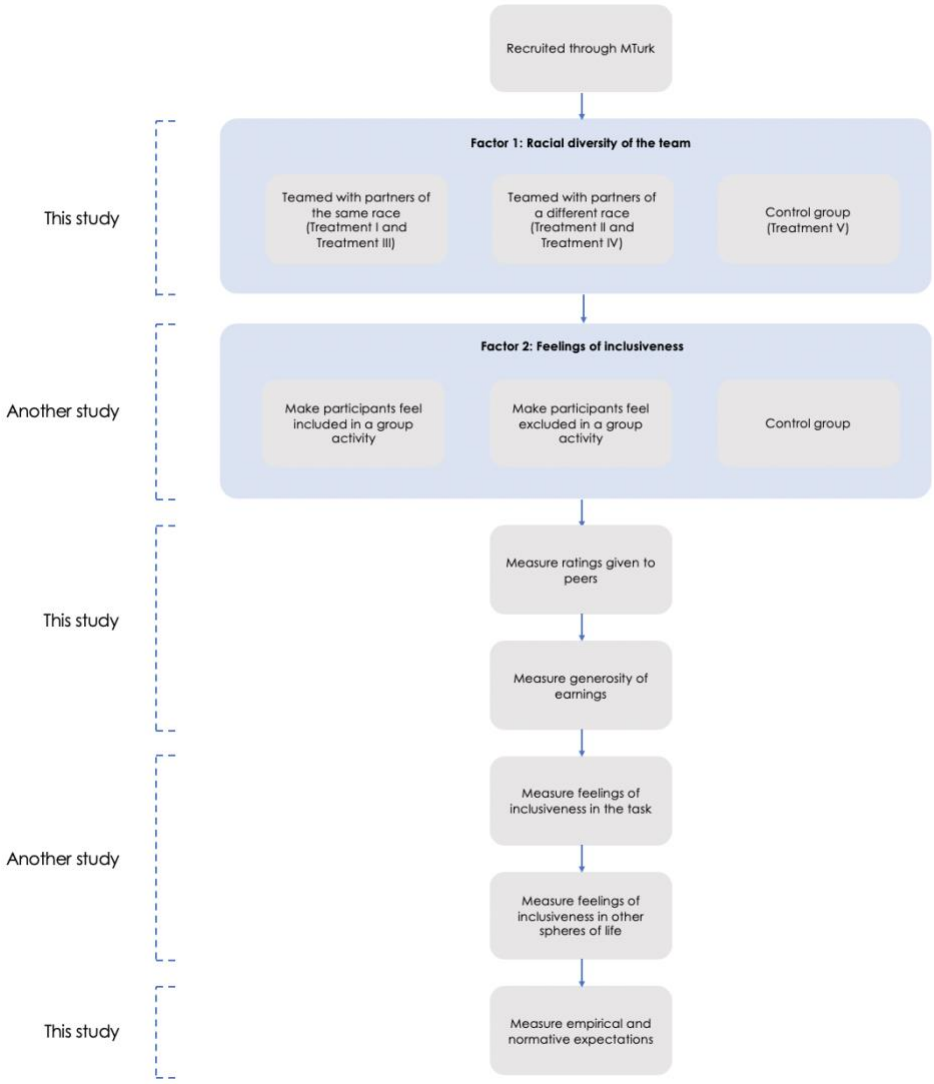


Figure 1: The experimental design flowchart. Specific sections relevant to this study have been identified on the left

4. Methodology

Data collection

The data were collected on Amazon Mechanical Turk (MTurk) on July 16, 2020. A total of 518 records were collected. While the total sample size was above the desired sample size determined by the power analysis described in Appendix E, there were mostly Black participants. Post hoc power analysis described in Appendix F showed that the study was underpowered.

Data exclusion

First, all responses which were incomplete were excluded. Then, for duplicate entries by the same person, their first response was retained assuming its authenticity and the subsequent responses were deleted. Then all non-males and those who didn't identify as White or Black were excluded. Also, an attention-check question was asked during the main task where participants were asked a comprehension question about the company description provided to them. Those that didn't pass this test were excluded. Finally, there were 341 participants remaining[8].

Data summary

Of the 341 data points used in the study, there were significantly more participants of Black race than of White race. However, the counts of in-group, out-group, and the control group were mostly similar. Table 1 summarizes the counts of various data points based on various groupings.

Table 1: Group-wise counts of data points collected

Group	Variable	Count
Gender of participant	Male	341
Race of participant	Black	268
	White	73
Race of team members	In-group	118
	Out-group	116
	No race	107
Treatment	T1: White-In-group	27
	T2: White-Out-group	25
	T3: Black-In-group	91
	T4: Black-Out-Group	91
	Control: No race	107

[8] Pre-registered on AsPredicted.org

Empirical strategy

To test the first hypothesis, the average evaluation scores for each of the treatment groups were compared with those of the control group in two-tailed t-tests for two samples[9].

To test the second hypothesis, the average generosity scores for each of the treatment groups were compared with those of the control group in two-tailed t-tests for two samples[10].

Analysis of the differences of evaluations scores and generosity scores given by people of different races were conducted using two-tailed t-tests for two samples between each treatment group and the control group. Similar analysis on the in-group/out-group nature of teams was also done using two-tailed t-tests for two samples. The relationship between evaluation scores and generosity with the race of the participant's team members was also done using t-tests. Finally, the relationship between social (empirical and normative) expectations and generosity was explored using a regression analysis (Appendix K).

[9] Pre-registered on AsPredicted.org
[10] Pre-registered on AsPredicted.org

5. Analysis Results

Hypothesis test I: Effect of the team's racial composition on evaluation scores

There were no significant differences between the evaluation scores given by participants in the other treatment groups and the control group for the two-tailed tests. Figure 2 shows the mean values of evaluation scores for each treatment along with their confidence intervals[11]. Table 2 shows the summarized results of the t-tests.

