Journal Pre-proofs

Full Length Article

Having Less, Giving More? Two Preregistered Replications of the Relationship Between Social Class and Prosocial Behavior

Angelos Stamos, Florian Lange, Szu-chi Huang, Siegfried Dewitte

PII: S0092-6566(19)30123-0
DOI: <https://doi.org/10.1016/j.jrp.2019.103902>
Reference: YJRPE 103902

To appear in: *Journal of Research in Personality*

Received Date: 27 October 2017
Revised Date: 8 November 2019
Accepted Date: 17 November 2019

Please cite this article as: Stamos, A., Lange, F., Huang, S-c., Dewitte, S., Having Less, Giving More? Two Preregistered Replications of the Relationship Between Social Class and Prosocial Behavior, *Journal of Research in Personality* (2019), doi: <https://doi.org/10.1016/j.jrp.2019.103902>

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2019 Published by Elsevier Inc.

Authors:

Angelos Stamos (corresponding author) is Postdoctoral Researcher at the KU Leuven, Behavioral Engineering Group, Campus Leuven, Naamsestraat 69 - box 3545, 3000 Leuven, Belgium.

Email: Angelos.Stamos@kuleuven.be, Phone: +3216326741

Florian Lange is Postdoctoral Researcher at the KU Leuven, Behavioral Engineering Group, Campus Leuven, Naamsestraat 69 - box 3545, 3000 Leuven, Belgium. Email:

florian.lange@kuleuven.be, Phone: +3216376933

Szu-chi Huang is an Associate Professor of Marketing at the Graduate School of Business, Stanford University, Stanford Graduate School of Business 655 Knight Way, Stanford, CA

94305. Email: huangsc@stanford.edu Phone: 650-725-5040

Siegfried Dewitte is Professor of Marketing at the KU Leuven, Behavioral Engineering Group, Naamsestraat 69 – box 3545 3000 Leuven, Belgium. Email: Siegfried.Dewitte@kuleuven.be,

Phone: +3216326949.

Having Less, Giving More?

Two Preregistered Replications of the Relationship Between Social Class and Prosocial Behavior

Abstract: In the present report, we describe two planned direct replications of studies on the relationship between social class and prosocial behavior. In the original studies, individuals with higher socioeconomic status have been shown to behave less prosocially across a variety of domains. This finding continues to influence both research and the public debate on the psychological correlates of social class. At the same time, the validity of the original findings has been contested. Against this background, pre-registered direct replication studies with sufficient statistical power are warranted to test the robustness of these influential findings. We conducted two replication studies to provide valuable diagnostic information with regard to the relationship between social class and prosocial behavior. Our results indicate that the generalizability of the original findings may be much more limited than suggested. In addition, they highlight the need for an increased reliance on psychometrically established measures to facilitate cumulative research on the relationship between social class and prosocial behavior.

Keywords: Replication; Prosocial behavior; Social class; socioeconomic status

1. Introduction

The financial turmoil of the past years and the heightened economic inequality (Wolff, 2012; World Health Organization, 2014; OECD, 2016) have triggered an increasing interest in understanding how differences in social class can influence individual and collective behavior (Côté, House, & Willer, 2015). Research examining the psychological significance of social class (e.g., by comparing individuals who differ with regard to socioeconomic status, SES) has linked differences in social class to a variety of attitudes, behaviors, and personality traits (Kraus, Piff, & Keltner, 2009; Johnson, Richeson, & Finkel, 2011; Piff et al., 2012; Zeckhauser, Trautmann, & van de Kuilen, 2013). For example, as compared to high-SES individuals, low-SES individuals show an increased present-orientation in many domains of their everyday behavior (Pepper & Nettle, 2017) and a greater psychological focus on the most pressing issues (Shah, Mullainathan, & Shafir, 2012). While these differences may reflect contextually appropriate adaptations to the life circumstances of low-SES individuals (Pepper & Nettle, 2017), they are also related to a problematic neglect of long-term problems and future consequences, which affects individuals and human society as a whole.

One important facet of this problem is the link between SES and prosocial behavior. Inhabitants of economically deprived neighborhoods have been reported to show less prosocial behavior than individuals from high-SES neighborhoods. Adults from low-SES neighborhoods were observed to make smaller donations in dictator games (Nettle, Colléony, & Cockerill, 2011; Silva & Mace, 2014, 2015), to return fewer lost letters (Holland, Silva, & Mace, 2012; Nettle et al., 2011; Silva & Mace, 2014), to reciprocate less in a trust game (Falk & Zehnder, 2007), and to be less likely to respond to a request to participate in a survey (Nettle et al., 2011). In addition, Korndörfer, Egloff, and Schmukle (2015) found high-SES individuals to report being more charitable, helpful, and likely to volunteer than individuals

with lower SES as well as to be more cooperative in a trust game in large and representative international samples.

A highly influential empirical counterweight to this evidence was provided by Piff and colleagues (Piff, Kraus, Cheng, & Keltner, 2010) who asserted that low-SES adults were actually more prosocial across a variety of domains (see also Jordan, Amir, & Rand, 2017; Kirkpatrick, Delton, Robertson, & de Wit, 2015). Although they describe their hypothesis as “inherently paradoxical” (p. 773), they found it supported in four studies, showing negative links between SES and generosity, donations to charity, trust, and helpfulness (see Table 1). The contribution by Piff and colleagues continues to influence research in psychological science and beyond, with the number of citations steadily increasing from 2010 to 2016. With a total 227 citations (24/10/2017), it has been placed in the top 1% of the most influential papers in the domain of Psychiatry/Psychology by the Web of Science. The article’s scientific impact has been accompanied by extensive coverage in popular media outlets (e.g., <http://www.economist.com/node/16690659>; http://www.nytimes.com/2010/08/22/magazine/22FOB-wwln-t.html?_r=0).

In view of the discrepancy between Piff and colleagues’ findings of a negative relationship between SES and prosocial behavior and the substantial evidence to the contrary cited above, the reliability of the studies reported by Piff and colleagues has been questioned (Korndörfer et al., 2015). From 2014 to 2016, we have run four independent studies that aimed to replicate the finding reported by Piff and colleagues (see Table 2 for an overview of their methods and results). Inspection of Table 2 reveals that all of these studies used measures that had also been used by Piff and colleagues to assess SES (i.e., the predictor variable) and prosocial behavior (i.e., the outcome variable). Just like the studies by Piff and colleagues, our studies examined the relationship between these variables in a laboratory context (with samples that predominantly consisted of students) and in an online context (with

samples that were more diverse with regard to participants' sociodemographic background). In addition, our data-analysis approach was designed to be as similar as possible to the one described by Piff and colleagues. However, we did not find any evidence for a negative relationship between high SES and prosocial behavior in any of our studies (see Matsuba, Hart, & Atkins, 2007, for another high-powered study that did not find a substantial link between SES and prosocial behavior). None of our analyses yielded significant results (see Table 2) and the overall effect size across our studies was as small as $r = .01$, with a 95% confidence interval $[-.109 - .131]$, calculated according to the Hedges-Vecchia random effects model) that included 0, but not the mean effect size reported by Piff et al. (i.e., $r = -.215$).

