ECONOMIC GAMES, (DIS)HONESTY AND TRUST

EDITED BY: Nikolaos Georgantzis, Tarek Jaber-Lopez and Ismael Rodriguez-Lara

PUBLISHED IN: Frontiers in Psychology

frontiers Research Topics

frontiers

ISSN 1664-8714
ISBN 978-2-88974-612-5
DOI 10.3389/978-2-88974-612-5

About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.
Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: frontiersin.org/about/contact

ECONOMIC GAMES, (DIS)HONESTY AND TRUST

Topic Editors:
Nikolaos Georgantzis, Burgundy School of Business, France
Tarek Jaber-Lopez, Université Paris Nanterre, France
Ismael Rodriguez-Lara, University of Granada, Spain

Citation: Georgantzis, N., Jaber-Lopez, T., Rodriguez-Lara, I., eds. (2022). Economic Games, (Dis)honesty and Trust. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-88974-612-5

Table of Contents

frontiers in Psychology

EDITORIAL
published: 09 February 2022
doi: 10.3389/fpsyg.2022.853653

Editorial: Economic Games, (Dis)honesty and Trust

Nikolaos Georgantzis[1,2]*, *Tarek Jaber-Lopez*[3] and *Ismael Rodriguez-Lara*[4,5]

[1] Burgundy School of Business, Dijon, France, [2] Laboratorio de Economia Experimental, Universitat Jaume I Castellon, Castellón de la Plana, Spain, [3] EconomiX, University Paris Nanterre, UPL, Paris, France, [4] Teoría e Historia Económica, Campus de la Cartuja, University of Granada, Granada, Spain, [5] Teoría e Historia Económica, Campus El Ejido, University of Malaga, Malaga, Spain

Keywords: trust, honesty and dishonesty, prosocial behavior, anti-social behavior, trustworthiness

Editorial on the Research Topic

Economic Games, (Dis)honesty and Trust

Trust is a central source of well-being in a society. When individuals feel that they can trust others, cooperative interactions become more likely, making a group of individuals able to enjoy better outcomes than the sum of individual stand-alone efforts would achieve. Opportunistic and dishonest behavior hinders trust by generating negative feedback to trusting behavior. In this Research Topic we collect cutting edge research on pro-social behavior, trust, and (dis)honesty. Below, we offer a brief discussion of the article included, under two general headings: (i) trust and trustworthiness and (ii) dishonesty and opportunistic behavior.

TRUST AND TRUSTWORTHINESS

Does the emergence of a crisis mitigate or substitute people's concerns regarding social issues? Blanco et al. suggest that donations aimed at addressing other social concerns are partially substituted by donations to COVID-19 funds. Yet, this substitution does not fully replace all other social concerns. Trusting the charitable organization is the most important factor to explain donations to a charity. These findings imply that the COVID-19 pandemic may substitute other social concerns, highlighting the importance of trust toward charitable institutions.

Which other societal factors foster trust in a society? Three contributions address the role of societal factors like culture, inequality, and social class in the emergence of trust. Rodrigo-González et al., find that inequality is an important explanatory factor of trust. In a trust game, trustors send more to those who have a higher endowment, probably under the belief that better performing people are more trustworthy. Trustees reciprocate more toward trustors who are richer when their money is determined by their effort. There is also evidence that trustees reciprocate more when they observe the history of decisions, and particularly trustor accumulated profits from past actions. Zylbersztejn et al. employ the hidden action game in Charness and Dufwenberg (2006) in two different locations, France and Japan. In both settings, observers are asked to predict the behavior of trustees in the hidden action game, after watching a mugshot picture or a muted video of the trustees, making a non-strategic statement independent of the hidden action game, or a loaded video in which the trustee made a strategic pre-play statement in front of the trustors. Their results suggest that observers account for morphological traits of the trustees and this bias persists across cultures. They also show that cultural distance is not *per se* helpful or detrimental for predicting trustworthiness. Rather, it affects ways in which people exploit observable information in social interactions. Finally, Qiang et al. find that social class may affect trust, but they also show that a social class-specific perception of control may be a mediating psychological mechanism in the association between social class and trust beliefs. Specifically, members of the upper social class are inclined to perceive high control over their outcomes, and they have a strong trust in daily life,

OPEN ACCESS

Edited and reviewed by:
Luis F. Martinez, Universidade NOVA de Lisboa, Portugal

***Correspondence:**
Nikolaos Georgantzis
nick.georgantzis@gmail.com

Specialty section:
This article was submitted to Personality and Social Psychology, a section of the journal Frontiers in Psychology

Received: *12 January 2022*
Accepted: *19 January 2022*
Published: *09 February 2022*

Citation:
Georgantzis N, Jaber-Lopez T and Rodriguez-Lara I (2022) Editorial: Economic Games, (Dis)honesty and Trust. Front. Psychol. 13:853653. doi: 10.3389/fpsyg.2022.853653

while members of the lower social class are more likely to feel a low sense of control, and in turn, low social trust. Focusing on another individual driver of cooperative behavior and trust, in a lab-in-the-field experiment with prison inmates, Balafoutas et al. investigate whether there is a connection between psychopathy and pro-sociality. They find that psychopathy correlates with anti-social behavior in its various forms, like weaker reciprocity to trust (trustworthiness), lower cooperation, lower benevolence, and more bribing.