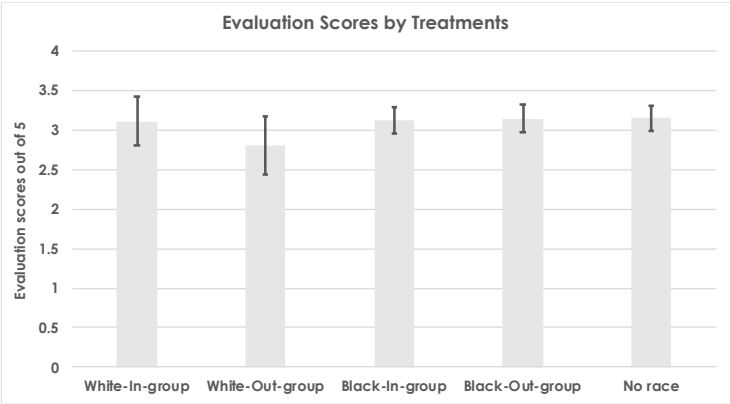


Figure 2: Average evaluation scores and their confidence intervals for the treatment groups and the control group

Table 2: Results from the t-tests for the first hypothesis (statistically significant values marked with *)

Test	Groups	Sample Size	Mean	Variance	P-value (one-tailed)	P-value (two-tailed)
White-In-group <> Control	White-In-group	27	3.111	0.915	0.434	0.868
	Control	107	3.147	1.019		
White-Out-group <> Control	White-Out-group	25	2.8	1.172	0.065 *	0.129
	Control	107	3.147	1.019		
Black-In-group <> Control	Black-In-group	91	3.118	0.896	0.418	0.836
	Control	107	3.147	1.019		
Black-Out-group <> Control	Black-Out-group	91	3.140	1.042	0.480	0.961
	Control	107	3.147	1.019		

[11] Confidence intervals were calculated at the pre-registered 90 percent confidence level

Hypothesis test II: Effect of the team's racial composition on generosity

There were no significant differences between the generosity scores of the participants in the treatment groups and the control group for the two-tailed t-tests. Figure 3 shows the mean values of generosity scores for each treatment along with their confidence intervals[12]. Table 3 shows the summarized results of the t-tests.

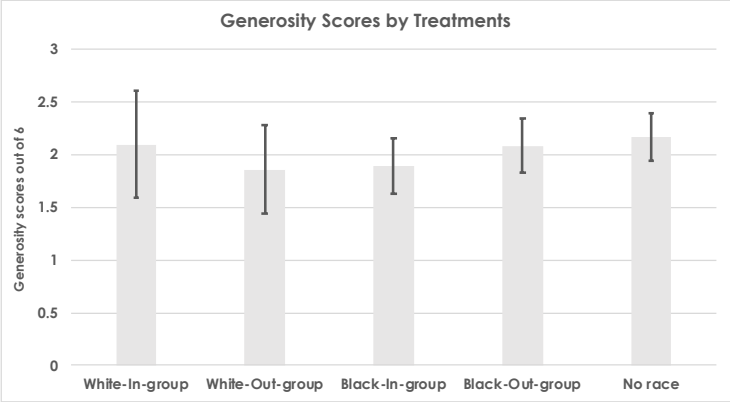


Figure 3: Average generosity scores and their confidence intervals for the treatment groups and the control group

Table 3: Results from the t-tests for the second hypothesis (statistically significant values marked with *)

Test	Groups	Sample Size	Mean	Variance	P-value (one-tailed)	P-value (two-tailed)
White-In-group <> Control	White-In-group	27	2.096	2.357	0.412	0.824
	Control	107	2.164	1.927		
White-Out-group <> Control	White-Out-group	25	1.856	1.501	0.155	0.310
	Control	107	2.164	1.927		
Black-In-group <> Control	Black-In-group	91	1.892	2.238	0.093 *	0.186
	Control	107	2.164	1.927		
Black-Out-group <> Control	Black-Out-group	91	2.084	2.164	0.345	0.691
	Control	107	2.164	1.927		

[12] Confidence intervals were calculated at the pre-registered 90 percent confidence level

Additional analyses

Additional analyses on the effect of participants' race, team's in-group/out-group nature, and race of team members on the dependent variables can be found in Appendix G, Appendix H, and Appendix I respectively. Finally, the relationship between the generosity score and social expectations can be found in Appendix J.

6. Discussion and Conclusion

For the purpose of this study, there were two hypotheses. First, that participants would give higher evaluation scores to team members of similar race as them, and second, that participants would be more generous with the collective rewards for team members of the same race. These hypotheses were determined by prior literature on homophily and generosity towards in-group members. However, in both the hypotheses, the evaluation scores and the generosity scores were higher for the groups without race prime compared with the different race prime treatments, but, these differences were not statistically significant.

The absence of any significant results may be indicative of the fact that there aren't any systemic effects of racial composition of teams on peer evaluations and generosity of collective rewards. Another possible explanation could be that people tend to form other in-group and out-group associations with peers that are different from race, like gender, age, etc. Yet another plausible explanation could be that the technique used in this study to prime people based on race was not strong enough and in order to prime people by race, other techniques may need to be used. A definite conclusion cannot be made since both the hypotheses were underpowered. The test of groups with the strongest achieved power was 76 percent which was less than the desired power of 80 percent.

The results of this study can show how people may behave in online communities especially in companies where most people work in teams. Specifically, it shows how race may play a minimal role in people's behavior in team projects. The null effects observed may suggest that avatars customized to reflect people's race may not have a strong effect on peer evaluations and generosity. Managers can use insights from this study to show that building racially diverse teams may not have much of an impact on generosity and peer evaluations. However, a study with a stronger power may be a better source of evidence for this claim.

Finally, the lack of differences in the dependent variables due to other race-related factors (discussed from Appendix G to Appendix J) may be evidence that systemic behaviors against certain races may not occur in such scenarios where virtual avatars are involved. Again, a study with more power will be needed to strengthen this claim.

7. Self-evaluation

This was the first time I designed an experiment on my own with a research question that I felt was interesting, useful, and relevant to today's time. There were many things that went well and many other areas that can be improved.

What went well

The survey design was efficient and compact. It took an average of less than 8 minutes (and a median of about 7 minutes) for the participants to complete the survey, which is a reasonable time for participants to maintain their attention span and for us to obtain authentic responses. Had the survey been longer, the attention of participants would have dwindled thereby reducing the quality of data. Also, the feedback provided by the survey participants in the open-ended question at the end contained positive responses for almost all of the cases. This is a good indicator that participants enjoyed doing our experiment.

This project was a joint effort by me and Eduardo G. Araujo. There is no doubt that working as a pair elevated the quality of both our experiments. We bounced ideas off of each other and were able to help each other whenever the other was stuck and a crucial decision related to the literature review, survey design, and experiment execution. I would recommend others to consider working in pairs, while ensuring that they meet the program's requirements of doing a complete capstone project.