However, the comparison of Table 1 and Table 2 also reveals that none of our studies can be regarded as a direct replication of one of the studies reported by Piff and colleagues, that is, in none of the studies did we directly follow the original authors in administering the same combination of predictor and outcome measures to participants from the same population. This feature of unsuccessful conceptual replications (like the ones reported by us or the ones reported by Korndörfer et al., 2015) leaves them open to explanations that make use of variations of the "hidden moderator" argument. When discussing the differences between the results presented by Piff and colleagues and Korndörfer and colleagues, Piff and Robinson (2017) argued that "[d]iscrepant findings may depend, in part, on the measures used" (p. 8). This is, of course, a viable explanation and at this moment, it is not possible to determine whether the original findings by Piff and colleagues are a false positive (as suggested by Korndörfer et al.) or the result of a very sensitive combination of measures (as suggested by Piff & Robinson).

(INSERT TABLE 1)

(INSERT TABLE 2)

Against this background, we conducted two high-powered, preregistered direct replications of studies reported in the original contribution by Piff and colleagues (2010). This approach was expected to provide diagnostic information with regard to the proposed negative effect of high SES on prosocial behavior. If we replicate the effects reported by Piff and colleagues, our studies would lend substantial support to the interpretation presented by Piff and Robinson (2017). They would speak to the existence and robustness of a negative effect of SES on prosocial behavior under carefully selected circumstances. In contrast, the absence of such an effect in direct replications of Piff and colleagues' studies would cast serious doubt not only on the generalizability of negative SES effects but also on the specific findings presented by Piff and colleagues. It would support Korndörfer and colleagues' interpretation that some sort of bias has contributed to the evidence reported by Piff and colleagues and, hence, that their data do not constitute a substantial reason to believe in a negative effect of high SES on prosocial behavior.

We selected Piff and colleagues' Study 1 and Study 2 for our preregistered attempt to replicate the proposed negative effect of high SES on prosocial behavior. Study 1 was a correlational study conducted under controlled laboratory conditions in a sample that, for the largest part, consisted of undergraduate students recruited from the University of California, Berkeley, CA. Two weeks prior to coming to the laboratory, participants reported their subjective SES in the context of an online questionnaire. Piff and colleagues found these SES scores to be negatively related to participants' prosocial allocations in a dictator game played in the laboratory. We attempted to replicate this finding in a sample recruited from the participant pool at Stanford University, Stanford, CA. Importantly, the location of our study is not only close to the location of the original Study 1, Berkeley and Stanford are also comparable with regard to social inequality (<http://www.civcdashboards.com/>), a variable that has been proposed to moderate the relationship between social class and prosocial

behavior (Côté, House, & Willer, 2015), a finding, however, that has been put into question by a more recent study (Schmukle, Korndörfer, & Egloff, 2019). The choice of our study location should thus ensure that the variance in participants' SES is comparable to SES variation in Study 1 by Piff and colleagues.

In contrast to the first study, the replicability of Study 2 by Piff and colleagues should depend to a lesser degree on the location of the study and the composition of the sample. In that study, Piff and colleagues actively manipulated participants' subjective perception of their relative social class. Participants were induced to compare themselves to people from a higher vs. lower socioeconomic background. In contrast to participants in the downward social comparison group, participants who engaged in an upward social comparison subsequently indicated that people should spend a larger percentage of their annual income on charitable donations. This outcome variable was also negatively related to an income-based measure of SES. Moreover, manipulated subjective SES and measured SES accounted for independent portions of variance in participants' support for charity. In our second study, we evaluated both these relationships as well as their independence in a sample of Belgian university students. We note that the relationship between measured SES and support for charity might be sensitive to cultural differences between Belgium and the location of the original study (i.e., "a major public university campus" in the US, Piff et al., 2000, p. 775). Finding a correlation between measured SES and prosocial behavior in Study 1 (Stanford), but not in Study 2 (Belgium) would support this notion. In contrast, we expect the effectiveness of manipulating subjective SES to be less dependent on sample characteristics because variance on the variable that Piff and colleagues proposed to be causally related to prosocial behavior is actively generated in the course of the experiment.

In sum, we tested the correlational relationship between high SES and low prosocial behavior in two samples (one very similar to the sample recruited by Piff et al. and one from a

different cultural background) and the causal relationship between these variables under controlled laboratory conditions in Belgium. These studies were expected to provide critical information with regard to the replicability and generalizability of the findings presented by Piff and colleagues. The preregistered study plan as well as all data and analysis scripts can be found at <https://osf.io/zha65/>.

2. Study 1

2.1. Participants and stopping rule

Following Piff and colleagues (2010, Study 1), we reached out to potential participants who are enlisted in the subject pool of a Californian university (in our case, Stanford University). Similar to the participant pool used by Piff and colleagues (2010), this participant pool is composed of undergraduate students and members of the broader community. To determine the target sample size, we followed the heuristic proposed by Simonsohn (2015) and multiplied the sample size of the original study (i.e., $N = 115$) with 2.5. With the resulting target sample size of 288 we had 98% power to detect an effect as large as reported by Piff and colleagues ($r = -.23$) in our study (G*Power 3.1.9.2, Faul, Erdfelder, Lang & Buchner, 2007; two-tailed, $\alpha = .05$).

We first offered as many testing slots as were required to reach the target sample size. If not all of the slots were filled, if participants canceled their participation, if they did not show up for their slot or if participants started but did not complete the study, additional slots were opened until the target sample size was reached. Only complete questionnaires without missing data for the predictor and the outcome variable were analyzed. No further participants were excluded as Piff and colleagues did not report any further exclusion criteria. We stopped data collection as soon as we noticed that the target sample size was reached. Due to conservative overbooking, we slightly exceeded the target sample size. Only after having

stopped data collection, we analyzed the data. Participants received \$10 in exchange for their participation.

2.2. Procedure

As in Piff and colleagues' study we administered an online questionnaire containing a number of demographic measures, including subjective SES, age, gender, ethnicity, and religiosity to the participants of the study. Participants received an email with the link of the questionnaire two weeks prior to visiting the lab for the completion of the prosocial behavior task (i.e., the dictator game). According to Piff and colleagues, collecting the data at separate time points should serve to reduce any potential biases that the completion of questionnaire measures might exert on behavior in the dictator game.

Upon arrival to the lab, participants were seated in front of a computer in semi-cubicles where they completed two unrelated tasks before taking part in the dictator game. All tasks were presented using the Qualtrics software. Participants first completed a picture rating task that has been used as a neutral filler task in a study by Lau, Molyneaux, Telman, and Belli (2011). Participants rated 60 emotionally neutral pictures for perceived pleasantness on a visual analogue scale. The scale ranged from 'very unpleasant' (0 cm) to 'very pleasant' (10 cm). Each picture was presented above the corresponding scale on a separate page of the Qualtrics questionnaire. This task took approximately ten minutes to complete. The second task was a word completion task that has been used in several studies in the past (Anderson, Carnagey, & Eubanks, 2003; Anderson et al., 2004). In this task, participants are given a list of word fragments that they are asked to complete. For example, they are presented with the fragment "t _ p _". Using the computer keyboard, they are then required to type in a word that exists in the English language and that can be formed by completing the presented fragment. Hence, in the example above, participants need to find a four-letter word that begins with the letter "t" and that has the letter "p" as a third letter. Valid solutions would be "type" or "tape".