Also, performing each step of a research project – literature review, problem definition, experiment design, experiment execution, data analysis, and reporting – was overall a rewarding experience. The research problem I explored was very interesting to me and therefore, I was able to give my 100 percent to it. I was also able to follow all my preregistered plans that I submitted to Aspredicted.org which was encouraging.

Finally, this project is a good launching pad for future studies in the realm of organizational behavior, diversity, and race. I would enjoy expanding on the insights from this project to perform a similar, but improved, study on a larger scale to come up with meaningful insights.

What could be improved

I felt there were many interesting analyses I could have preregistered, but did not consider earlier. I should plan ahead in future projects to ensure I preregister and plan for such interesting analyses. In my pre-registration, I mentioned using a t-test for all my analyses, however, for the collected sample size, a non-parametric test would have been more suitable.

There were some advantages and disadvantages to working in a shared project with Eduardo. While the advantages included a better quality of the overall survey due to multiple minds working on the same output, it was challenging to clearly delineate our two projects. But, because of his additional questions, we were able to conduct some interesting analyses on inclusiveness, social expectations, and race, which would not have been possible otherwise. In spite of our survey being a meant for two separate studies, we did a great job a minimizing the survey length and, in the end, it was extremely manageable. Future students should weight the advantages and disadvantages to determine if doing a capstone project in pairs is a good idea or not.

The lack of statistically significant results suggests that perhaps the race prime was not strong enough to induce a reaction. Future studies should keep that in mind and try to apply a stronger race prime.

Also, due to the unique design of the study that involved deception, I did not conduct a full-fledged pre-test among my peers. Instead, I sent this survey to some people for feedback. While the feedback was tremendously useful, perhaps I could have spent a small amount of money to conduct a pre-test on MTurk to foresee the imitations of this experiment, specifically the lack of a strong race prime.

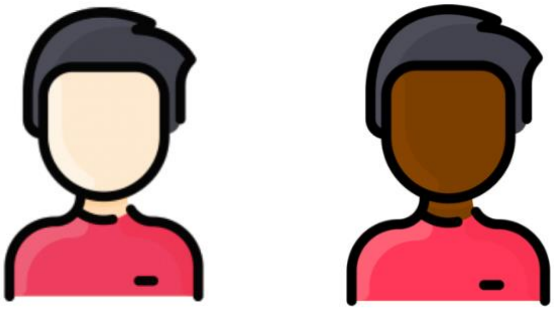
Finally, I made a technical error when designing the questions measuring the empirical expectations. I asked two questions to understand that out of 10 people, how many would (i) split the rewards equally, and (ii) maximize their own rewards. The total of these two answers should by less than or equal to 10, since out of 10 people, a fraction would split, while the rest would maximize or minimize the rewards. However, many responses for these two questions had a total of more than 10, which is paradoxical. I should have added some constraints or explained the questions more clearly for better-quality responses. Because of this technical error, the analysis on social expectations was incomplete.

8. Appendix

Appendix A: Avatars used to represent race

Since only males were studied in this experiment, the avatars were only designed keeping male participants in mind. The skin color of the digital avatars was varied to reflect the race of a participant.

Virtual avatars



White race Black race

Appendix B: Complete survey

Step 1: Consent form

Informed Consent to Participate in Research

What is the purpose of this study?
You are invited to take part in a study named Team Decision Task. Generally, the purpose of this study is to understand human decision-making in groups. A fully detailed disclosure of the purpose and intent of this study will be shared at the end of the survey when all steps have been completed. When you are done, we will explain why we are doing this study, what we are looking at, and any other information you should know about this study. You will also be able to ask any questions you might have about the study's purpose and the tasks you did. Though we may not be able to explain the real purpose of the study until after you complete the tasks, there are no additional risks to those that have been described in this consent form.

Participation
Your participation in this study is completely voluntary, and you may refuse to participate. If you withdraw from the study before completing it, you will not be paid. Compensation will be awarded upon completion of the entire study. You will be paid within 10 days of completing the study via MTurk. Your MTurk ID will only be used to pay you correctly and will be deleted permanently from the experimenter's data after payment is complete. You will also receive a debriefing form after the survey.

What will you do?
You will be paired with two more participants to perform a task. You will all be asked to complete a task and then evaluate each other's work. We will also ask you to provide demographic information. We will not ask for your name or any information that will make you identifiable. Overall, this study will take approximately 8 to 12 minutes.

Risks
The risks to participating are no greater than those encountered in everyday life. The major risk in this study concerns breach of confidentiality, whereby unauthorized individuals might obtain the data collected. We minimize this risk by using a secure, password-protected server that only the researches at the University of Pennsylvania will have access to. Make sure you have read Amazon's MTurk participant and privacy agreements to understand how your personal information may be used or disclosed. The data collected through this survey will be visible to only the two researchers involved in this study.

If you have any questions about this study, you may contact us at the emails provided at the end.

Please feel free to print or save a copy of this consent form.

Step 2: Demographic information

Q1. What gender do you identify with?

- Female
- Male
- Other (please specify)

Q2. What race do you identify with?

- Black or African American
- White
- Other

Q3. What is your age?

- Under 18
- 18-24 years old
- 25-34 years old
- 35-44 years old
- 45-54 years old
- Above 54

Step 3: Task instructions

Instructions. Please carefully read the following instructions about your task.

You will be put together into groups of three. Two more people will be assigned to your team to perform the task.

You and your team members will be shown a short description of a company. Each one of you will first read the description of the company, and then submit two possible names for the company. Be creative! Remember, there are no wrong answers.

After that, all of you will evaluate each other's suggested names, voting on the best and the worst names in a scale from 1 (worst) to 5 (best). Finally, we will reveal the voting results. Results will be shown as the average score given by all participants to a given name.

After the task, you will be chosen as the team leader and will be given the entire team's earnings. You will be asked to keep a certain amount of it and distribute the remaining earnings among your team members.

Please proceed if you have read the instructions and are ready to be assigned to your team.

Step 4a: Team assignment (waiting)

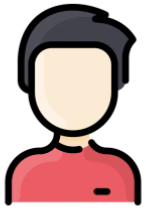
Instructions. Please wait while we assign you to your team.

Step 4b: Team assignment (No race prime)

Instructions. You have been assigned to a team.

Step 4b: Team assignment (White-Out-group)

Instructions. Based on your gender and race, we have selected the following avatar for you:



Instructions. You have been assigned to the following team members:

Participant A	(YOU) Participant B	Participant C

Step 4b: Team assignment (White-In-group)

Instructions. Based on your gender and race, we have selected the following avatar for you:

Instructions. You have been assigned to the following team members:

Participant A	(YOU) Participant B	Participant C

Step 4b: Team assignment (Black-Out-group)

Instructions. Based on your gender and race, we have selected the following avatar for you:

Instructions. You have been assigned to the following team members:

Participant A	(YOU) Participant B	Participant C

Step 4b: Team assignment (Black-In-group)

Instructions. Based on your gender and race, we have selected the following avatar for you:

Instructions. You have been assigned to the following team members:

Participant A	(YOU) Participant B	Participant C

Step 5: Team task

Instructions. A company has been selling office supplies like tables, chairs, and power cables to other workplaces for the past 2 years and has experienced tremendous popularity and growth in the United States.

Now, they would like to launch a spinoff company that would focus on selling stationery items to offices and schools. They will produce items like pens, notepads, rulers, pushpins, and so on. Some of these items have been shown in the image below.

Q4. Which of the following items below is an example of what the **spinoff company** will sell?

- Shoes
- Notepads
- Vegetables

Q5. What could be a good company name for the spinoff stationery company?

Please provide your best idea in the text box below.

Q6. Please provide one more possible name for the spinoff company in the text box below.

Instructions. You will be allowed to proceed to the evaluation step after your team members have submitted their responses. The button to proceed will become visible when your team members complete their tasks. Remember, each of you has a maximum of **5 minutes** to complete the questions above. The time elapsed is visible below.

Step 6: Evaluation

Q7. Please evaluate the company name ideas submitted by you and your team members. Each of you has a maximum of **3 minutes** to evaluate these names. The button to proceed will become visible when your team members complete their evaluations.

	Excellent - 5 points	Good - 4 points	Fair - 3 points	Poor - 2 points	Very bad - 1 point
Participant A: Stationeroo	O	O	O	O	O
Participant A: Workplace Supplies	O	O	O	O	O
Participant C: Craftsmen	O	O	O	O	O
Participant C: Materialistics	O	O	O	O	O

Step 7: Results (Excluded)

Results.

Submitted Company Name	Average Score
Participant A: Stationeroo	4.8
Participant C: Materialistics	4.7
Participant A: Workplace Supplies	4.2
Participant C: Craftsmen	3.6
You:	3.1
You:	2.8

Step 7: Results (Included)

Results.

Submitted Company Name	Average Score
You:	4.8
You:	4.7
Participant A: Workplace Supplies	4.2
Participant C: Craftsmen	3.6
Participant C: Materialistics	3.1
Participant A: Stationeroo	2.8

Step 8: Rewards distribution

Q8. Congratulations! Your team has completed the necessary task. Since you are the team leader, you are being provided **US$ 6** as compensation for **your entire team**. Please use the slider below to indicate how much money **you would like to keep**. Whatever amount you decide **not to keep** will be **equally distributed** among the other three teammates. For example, if you keep US$ 2, then all team members get US$ 2.

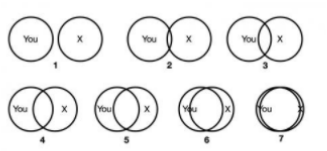
0 1 2 3 4 5 6

Money to keep

Step 9: Measuring inclusiveness

Q9 In the following figure, we ask you to consider which of these pairs of circles best describes how you felt about your relationship with the team you were matched with.

In the figure, you should think of "X" as your team.



1 - Not close at all

2 - Hardly close

3 - A little close

4 - Somewhat close

5 - Quite close

6 - Very Close

7 - Extremely close

Q10. Using the same figure scale, how do you think the **other participants** in your team felt about their relationship among themselves?

1 - Not close at all

2 - Hardly close

3 - A little close

4 - Somewhat close

5 - Quite close

6 - Very close

7 - Extremely Close

Q11. Using the same figure scale, how do you think the **other participants** in your team felt about their relationship with you?

- 1 - Not close at all
- 2 - Hardly close
- 3 - A little close
- 4 - Somewhat close
- 5 - Quite close
- 6 - Very close
- 7 - Extremely close

Q12. Please answer the following questions based on your best judgement.

	5 - Almost everyday	4 - At least once a week	3 - A few times a month	2 - A few times a year	1- Less than once a year	0 - Never
You are treated with less courtesy or respect than other people.	O	O	O	O	O	O
You receive poorer service than other people at restaurants or stores.	O	O	O	O	O	O
You are threatened or harassed.	O	O	O	O	O	O

Step 10: Measuring social expectations

Instructions. Please answer the following questions based on your best judgement.

Q13. Of 10 people, how many people do you think **split the money evenly** across the four team members?

0 1 2 3 4 5 6 7 8 9 10

Number of people

Q14. Of 10 people, how many people do you think tried to **keep the maximum amount** of the money across the four team members?

0 1 2 3 4 5 6 7 8 9 10

Number of people

Q15. Do you believe the participant should have split the money equally across the four members, i.e., a payoff of US$ 2 for each member?

- Yes
- No
- I don't know

Q16. Of 10 people, how many people do you think answered **yes** to the above questions?

0 1 2 3 4 5 6 7 8 9 10

Number of people

Step 10: Measuring social expectations

University of Pennsylvania
Debriefing Form to Participate in Research

Thank you for participating in this study! Your involvement is greatly appreciated. The material on this form details information about your experience in this study

The procedure of this study
Earlier in our consent form, we informed you that this study involves performing a team-based task where you will be partnered with two other people. However, we did not actually partner you with anyone in real-time. The spinoff company names and scores displayed after the evaluation phase were hypothetical and predetermined. They were not actually scored by any peers in real-time. Each participant like you was shown one of the two predetermined scenarios, where people gave your submission either the highest average score or the lowest average score. This was done to vary the level of inclusiveness felt by the participants and observe how that affected the participant's responses.

All of these measures were adopted to simplify the logistics of executing the experiment.