Anderson and colleagues provide a list of 98 fragments that can be retrieved from public.psych.iastate.edu/caa/Scales/WordComp.pdf. We only used 48 of these fragments (which have been tested to yield neutral completions like the above example), as the remaining 50 fragments can yield completions with aggressive connotations. For example, “b _ h _ _ _” can be completed to yield neutral words such as “behalf” but also to yield an aggressive word (i.e., “behead”). By limiting ourselves to the neutral fragments (e.g., “t _ p _”), we intended to keep this filler task as neutral as possible. Participants were informed that their task was to complete as many of the word fragments as possible in ten minutes. They were instructed that it was no problem if they did not complete all fragments in that time and that, if they completed all items before the time was up, they could use the remaining time to check their answers or to relax. All 48 fragments were presented on the same page of the Qualtrics questionnaire. Next to each fragmented word, participants found a blank field to insert one of the possible solutions. After ten minutes, participants were presented with the subsequent page, which introduced the next task with the words “Very good! You will now go on to play a very quick game.” Then, they completed the dictator game as described by Piff and colleagues (2010, see below). After completing the game, participants were asked if they could imagine the purpose of the study. Finally, they were asked to briefly describe how they imagined the person they interacted with in the dictator game. Data from these questions were used to evaluate the transparency of the study and the believability of the game. However, as Piff and colleagues have not reported the results of comparable checks for Study 1, these data cannot be used to directly qualify the results of our replication attempt.

2.2.1 Subjective SES. Participants completed an online version of the MacArthur Scale of subjective SES (e.g., Adler et al., 2000; Kraus et al., 2009). This scale comprises of a ladder that assesses where participants put themselves within the hierarchy of their community. Participants were presented with a drawing of a ladder with 10 different rungs,

which represent different levels of occupational status, education, and income. Participants were called to indicate where they thought they stand compared to the other individuals in their community. Higher rungs reflect higher placement and higher SES.

2.2.2 Dictator game. Participants were informed that they had been paired with an anonymous partner who was located in a different room (Piff et al. 2010; Forsythe et al., 1994; Fowler & Kam, 2007). They were given 10 points and instructed that their task was to decide how many of the points they wished to keep for them and how many points they wanted to transfer to the anonymous partner (they were told that this amount could vary from 0 to 10 points). Participants were further informed that their cash payout depended on the amount of points that they kept for themselves and that their partner would have no strategic input into the game's outcome. Furthermore, they were told that their responses in the game would remain anonymous. Higher allocations show higher levels of pro-social behavior. After the completion of the tasks, participants were debriefed and paid the normal fee of participation (\$20 per hour) regardless of what they scored in the dictator game.

2.3. Data analysis

We ran two different analyses. The first included only the focal variables, subjective SES and dictator game scores. We ran an ordinary least squares (OLS) regression using subjective SES as predictor and dictator game scores as outcome variable. In the second analysis, we tested the effect of subjective SES on dictator game scores controlling for age, religiosity, and ethnicity (similar to Piff and colleagues' study). The replication will be regarded as successful when both analyses reveal a significant negative relationship between subjective SES and dictator game scores ($\alpha = .05$, two-sided). In addition, we tested whether the effect sizes obtained in our study significantly differ from the effect sizes reported by Piff and colleagues (using <http://vassarstats.net/rdiff.html>). Furthermore, we compared the distributions of subjective SES, dictator game scores, and demographics with the data

reported in the original study. Significant differences (in measures of dispersion or central tendency) between the studies were taken into account when interpreting the results. In case of substantial between-study differences, we conducted exploratory subgroup analyses to test whether any potential difference in the study outcomes can be attributed to these sample differences.

2.4 Results

A total sample of $N = 300$ participants completed both the short online survey and the laboratory session assessing dictator game performance. The average time delay between the two assessments was approximately 15 days.

The OLS regression analysis did not reveal a significant link between SES and dictator game performance, $\beta=0.007$; $t(298)=0.115$, $p=0.91$, 95% CI $[-0.11, 0.12]$. This relationship was significantly weaker than the one reported by Piff and colleagues, $z=-2.180$, $p=0.029$ (see also Table 6). Results did not markedly change when we included age, religiosity, and ethnicity as covariates, $\beta=0.005$; $t(295)=0.08$, $p=0.93$, 95% CI $[-0.11, 0.12]$. Similarly, the link between SES and dictator game performance was not significant ($\beta=-0.001$; $t(247)=-0.02$, $p=0.98$, 95% CI $[-0.13, 0.12]$) when we excluded 51 participants who indicated to believe that they were not actually paired with another participant during the dictator game.

Table 3 reveals a number of potentially relevant differences between our sample and the sample tested by Piff and colleagues. Participants in our study were older, $t(413)=2.70$, $p=0.007$, ranked themselves higher on the MacArthur Scale of subjective SES, $t(413)=5.56$, $p<0.001$, and donated less money in the dictator game, $t(413)=3.18$, $p=0.002$ (as indicated by independent-samples t-tests, <https://www.graphpad.com/quickcalcs/ttest2/>). In addition, age, $F(114, 299)=0.68$, $p=0.019$, and dictator game performance, $F(114, 299)=0.42$, $p<0.001$, but not subjective SES, $F(114, 299)=1.09$, $p=0.575$, showed significantly larger variation in our sample than in the sample by Piff and colleagues (as indicated by F-tests,

<http://www.statskingdom.com/220VarF2.html>). The ratio of female to male participants did not differ between studies, $\chi^2(1) = 0.002$, $p = 0.968$ (as indicated by a chi-square test with Yates correction, <https://www.graphpad.com/quickcalcs/contingency1/>).

(INSERT TABLE 3)

To explore whether the differences in sample characteristics between our sample and the sample tested by Piff and colleagues can account for the difference in study outcomes, we conducted systematic subgroup analyses using a multiverse approach (Steege et al., 2016). A multiverse analysis involves conducting the analysis of interest across a set of plausible data sets that can be constructed from the same raw data set (Steege et al., 2016). It is considered as a more extensive outlier analysis (Steege et al., 2016; Ramsey & Schafer, 2012; Simmons et al., 2011). To conduct the multiverse analysis, we constructed a number of different data sets by systematically excluding cases with extreme scores on the variables that were found to be significantly different from the study by Piff and colleagues. As participants' age was higher in our study than in theirs, we trimmed our age distribution at levels of 56, 46, and 36 years. As can be seen in Table 4, this trimming not only reduced the mean age of our sample, but also its standard deviation to be more similar to the sample tested by Piff and colleagues. Participants with high scores on the MacArthur Scale of subjective SES were excluded at three levels (10, ≥ 9 , ≥ 8), because subjective SES was found to be higher in our study. Finally, participants with minimum donations in the dictator game (i.e., 0 points) were excluded as dictator game scores were lower in our sample. Results for the association between the MacArthur Scale of subjective SES and the number of donated points in the dictator game, for all combinations of the exclusion criteria listed above, are displayed in Table 4. Subjective SES did not significantly relate to dictator game performance in any of the 32 data sets (all $p > .34$).