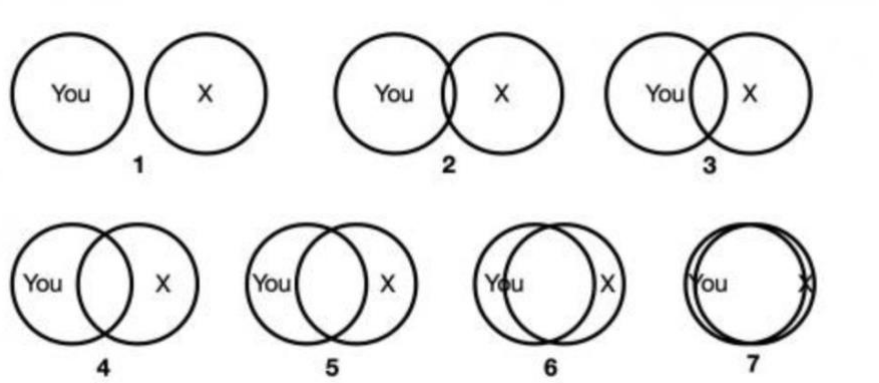
Confidentiality
Please note that your information will still remain confidential and no identifiable information will be retained by the two researchers involved. Your MTurk IDs will be deleted as soon as your payment is processed.

If you would like to ask us further questions or would like to receive a summary of our findings, please feel free to contact us at asaxena7@sas.upenn.edu or earaujo@sas.upenn.edu.

Appendix C: The oneness scale

The goal of the oneness scale (IOS) is to test the primacy manipulation and neutrally assess the closeness of individuals without requiring disclosure of subjective perceptions or personal motivations (Gachter, et al., 2015; Aron and Smollan, 1992). Concurrently, it is important to acknowledge that the survey is a self-reported measure and is subject to experiment demand effects (Zizzo, 2010).

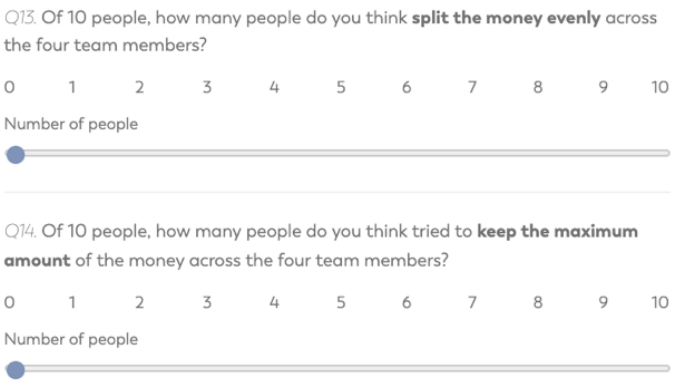
Each pair of circles differs in the level of overlap, representing the closeness of the relationship between the participant and their counterpart.



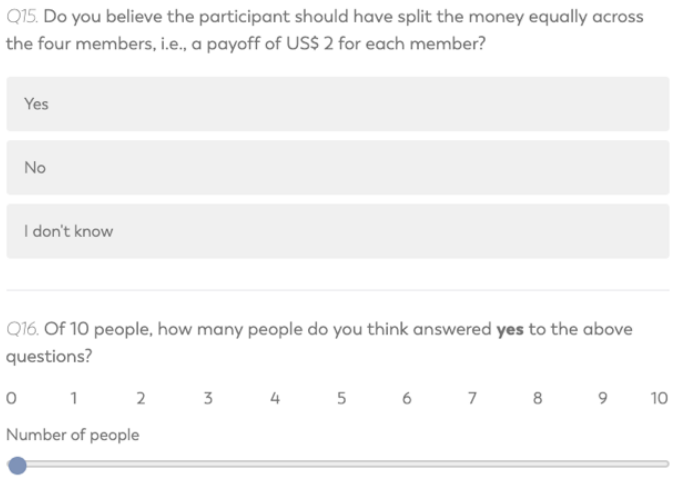
Appendix D: Measuring normative and empirical expectations

Bicchieri and Xiao (2009) show that there are two dominant aspects of social norms that play an important role in individual decision making. The first is one's empirical expectations, which refers to what we think others will do. The second is normative expectations, which refers to what we believe others think we should do. The following questions were designed to measure these two expectations:

Empirical expectations



Normative expectations

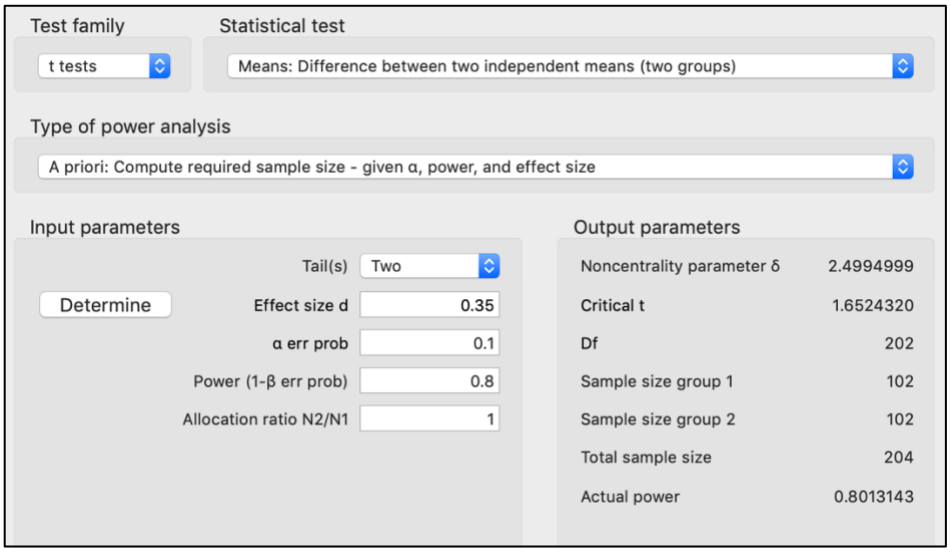


Appendix E: A priori power analysis for sample size determination

The desired sample size for the hypothesis tests was determined to be 500, i.e., 100 per treatment group and the control group.

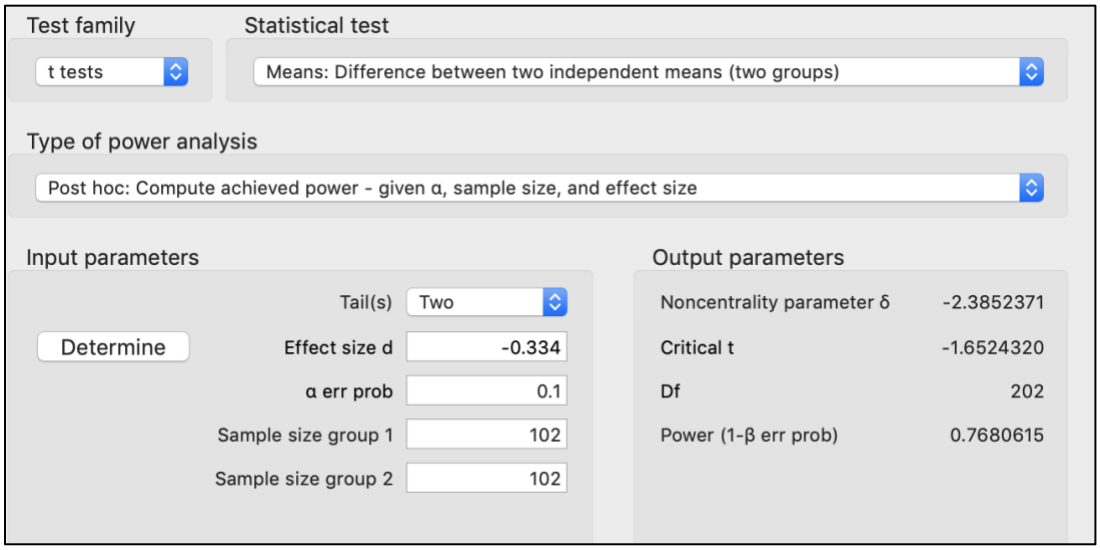
Assumptions

- Desired power: 80 percent
- Confidence level: 90 percent
- Effect size: 0.35



Appendix F: Post hoc power analysis for calculating achieved power

Post hoc power calculations reveal that the hypothesis tests were underpowered with an achieved power of 76 percent for the comparison of groups with the highest effect size and sample sizes. This means that the other group wise comparisons were more underpowered than 76 percent.



Appendix G: Participants' race on evaluation and generosity scores

Evaluation scores

No statistically significant difference of evaluations scores between Black and White participants based on a two-tailed t-test at 90 percent confidence level.

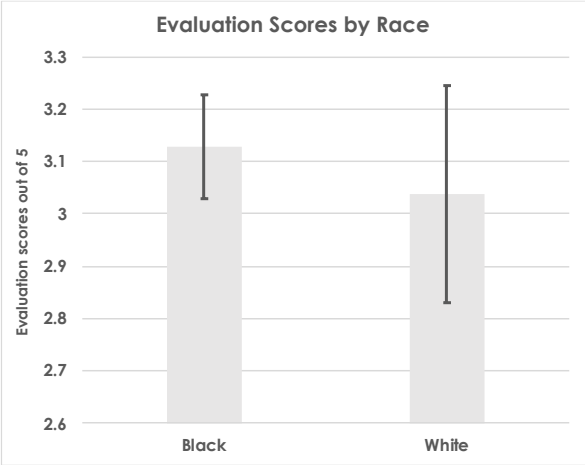


Figure: Average evaluation scores and their confidence intervals for Black and White participants

Table: Results from the t-tests (statistically significant values marked with *)

Test	Groups	Sample Size	Mean	Variance	P-value (one-tailed)	P-value (two-tailed)
White <> Black	White participants	73	3.038	1.123	0.245	0.489
	Black participants	268	3.129	0.958		

Generosity scores

No statistically significant difference of generosity scores between Black and White participants based on a two-tailed t-test at 90 percent confidence level.

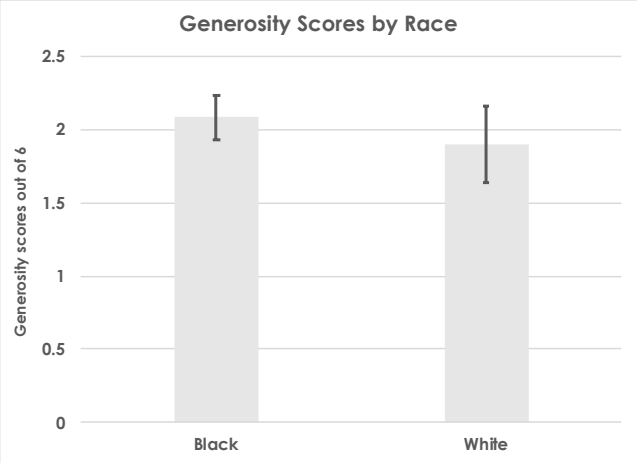


Figure: Average generosity scores and their confidence intervals for Black and White participants

Table: Results from the t-tests (statistically significant values marked with *)

Test	Groups	Sample Size	Mean	Variance	P-value (one-tailed)	P-value (two-tailed)
White <> Black	White participants	73	3.038	1.123	0.245	0.489
	Black participants	268	3.129	0.958		

Appendix H: Effect of team's in-group/out-group nature on evaluation and generosity scores

In-group refers to White participants with White team members and Black participants with Black team members. Out-group refers to White participants with Black team members and Black participants with White team members

Evaluation scores

No statistically significant difference of evaluations scores between in-group and out-group treatments and the control group based on a two-tailed t-test at 90 percent confidence level.

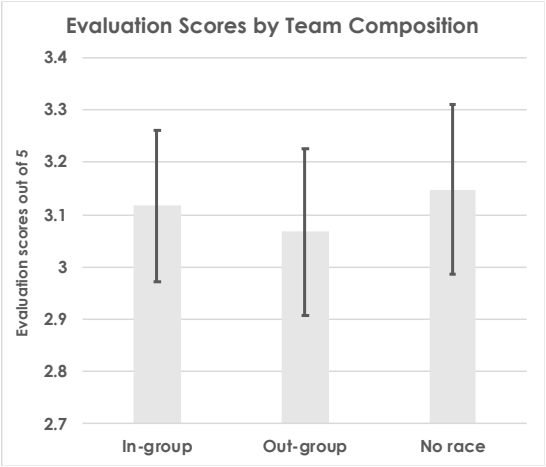


Figure: Average evaluation scores and their confidence intervals

Table: Results from the t-tests (statistically significant values marked with *)

Test	Groups	Sample Size	Mean	Variance	P-value (one-tailed)	P-value (two-tailed)
In-group <> Control	In-group	118	3.117	0.893	0.407	0.814
	Control	107	3.147	1.019		
Out-group <> Control	Out-group	116	3.067	1.080	0.280	0.560
	Control	107	3.147	1.019		

Generosity scores

No statistically significant difference of generosity scores between in-group and out-group treatments and the control group based on a two-tailed t-test at 90 percent confidence level.