(INSERT TABLE 4)

3. Study 2

3.1. Participants and stopping rule

We reached out to students who were enlisted in the subject pool of the Faculty of Economic and Business at KU Leuven, Leuven, Belgium. Multiplying the sample size of Study 2 by Piff and colleagues (i.e., $N = 88$) with 2.5 yielded a target sample size of $N = 220$. Note that we used the sample size before data exclusion for this calculation which ensured that our study would have been sufficiently powered if we had to exclude a similar proportion of participants as Piff and colleagues. With a sample size of 220 we had 97% power to detect a group difference (upward social comparison vs. downward social comparison) as large as reported by Piff and colleagues ($d = 0.53$, two-tailed, $\alpha = .05$) in our study. At the same time, we had 93% power to detect a correlation of $r = -.23$ (two-tailed, $\alpha = .05$) as reported by Piff and colleagues for the relationship between measured SES and support for charity.

We first offered as many testing slots as were required to reach the target sample size. If not all of the slots were filled, if participants canceled their participation, if they did not show up for their slot, or if participants started but did not complete the study, additional slots were opened until the target sample size was reached. Only complete questionnaires without missing data for the predictor and the outcome variable were analyzed. Data from participants who correctly guessed the purpose of the study or who did not follow the instructions in the writing task that was part of the manipulation of the independent variable were excluded.

3.2. Procedure

Participants were welcomed and seated in front of the computers in semi-cubicles. They were randomly allocated to one of the two conditions of the relative social class manipulation. After that, participants received the word completion task described in Study 1 as a filler task (Anderson, Carnagey, & Eubanks, 2003; Anderson et al., 2004). We then measured the dependent variable by asking participants for their preferred allocation of money to various

purposes (including charity). Participants then answered a series of demographic questions (age, highest level of education, ethnicity¹), including a measure of family income that allowed us to determine their social class. Last, participants responded to a question about the purpose of the study, were thanked and debriefed. After the study, participants completed a number of measures for other projects. The entire laboratory session lasted approximately 30 minutes, for which participants were compensated with a payment of €6.

3.2.1 Manipulation of subjective SES. Participants took part in the manipulation of subjective SES that was used in the original paper. Participants were presented with an image of a ladder with 10 rungs. Participants were instructed to “Think of the ladder above as representing where people stand in Belgium.” Participants who were assigned to the high(low) subjective SES condition received the following instructions: “Now, please compare yourself to the people at the very bottom (top) of the ladder. These are people who are the worst (best) off—those who have the least (most) money, least (most) education, and the least (most) respected jobs. In particular, we’d like you to think about how you are different from these people in terms of your own income, educational history, and job status. Where would you place yourself on this ladder relative to these people at the very bottom (top)?”

Participants placed themselves on the ladder relative to people at the very top or bottom (10 - top rung, 1 - bottom rung). After participants had placed themselves on the ladder, they were asked to imagine themselves in a “getting acquainted interaction” with one of the people they had just thought about when placing themselves on the ladder. They were instructed to “please think about how the differences between you might impact what you would talk about, how the interaction is likely to go, and what you and the other person might say to each other.” Participants were instructed to write no more than five sentences.

¹In contrast to our pre-registration, we did not measure ethnicity due to a technical error.

3.2.2 Attitude towards charitable donations. To assess attitudes toward charitable donations, participants were asked to complete a survey assessing their views on “how you think people should spend their annual income?” Participants were instructed to indicate the percentage of their income people should spend annually on a number of expenses: food, luxury items, recreation, clothing, gifts, bills, education, travel, and charitable donations. Participants were instructed to allocate a part of 100% of total annual salary expenditures to each category. The dependent variable was the percentage of annual income that participants thought people should spend on charitable donations

3.2.3 SES measurement. To measure SES we used participants’ estimates of their annual family income. Participants rated their family income on the basis of eight categories: (1) less than €15,000, (2) \$15,001–€25,000, (3) €25,001–\$35,000, (4) €35,001–\$50,000, (5) €50,001–€75,000, (6) €75,001–€100,000, (7) €100,001–€150,000, or (8) more than €150,000.

3.3 Data analysis

3.3.1 Sample characteristics. First, we compared the distributions of all study variables with the data reported in the original study. Significant differences (in measures of dispersion or central tendency) between the studies were taken into account when interpreting the results.

3.3.2 Manipulation of subjective SES. Following Piff and colleagues, we compared the ladder rankings of participants in the high SES condition and participants in the low SES condition using an independent-samples t-test. If the manipulation is successful, participants who compare themselves with people at the bottom of the ladder (high SES condition) will place themselves in a significantly higher rank than participants who compare themselves with people at the top of the ladder (low SES condition). We also compared the effect size of the between-group difference with the effect size reported in the original study.

3.3.3 Manipulated subjective SES and charitable donations. To test whether the manipulation affected participants’ support for charity, we conducted an independent samples

t-test to assess whether the means of the percentage of the annual salary that should go to charities were significantly different between the two experimental groups. We also compared the effect size of the between-group difference with the effect size reported in the original study.

3.3.4 Measured SES and charitable donations. We analyzed the bivariate correlation (Pearson's r) between participants' family income and the dependent measure of charitable donations. To be consistent with Piff and colleagues (2010), we planned to control for ethnicity (coded as 0 for non-European American and 1 for European American) in an additional partial correlation analysis². Correlations were tested for significance ($\alpha = .05$, two-sided). The zero-order Pearson correlation coefficient was compared against the correlation coefficient reported by Piff and colleagues (using <http://vassarstats.net/rdiff.html>).

3.3.5 Test of independent effects of measured SES and manipulated subjective SES on charitable donations. Finally, we tested whether measured SES and manipulated subjective SES had independent effects on endorsements of charitable donations. We used an OLS regression analysis. We regressed charitable donations² on experimental condition (high vs. low subjective SES), measured SES, and the interaction between experimental condition and measured SES. Potential contributions of measured SES and manipulated subjective SES to the prediction of the outcome were considered independent when both predictors significantly ($\alpha = .05$) accounted for unique variance in participants' endorsements of charitable donations.

3.4 Results

We collected data from 221 students ($M_{age} = 22.89$, $SD = 4.80$; 63.8% female). One participant in the high SES condition was excluded for correctly identifying the purpose of the

²We did not perform the analysis controlling for ethnicity as due to a technical error we did not include a measure of ethnicity in the study.

study. Twenty additional participants (nine in the high SES condition and eleven in the low SES condition) were excluded for not following the instructions of the writing task described above. These exclusions resulted in a final sample size of $N = 200$. The proportion of excluded participants was similar in our study and in the study by Piff and colleagues (10% vs. 8%).