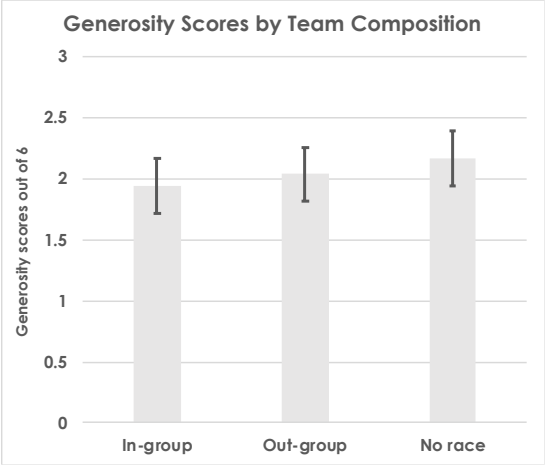


Figure: Average generosity scores and their confidence intervals

Table: Results from the t-tests (statistically significant values marked with *)

Test	Groups	Sample Size	Mean	Variance	P-value (one-tailed)	P-value (two-tailed)
In-group <> Control	In-group	118	1.940	2.253	0.122	0.244
	Control	107	2.165	1.930		
Out-group <> Control	Out-group	116	2.035	2.016	0.245	0.490
	Control	107	2.165	1.930		

Appendix I: Effect of race of team members on evaluation and generosity scores

Evaluation scores

No statistically significant difference of evaluations scores for participants assigned to teams with all Black team members and those assigned to teams with all White team members versus the control group based on a two-tailed t-test at 90 percent confidence level.

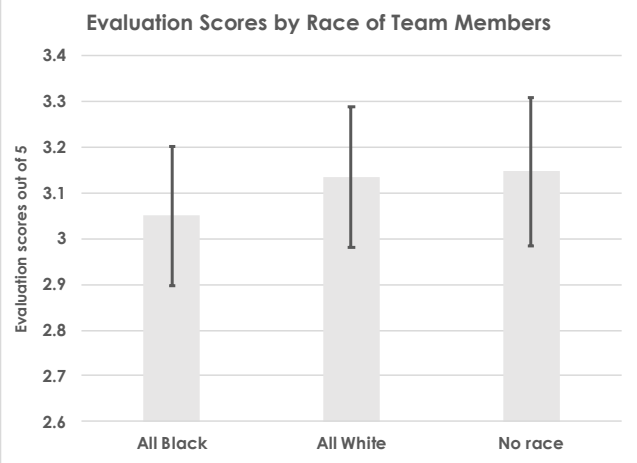


Figure: Average evaluation scores and their confidence intervals

Table: Results from the t-tests (statistically significant values marked with *)

Test	Groups	Sample Size	Mean	Variance	P-value (one-tailed)	P-value (two-tailed)
All Black <> Control	**All Black**	116	3.050	0.963	0.233	0.465
	Control	107	3.147	1.019		
All White <> Control	**All White**	118	3.133	1.005	0.460	0.919
	Control	107	3.147	1.019		

Generosity scores

No statistically significant difference of generosity scores for participants assigned to teams with all Black team members and those assigned to teams with all White team members versus the control group based on a two-tailed t-test at 90 percent confidence level.

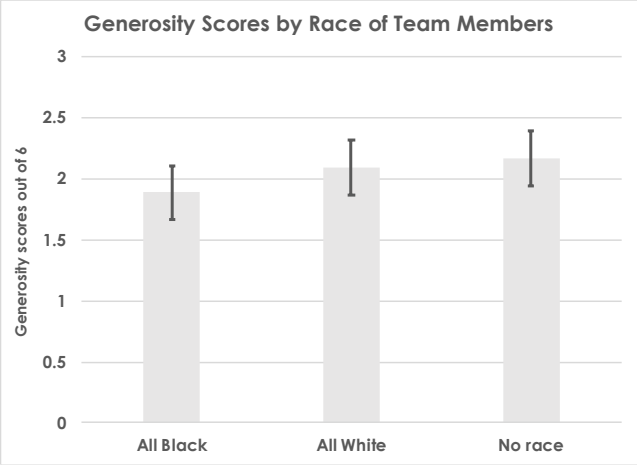


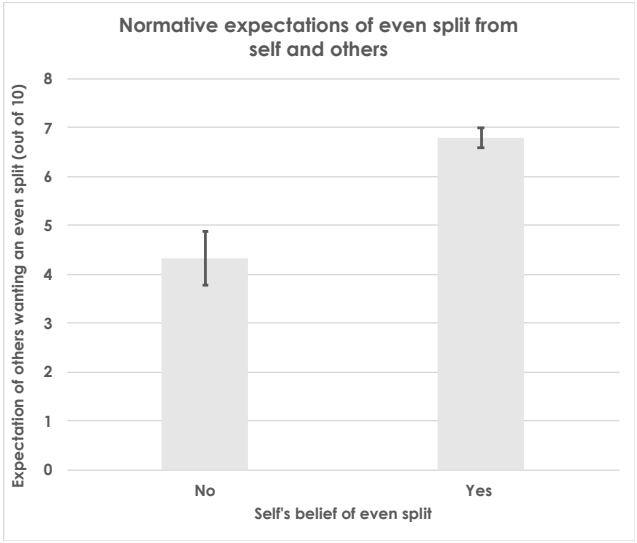
Figure: Average generosity scores and their confidence intervals

Table: Results from the t-tests (statistically significant values marked with *)