We preregistered to compare the distributions of all study variables with the data reported in the original study. However, as revealed in Table 5, Piff and colleagues did not provide distribution data for age and gender of their sample, thus not allowing for a direct comparison of distributions between samples. With regard to measured SES, participants reported a median annual family income in our study that fell into the third lowest category, while in the study by Piff and colleagues, median annual family income fell into the fourth lowest category. Note that the income scales used in their study and ours are not directly comparable due to the Euro to Dollar exchange rate being different from 1 (e.g., an annual income of €32000 ($\approx$ \$36000 at the time of the study) fell into the third lowest category in our study but into the fourth lowest category in the study by Piff and colleagues). The two samples did not differ with regard to the mean percentage of annual income that participants thought people should spend on charitable donations, $t(279) = 0.05$, $p = .96$ (as indicated by an independent-samples t-test, <https://www.graphpad.com/quickcalcs/ttest2/>). Similarly, we found no significant differences in the standard deviation of attitude towards donations to charity between the two samples, $F(80, 199) = 1.17$, $p = 0.37$ (as indicated by an F-test, <http://www.statskingdom.com/220VarF2.html>).

(INSERT TABLE 5)

Manipulation check analyses revealed that the participants primed to feel higher in subjective SES placed themselves significantly higher on the SES ladder ($N=100$, $M=6.85$, $SD=1.40$) than the participants primed to feel lower in subjective SES ($N=100$, $M=5.84$,

$SD=1.56$), $t(198)=4.82$, $p<0.001$, $d=0.65$, 95% CI [0.40, 0.97]. This size of the effect was comparable to the one reported by Piff and colleagues, $d=0.54$, 95% CI [0.08, 0.99], with large overlap between the corresponding confidence intervals.

As compared to participants in the low subjective SES condition ($M=3.20$, $SD=2.54$), participants in the high subjective SES condition indicated to think that people should give a higher proportion of their annual income to charity ($M=4.31$, $SD=3.52$), $t(198)=2.56$, $p=.01$, $d=0.36$, 95% CI [0.08, 0.64]. Note that this effect is in the opposite direction than the one reported by Piff and colleagues. Piff et al. reported that participants induced to experience a lower sense of subjective SES allocated a higher percentage of the annual salary to charitable donations ($M=4.65$) than the participants primed to feel that they belong to a higher SES rank ($M=2.95$), $d=-0.53$, 95% CI [-0.99, -0.07] (see Table 6).

(INSERT TABLE 6)

The association between measured SES and charitable donations was found to be non-significant, $r=-0.02$, $p=0.81$, 95% CI [-0.16, 0.12]. The zero-order Pearson correlation coefficient was compared (using <http://vassarstats.net/rdiff.html>) against the correlation coefficient reported by Piff and colleagues (i.e., $r(72)=-0.23$). Note that the latter coefficient is not a zero-order Pearson correlation coefficient, but a partial correlation coefficient (controlled for ethnicity). For the comparison of correlation coefficients, we assumed that the zero-order Pearson correlation between measured SES and charitable donations was $r=-.23$ in the sample of 81 participants tested by Piff and colleagues. The comparison indicated that the correlation was weaker in our study than in the study by Piff and colleagues, but not to a significant extent, $z=-1.60$, $p=0.11$.

Regressing charitable donations on measured SES (reported annual family income), manipulated subjective SES (1 = high SES, 2 = low SES), and the interaction of measured SES and manipulated subjective SES yielded a significant main effect of manipulated

subjective SES, $b=-1.18$, $t(196)=-2.70$, $p=0.01$, $CI [-2.04, -0.32]$. Neither the main effect of measured SES $b=-0.12$, $t(196)=1.54$, $p=0.12$, $CI [-0.35, 0.11]$, nor the interaction term, $b=-0.44$, $t(196)=-1.86$, $p=0.06$, $CI [-0.90, 0.03]$, were statistically significant.

4. General Discussion

In two high-powered and preregistered replication studies, we did not find evidence in support of the relationship between social class and prosocial behavior reported by Piff et al. (2010). In our first study, we closely mimicked the procedure of Piff et al.'s Study 1 by measuring SES and dictator game performance in Californian university students. The study revealed a non-significant correlation between SES and dictator game performance. Our second study, a close replication of Piff et al.'s Study 2 in a West European context, revealed a similar non-significant correlation between SES and participants' attitude towards charitable donations. This study also involved a manipulation of participants' perception of their relative SES. In contrast to Piff and colleagues, we did not find a negative effect of priming participants to identify with a higher SES on their attitude towards charitable donations. Instead, participants primed in this way even indicated that people should give a higher proportion of their annual income to charity.

In addition to these preregistered analyses, we added a set of exploratory analyses to further test the robustness of our findings. Because some sample characteristics differed between our first study and the corresponding original study, we gradually pruned our sample to make it more similar to the one tested by Piff and colleagues. We then repeated our main analysis for each of the samples created that way, following the procedure of a multiverse analysis (Steege et al. 2016). In all of these pruned samples, sample size exceeded that of the original study. In none of these analyses did we find a negative correlation between SES and prosocial behavior. Note that we have not been able to subject the data from our second study to similar robustness analyses because specific information about the sample of the original

study was lacking. We also note that we failed to measure ethnicity due to a technical error in that study. As a corollary, we were not able to follow Piff and colleagues in examining the correlation between measured SES and charitable donations when controlling for ethnicity. However, due to the unclear role of ethnicity for the relationship of interest in the original sample and the unclear relation between ethnicity in US vs. European samples, we do not think that the lack of this additional analysis limits the interpretability of our results.

The null correlations between SES and prosocial behavior are relevant for the debate between Piff and colleagues, who advocate a negative relation between social class and prosocial behavior, and those who doubt such a relation. Piff and Robinson (2017) suggest that discrepancies in the empirical literature might result from methodological differences. The main aim of our investigation was to put this claim to a fair test. Although we closely followed and replicated the procedures of the original studies, we did not find a negative correlation between social class and prosocial behavior. Hence, it does not seem likely that the relationship observed in Piff et al. (2010), or the lack thereof, would be due to the use of a particular combination of methods.

It is important to note that the possibility remains that differences in factors other than the study materials account for our failure to replicate the results of Piff and colleagues (2010). For example, the population studied by Piff and colleagues might differ from the population studied by us in ways that are relevant for the association of social class and prosocial behavior. This could explain the lack of a relationship in the Belgian sample in our Study 2, but this argument is weakened by the fact that the correlation was not larger in Study 1, where we followed Piff and colleagues in testing Californian university students. Similarly, we did not observe a negative correlation between social class and prosocial behavior when we pruned the sample of Study 1 to approximate the sample characteristics reported by Piff and colleagues.

Other moderators that could account for the difference between our results and the data reported by Piff and colleagues include time or circumstantial factors, which happened to be present in Piff et al.'s studies, but not in ours. The influence of such circumstantial factors (e.g., air pressure, temperature, NO₂ density in the room, color of the table, height of the building, voice pitch of the experimenter) might often go undetected as they do not play a prominent role in the development of theoretical frameworks. There is no way for us to know if such factors play a meaningful moderating role for the relationship between social class and prosocial behavior, because it is not standard practice to specify and register, let alone report such background variables. An increased focus on the study environment (as shown, for example, by ecological psychologists, Barker, 1968; Gibson, 1979; and behavior analysts, Skinner, 1953) may be of tremendous value to increase researchers' insights into the factors that make findings externally valid (e.g., generalizable to other situations, Lynch, 1999).

Irrespective of whether population, time, circumstantial moderators, or just noise may explain why Piff and colleagues observed a relationship that we did not, it seems reasonable to conclude that the generalizability of their findings is much more limited than suggested by their conclusions. The question then arises to what extent our findings may bear on the broader question of whether social class and prosociality are related at all. Our findings add to a mixed research literature involving reports of negative correlations (e.g., Jordan et al. 2017), positive correlations (e.g., Korndörfer, et al. 2015), and null correlations (e.g., Matsuba et al., 2007) between social class and prosocial behavior. In view of the methodological heterogeneity across these studies, we call for further systematic and cumulative examinations of the relationship between social class and prosocial behavior. A promising approach might be to first establish the robustness of the relationship under specifiable conditions, before gradually enlarging its scope in the course of systematic follow-up studies.

Increasing the robustness of research on the relationship between social class and prosocial behavior will also require a shift towards psychometrically established measures. Many studies in the literature rely on single-item/single-trial measures of unknown construct validity. Dictator game performance and support for charitable donations, for example, are unlikely to capture more than a few percent of interindividual variance in prosociality. In a study by Baumert, Schlösser, & Schmitt (2014), participants' dictator game offers and real-life donations were correlated as weakly as $r = .09$. If such measures do not reliably distinguish between individuals based on their prosociality, it is unreasonable to expect them to robustly relate to individual differences in another variable (such as social class). A recent study on the development of a measurement model for prosociality concluded "that further work to improve psychometric properties of several indicators is strongly required" (Wilhelm, Kaltwasser, & Hildebrandt, 2018, p. 498). It is only by addressing such assessment-related issues that we can expect future research to yield conclusive results about the relationship between social class and prosocial behavior.

Literature

- Adler, N. E., Epel, E. S., Castellazzo, G., & Ickovics, J. R. (2000). Relationship of subjective and objective social status with psychological and physiological functioning: Preliminary data in healthy White women. *Health Psychology*, 19, 586–592.
- Anderson, C.A., Carnagey, N.L., & Eubanks, J. (2003). Exposure to violent media: The effects of songs with violent lyrics on aggressive thoughts and feelings. *Journal of Personality and Social Psychology*, 84, 960–971.
- Anderson, C. A., Carnagey, N. L., Flanagan, M., Benjamin, A. J., Eubanks, J., & Valentine, J. C. (2004). Violent video games: Specific effects of violent content on aggressive thoughts and behavior. *Advances in Experimental Social Psychology*, 36, 199–249.
- Balliet, D., Parks, C., & Joireman, J. (2009). Social values orientation and cooperation in social dilemmas: A meta-analysis. *Group Processes & Intergroup Relations*, 12, 533–547.
- Baron, R. M., & Kenny, D. A. (1986). The moderator—mediator variable distinction in social psychology research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, 51, 1173–1182.
- Berg, J., Dickhaut, J., & McCabe, K. (1995). Trust, reciprocity, and social history. *Games and Economic Behavior*, 10, 122–142.
- Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's Mechanical Turk: A new source of inexpensive, yet high-quality, data?. *Perspectives on psychological science*, 6(1), 3–5.
- Côte S., House J. & Willer R. (2015). High economic inequality leads higher-income individuals to be less generous. *Proceedings of the National Academy of Science*, 112, 15828–15843.
- Falk, A., & Zehnder, C. (2007). *Discrimination and in-group favoritism in a citywide trust experiment*. Working Paper, Institute for Empirical Research in Economics. Zurich.
- Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior research methods*, 39(2), 175–191.
- Forsythe, R., Horowitz, J. L., Savin, N. E., & Sefton, M. (1994). Fairness in simple bargaining experiments. *Games and Economic Behavior*, 6, 347–369.
doi:10.1006/game.1994.1021
- Fowler, J. H., & Kam, C. D. (2007). Beyond the self: Social identity, altruism, and political participation. *The Journal of Politics*, 69, 813–827.
- Griskevicius, V., Delton, A. W., Robertson, T. E., and Tybur, J. M. (2011). Environmental contingency in life history strategies: The influence of mortality and socioeconomic status on reproductive timing. *Journal of Personality and Social Psychology*, 100, 241–254.
- Holland, J., Silva, A. S., & Mace, R. (2012). Lost letter measure of variation in altruistic behaviour in 20 neighbourhoods. *PLoS one*, 7(8), e43294.
- Johnson, S. E., Richeson, J. A., & Finkel, E. J. (2011). Middle class and marginal? Socioeconomic status, stigma, and self-regulation at an elite university. *Journal of Personality and Social Psychology*, 100, 838–852.
- Jordan, Amir, D. & Rand (2017). Childhood socioeconomic status has lasting effects on social, risk, and time preferences. Manuscript submitted for publication.
- Kirkpatrick M, Delton AW, de Wit H, et al. (2015) Prosocial effects of MDMA: a measure of generosity. *J Psychopharmacol* 29: 661–668
- Kraus, M. W., Piff, P. K., & Keltner, D. (2009). Social class, the sense of control, and social explanation. *Journal of Personality and Social Psychology*, 97, 992–1004.