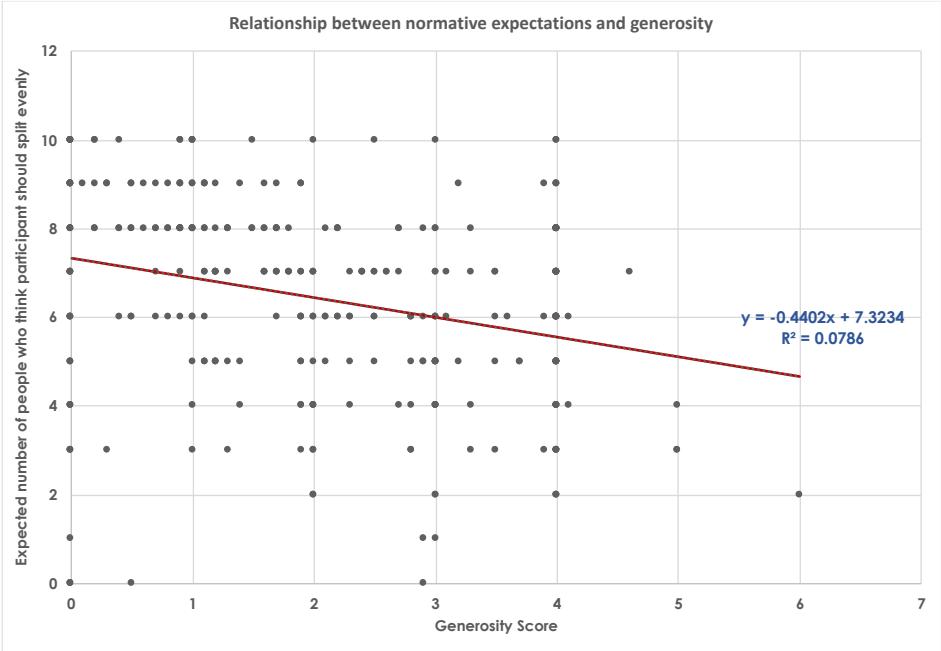
Test	Groups	Sample Size	Mean	Variance	P-value (one-tailed)	P-value (two-tailed)
All Black <> Control	**All Black**	116	1.885	2.065	0.070 *	0.141
	Control	107	2.165	1.930		
All White <> Control	**All White**	118	2.086	2.189	0.342	0.684
	Control	107	2.165	1.930		

Appendix J: Relationship between generosity and normative expectations

There is a clear distinction in the normative expectations of dividing the rewards equally between people who thought they ought to split the rewards equally versus those who did not think so, as seen in the below figure.



Additionally, there is a weak linear relationship between the generosity of collective rewards and the normative expectations of an even split of rewards.



9. Bibliography

Adams, J. S., & Freedman, S. (1976). Equity theory revisited: Comments and annotated bibliography. In Advances in experimental social psychology (Vol. 9, pp. 43-90). *Academic Press*.

Aron, A., Aron, E. N., & Smollan,D. (1992). Inclusion of other in the self scale and the structure of interpersonal closeness. *Journal of personality and social psychology*, 63(4), 596.

Blazovich, J. L. (2013). Team identity and performance-based compensation effects on performance. *Team Performance Management: An International Journal*.

Bicchieri, C., & Xiao, E. (2009). Do the right thing: but only if others do so. *Journal of Behavioral Decision Making*, 22(2), 191-208.

Brown, M. A. (2011). The power of generosity to change views on social power. *Journal of Experimental Social Psychology*, 47(6), 1285-1290.

Bolton, G. E., Katok, E., & Zwick, R. (1998). Dictator game giving: Rules of fairness versus acts of kindness. *International journal of game theory*, 27(2), 269-299.

Cramer, R. E., McMaster, M. R., Lutz, D. J., & Ford, J. G. (1986). Sport fan generosity: A test of mood, similarity, and equity hypotheses. *Journal of sport Behavior*, 9(1), 31.

Darity Jr, W. A., Mason, P. L., & Stewart, J. B. (2006). The economics of identity: the origin and persistence of racial identity norms. *Journal of Economic Behavior & Organization*, 60(3), 283-305.

Duclos, R., & Barasch, A. (2014). Prosocial behavior in intergroup relations: How donor self-construal and recipient group-membership shape generosity. *Journal of Consumer Research*, 41(1), 93-108.

Gächter, S., Starmer, C., & Tufano, F. (2015). Measuring the closeness of relationships: a comprehensive evaluation of the 'Inclusion of the Other in the Self' scale. *PloS one*, 10(6).

Gomez, C., Kirkman, B. L., & Shapiro, D. L. (2000). The impact of collectivism and in-group/out-group membership on the evaluation generosity of team members. *Academy of management Journal*, 43(6), 1097-1106.

Hasler, B. S., Spanlang, B., & Slater, M. (2017). Virtual race transformation reverses racial in-group bias. *PloS one*, 12(4), e0174965.

Joshi, A., Kale, S., Chandel, S., & Pal, D. K. (2015). Likert scale: Explored and explained. *Current Journal of Applied Science and Technology*, 396-403.

May, G. L., & Gueldenzoph, L. E. (2006). The effect of social style on peer evaluation ratings in project teams. *The Journal of Business Communication* (1973), 43(1), 4-20.

Merrill, D. W., & Reid, R. H. (1981). Personal styles & effective performance. *CRC Press*.

Ohland, M. W., Loughry, M. L., Woehr, D. J., Bullard, L. G., Felder, R. M., Finelli, C. J., ... & Schmucker, D. G. (2012). The comprehensive assessment of team member effectiveness: Development of a behaviorally anchored rating scale for self-and peer evaluation. *Academy of Management Learning & Education*, 11(4), 609-630.

Pratto, F., Sidanius, J., Stallworth, L. M., & Malle, B. F. (1994). Social dominance orientation: A personality variable predicting social and political attitudes. *Journal of personality and social psychology*, 67(4), 741.

Shrum, W., Cheek Jr, N. H., & MacD, S. (1988). Friendship in school: Gender and racial homophily. *Sociology of Education*, 227-239.

Taylor, D. M., & Doria, J. R. (1981). Self-serving and group-serving bias in attribution. *The Journal of Social Psychology*, 113(2), 201-211.

TRACOM. (1991). The social style profile–technical report: Development, reliability and validity. *Denver, CO: The TRACOM Corporation*.

Yee, N., Bailenson, J. N., & Ducheneaut, N. (2009). The Proteus effect: Implications of transformed digital self-representation on online and offline behavior. *Communication Research*, 36(2), 285-312.

Zizzo, D. J. (2010). Experimenter demand effects in economic experiments. *Experimental Economics*, 13(1), 75-98