- Kraus, M. W., & Keltner, D. (2009). Signs of socioeconomic status: A thin-slicing approach. *Psychological Science*, 20, 99–106.
- Korndörfer, M., Egloff, B., & Schmukle, S. C. (2015). A large scale test of the effect of social class on prosocial behavior. *PloS one*, 10(7), e0133193.
- Lynch, J. G. (1999). Theory and external validity. *Journal of the Academy of Marketing Science*, 27(3), 367-376.
- Matsuba, M. K., Hart, D., & Atkins, R. (2007). Psychological and social-structural influences on commitment to volunteering. *Journal of research in personality*, 41(4), 889-907.
- Nettle, D., Colléony, A., & Cockerill, M. (2011). Variation in cooperative behaviour within a single city. *PloS one*, 6(10), e26922.
- OECD (2016). Income inequality remains high in the face of weak recovery. Center for opportunity and equality.
- Pepper, G. V., & Nettle, D. (2017). The behavioural constellation of deprivation: causes and consequences. *Behavioral and Brain Sciences*, 1-72.
- Piff, P. K., Kraus, M. W., Côté, S., Cheng, B. H., & Keltner, D. (2010). Having less, giving more: the influence of social class on prosocial behavior. *Journal of personality and social psychology*, 99(5), 771-784.
- Piff, P. K., & Robinson, A. R. (2017). Social class and prosocial behavior: current evidence, caveats, and questions. *Current Opinion in Psychology*, 18, 6-10.
- Ramsey, F., Schafer, D. (2012). The statistical sleuth: A course in methods of data analysis (3rd ed.). Stamford, CT: Cengage Learning.
- Safra, L., Tecu, T., Lambert, S., Sheskin, M., Baumard, N., & Chevallier, C. (2016). Neighborhood Deprivation Negatively Impacts Children's Prosocial Behavior. *Frontiers in psychology*, 7.
- Shah, A. K., Mullainathan, S., & Shafir, E. (2012). Some consequences of having too little. *Science*, 338(6107), 682-685.
- Silva, A. S., & Mace, R. (2014, October). Cooperation and conflict: field experiments in Northern Ireland. In *Proc. R. Soc. B* (Vol. 281, No. 1792, p. 20141435). The Royal Society.
- Silva, A. S., & Mace, R. (2015). Inter-group conflict and cooperation: Field experiments before, during and after sectarian riots in Northern Ireland. *Frontiers in Psychology*, 6.
- Simmons, J. P., Nelson, L. D., Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, 22, 1359–1366.
- Simonsohn, U. (2015). Small telescopes: Detectability and the evaluation of replication results. *Psychological Science*, 26(5), 559-569.
- Schmukle, S. C., Korndörfer, M., and Egloff, B. (2019). No evidence that economic inequality moderates the effect of income on generosity. *Proc. Natl. Acad. Sci. U.S.A.* 116, 9790–9795.
- Steege, S., Tuerlinckx, F., Gelman, A., & Vanpaemel, W. (2016). Increasing transparency through a multiverse analysis. *Perspectives on Psychological Science*, 11, 702–712.
- Trautmann, S. T., Kuilen, G. v. d., & Zeckhauser, R. J. (2013). Social Class and (Un)Ethical Behavior A Framework, With Evidence From a Large Population Sample. *Perspectives on Psychological Science*, 8(5), 487-497.
- Van Lange, P. A. M. (1999). The pursuit of joint outcomes and equality in outcomes: An integrative model of social value orientation. *Journal of Personality and Social Psychology*, 77, 337–449.
- WHO, World Health Statistics 2014.
- Wolff, E. (2012). The Asset Price Meltdown and the Wealth of the Middle Class. National Bureau Of Economic Research. Nber Working Paper Series. Working Paper 18559.

Table 1.

Overview of the studies reported by Piff and colleagues (2010)

	Predictor (i.e., operationalization of social class)	Outcome (i.e., operationalization of prosocial behavior)	Sample size and characteristics	<i>p</i>	Effect size in <i>r</i>
Study 1	Online version of MacArthur Scale of subjective SES (Adler et al., 2000)	Dictator game	$N = 115$ (undergraduates) $M_{\text{age}} = 22.57$, $SD = 7.95$	.01	-.23
Study 2	Manipulation of social class using ladder from MacArthur Scale	Percentage of annual income that should be given to charities	$N = 81$ (undergraduates) Median _{income} = between \$35.000 and \$50.000	<.05	-.24 (averaged across the two predictors)
	Annual Family Income (8-point scale)			<.05	
Study 3	Annual Family Income (8-point scale) + Level of Education (4-point)	Trust Game (Berg, Dickhaut, & McCabe, 1995)	$N = 155$ (online sample) $M_{\text{age}} = 37$ Median _{income} = between \$25.000 and \$35.000	<.05	-.18
Study 4	Childhood Family Income (7-point scale) + Current Annual Family Income	Helping task (Chen et al., 2001)	$N = 102$ (students and adults) $M_{\text{age}} = 21.64$, $SD = 3.05$ $M_{\text{income}} = 5.86$, $SD = 2.06$	<.01	-.23
Total					-.22 95%CI: [.13 - .30]

Note. Mean effect sizes have been calculated according to the Hedges-Vecchia random effects model.

Table 2.

Overview of the studies conducted from our group between 2014 and 2016

	Predictor (i.e., operationalization of social class)	Outcome (i.e., operationalization of prosocial behavior)	Sample size and characteristics	<i>p</i>	Effect size in <i>r</i>
Study 1	Annual Family Income (6-point scale)	Percentage of annual income that should be given to charities (Piff et al., study 2)	<i>N</i> = 176 (Mturk, online) <i>M</i> _{age} = 36.46, <i>SD</i> = 12.45	.52	-.05
Study 2	Piff et al. 2012 (participants reported their childhood and current social class by rating their agreement with five statements on a 7-point scale (e.g., Griskevicius, Delton, et al., 2011; 1 Strongly disagree, 7 Strongly agree).	Percentage of annual income that should be given to charities	<i>N</i> = 202 (Mturk, online) <i>M</i> _{age} = 22.47, <i>SD</i> = 5.8	.10	.12
Study 3	Piff et al. 2012	Percentage of annual income that should be given to charities	<i>N</i> = 85 (faculty pool, lab) <i>M</i> _{age} = 37.87, <i>SD</i> = 11.96	.41	.09
Study 4	Annual Family Income (6-point scale)	Trust Game (Berg, Dickhaut, & McCabe, 1995)	<i>N</i> = 68 (faculty pool, lab) <i>M</i> _{age} = 22.65, <i>SD</i> = 4.1	.17	.17
Total					.01 95% CI [-.11 -.13]

Table 3: Descriptive Statistics

	Piff et al.	Study 1
SES (rank)	$M=5.82, SD=1.47$	$M=6.69, SD=1.41$
Dictator Game	$M=4.14, SD=1.91$	$M=3.20, SD=2.94$
Age	$M=22.57, SD=7.95$	$M=25.29, SD=9.61$
Gender	63.5% female, 33% male, 3.5% preferred not say	65.3%female, 34.3% male, 0.3% preferred not say

Table 4: Multiverse Analysis

Dictator=Full Range				Dictator>0				
Age=Full Range	Age<56	Age<46	Age<36	Age=Full Range	Age<56	Age<46	Age<36	
N=300 $\beta=.007$, $p=.91$ $M_{SES}=6.69$, $SD=1.41$ $M_{dict}=3.20$, $SD=2.94$ $M_{age}=25.29$, $SD=9.61$	N=291 $\beta=.018$, $p=.755$ $M_{SES}=6.71$, $SD=1.40$ $M_{dict}=3.16$, $SD=2.96$ $M_{age}=24.22$, $SD=7.51$	N=281 $\beta=.010$, $p=.862$ $M_{SES}=6.70$, $SD=1.42$ $M_{dict}=3.12$, $SD=2.93$ $M_{age}=23.24$, $SD=5.51$	N=268 $\beta=.017$, $p=.783$ $M_{SES}=6.72$, $SD=1.43$ $M_{dict}=3.13$, $SD=2.95$ $M_{age}=22.51$, $SD=4.47$	N=193 $\beta=-.042$, $p=.561$ $M_{SES}=6.73$, $SD=1.40$ $M_{dict}=4.97$, $SD=2.96$ $M_{age}=25.84$, $SD=10.41$	N=185 $\beta=-.047$, $p=.524$ $M_{SES}=6.77$, $SD=1.40$ $M_{dict}=4.97$, $SD=2.18$ $M_{age}=24.34$, $SD=7.63$	N=178 $\beta=-.062$, $p=.414$ $M_{SES}=6.75$, $SD=1.42$ $M_{dict}=4.93$, $SD=2.14$ $M_{age}=23.29$, $SD=5.55$	N=170 $\beta=-.072$, $p=.348$ $M_{SES}=6.79$, $SD=1.43$ $M_{dict}=4.94$, $SD=2.18$ $M_{age}=22.60$, $SD=4.63$	SES=Full Range
N=297 $\beta=.017$, $p=.772$ $M_{SES}=6.66$, $SD=1.37$ $M_{dict}=3.21$, $SD=2.95$ $M_{age}=25.33$, $SD=9.65$	N=288 $\beta=.029$, $p=.626$ $M_{SES}=6.68$, $SD=1.37$ $M_{dict}=3.17$, $SD=2.97$ $M_{age}=24.25$, $SD=7.54$	N=278 $\beta=.021$, $p=.731$ $M_{SES}=6.66$, $SD=1.39$ $M_{dict}=3.13$, $SD=2.93$ $M_{age}=23.26$, $SD=5.53$	N=265 $\beta=.028$, $p=.653$ $M_{SES}=6.68$, $SD=1.34$ $M_{dict}=3.15$, $SD=2.96$ $M_{age}=22.52$, $SD=4.49$	N=191 $\beta=-.020$, $p=.781$ $M_{SES}=6.69$, $SD=1.37$ $M_{dict}=4.99$, $SD=2.14$ $M_{age}=25.88$, $SD=10.45$	N=183 $\beta=-.025$, $p=.736$ $M_{SES}=6.73$, $SD=1.36$ $M_{dict}=4.99$, $SD=2.20$ $M_{age}=24.37$, $SD=7.66$	N=176 $\beta=-.040$, $p=.522$ $M_{SES}=6.72$, $SD=1.38$ $M_{dict}=4.95$, $SD=2.14$ $M_{age}=23.31$, $SD=5.57$	N=168 $\beta=-.050$, $p=.518$ $M_{SES}=6.75$, $SD=1.39$ $M_{dict}=4.96$, $SD=2.18$ $M_{age}=22.61$, $SD=4.65$	SES<10
N=281 $\beta=-.020$, $p=.774$ $M_{SES}=6.53$, $SD=1.29$ $M_{dict}=3.15$, $SD=2.95$ $M_{age}=25.59$, $SD=9.85$	N=272 $\beta=-.009$, $p=.889$ $M_{SES}=6.54$, $SD=1.28$ $M_{dict}=3.11$, $SD=2.96$ $M_{age}=24.45$, $SD=7.70$	N=262 $\beta=-.021$, $p=.735$ $M_{SES}=6.52$, $SD=1.30$ $M_{dict}=3.06$, $SD=2.93$ $M_{age}=23.41$, $SD=5.64$	N=249 $\beta=-.014$, $p=.830$ $M_{SES}=6.53$, $SD=1.31$ $M_{dict}=3.08$, $SD=2.95$ $M_{age}=22.63$, $SD=4.58$	N=177 $\beta=-.014$, $p=.854$ $M_{SES}=6.51$, $SD=1.25$ $M_{dict}=5.01$, $SD=2.12$ $M_{age}=26.30$, $SD=10.73$	N=169 $\beta=-.020$, $p=.796$ $M_{SES}=6.54$, $SD=1.24$ $M_{dict}=5.00$, $SD=2.15$ $M_{age}=24.68$, $SD=7.87$	N=162 $\beta=-.040$, $p=.610$ $M_{SES}=6.51$, $SD=1.26$ $M_{dict}=4.96$, $SD=2.12$ $M_{age}=23.54$, $SD=5.72$	N=162 $\beta=-.052$, $p=.521$ $M_{SES}=6.54$, $SD=1.27$ $M_{dict}=4.97$, $SD=2.16$ $M_{age}=22.79$, $SD=4.77$	SES<9
N=210 $\beta=-.011$, $p=.874$ $M_{SES}=6.03$, $SD=1.12$ $M_{dict}=3.19$, $SD=2.84$ $M_{age}=25.73$, $SD=10.1$	N=203 $\beta=-.006$, $p=.927$ $M_{SES}=6.04$, $SD=1.11$ $M_{dict}=3.12$, $SD=2.86$ $M_{age}=24.51$, $SD=7.78$	N=195 $\beta=-.023$, $p=.750$ $M_{SES}=6.02$, $SD=1.12$ $M_{dict}=3.08$, $SD=2.84$ $M_{age}=23.42$, $SD=5.68$	N=184 $\beta=-.012$, $p=.870$ $M_{SES}=6.02$, $SD=1.14$ $M_{dict}=3.09$, $SD=2.87$ $M_{age}=22.51$, $SD=4.40$	N=134 $\beta=-.037$, $p=.672$ $M_{SES}=6.04$, $SD=1.06$ $M_{dict}=4.99$, $SD=1.91$ $M_{age}=26.69$, $SD=8.17$	N=127 $\beta=-.050$, $p=.577$ $M_{SES}=6.06$, $SD=1.06$ $M_{dict}=4.98$, $SD=1.93$ $M_{age}=24.80$, $SD=11.36$	N=121 $\beta=-.061$, $p=.504$ $M_{SES}=6.01$, $SD=1.06$ $M_{dict}=4.97$, $SD=1.90$ $M_{age}=23.50$, $SD=5.85$	N=114 $\beta=-.076$, $p=.423$ $M_{SES}=6.03$, $SD=1.08$ $M_{dict}=4.99$, $SD=1.94$ $M_{age}=22.61$, $SD=4.68$	SES<8

Table 5: Descriptive Statistics

	Piff et al.	Study 2
SES (income)	<i>Median</i> = \$35,001–\$50,000	<i>Median</i> = €25,001–€35,000
Donations	<i>M</i> =3.78, <i>SD</i> =3.37	<i>M</i> =3.76, <i>SD</i> =3.11
Age	not reported	<i>M</i> =22.84, <i>SD</i> =4.76
Gender	not reported	64.5%female, 35.5% male

Table 6: Comparison of results

	Piff et al.	Present Study
Study 1	$\beta=-0.23$; $t(113)=-2.52$, $p=0.01$, 95% <i>CI</i> [-.40, -.05]	$\beta=0.007$; $t(298)=0.115$, $p=0.91$, 95% <i>CI</i> [-0.11, 0.12]
Study 2	$t(76)= -2.29$, $p<.05$, $d=-0.53$, 95% <i>CI</i> [-0.99, -0.07], $M_{\text{HighSES}}=2.95$, $SD=\text{Not reported}$ $M_{\text{LowSES}}=4.65$, $SD=\text{Not reported}$	$t(198) = 2.56$, $p = .01$, $d = 0.36$, 95% <i>CI</i> [0.08, 0.64], $M_{\text{HighSES}}=4.31$, $SD=3.52$ $M_{\text{LowSES}}=3.20$, $SD=2.54$