DISHONESTY AND OPPORTUNISTIC BEHAVIOR

In order to improve our understanding of the determinants of cheating behavior and expectations about it, the following contributions address the role of the emotional state, gender and the environment in the emergence of dishonest behavior.

Medai and Noussair induce emotional states to participants by asking them to watch a video prior to rolling a die. The authors consider two different treatments, depending on whether or not the video induces a positive emotional state (Happiness) or does not have any effect on emotional state (Neutral). The main result of their paper is that the level of dishonesty (opportunistic misreporting of the die rolling task) is lower in the Happiness treatment, compared with the Neutral treatment. They further argue that there are no differences in lying behavior when looking at the behavior of men and women. A further examination of gender differences in lying behavior is pursued by Muñoz García et al.. In their article, they employ a modified die-under-the-cup task, in which the experimenter can observe the real distribution of the rolls. They find gender differences in cheating behavior in that women are satisfied with lower earnings than men. The frequency of radically dishonest subjects (those who did not even roll the die) is larger among men, while the proportion of "lucky honest" (rolling, but misreporting) is larger among women. Gender differences are also reported by Monzani et al., who study the drivers of anti-social behavior among entrepreneurs. Their results revealed that displaying authentic leadership reduced the likelihood of entrepreneurs (vs. managers) and men (vs. women) of engaging in antisocial behaviors such as lying to harm one's competition or seeking an unfair advantage by cheating.

Pascual-Ezama et al. find that different types of cheaters exhibit different abilities to detect unethical behavior. In their online experiment, participants are shown videos from Golder Balls, one of the most popular TV shows in the UK and they are asked to predict whether or not contestants will be dishonest. Their participants do not beat randomness in detecting dishonest behavior, but some types of cheaters are better at detecting honesty than others. The authors also highlight the importance of (non-)verbal cues and information to detect unethical behavior, and provide evidence of a "preconceived honesty bias" (i.e., people tend to think that honesty prevails). Chapkovski et al. in a sequential version of the die-rolling task, find that the likelihood to cheat increases in a "collaborative" setting, in comparison with an individual one. As the game is repeated across 45 rounds, participants become more dishonest over time in the collaborative treatment, whereas there is no such trend in the individual condition.

In a tax-evasion experiment, Du et al. randomly assigned a gross income to be declared to a central tax authority. One of the subjects in the group is randomly selected in each round to be audited. If the subject has misreported his/her income, then she will need to pay a fine. A whistleblowing mechanism is shown to be effective in both curbing tax evasion and improving the precision of tax auditing. In addition, the authors find no evidence of spillover effects of whistleblowing on ingroup cooperation in the subsequent generalized gift exchange game.

Finally, in a theoretical contribution, Spiegelman addresses academic dishonesty in the presence of open data practices. A signaling model is presented to show that both high- and low-quality results may be published in both open and closed data regimes, but open data is favored by high-quality results. A measure of "science welfare" is proposed, to show that open data will always improve the aggregate state of knowledge.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

REFERENCES

Charness, G., and Dufwenberg, M. (2006). Promises and partnership. *Econometrica* 74, 1579–1601. doi: 10.1111/j.1468-0262.2006.00719.x

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychology

ORIGINAL RESEARCH
published: 14 May 2021
doi: 10.3389/fpsyg.2021.650032

Collaborative Settings Increase Dishonesty

Youhong Du[1,2,3†], Weina Ma[4†], Qingzhou Sun[5] and Liyang Sai[1,2,3]**

[1]Center for Cognition and Brain Disorders, The Affiliated Hospital of Hangzhou Normal University, Hangzhou, China, [2]Institute of Psychological Science, Hangzhou Normal University, Hangzhou, China, [3]Zhejiang Key Laboratory for Research in Assessment of Cognitive Impairments, Hangzhou, China, [4]Department of Education, Institute of Psychological Sciences, Hangzhou Normal University, Hangzhou, China, [5]School of Management, Zhejiang University of Technology, Hangzhou, China

The present study examines whether collaborative situations make individuals more dishonest in face-to-face settings. It also considers how this dishonesty unfolds over time. To address these questions, we employed a sequential dyadic die-rolling task in which two participants in a pair sitting face-to-face received a payoff only if both reported the same outcome when each one rolled their die. In each trial, one participant (role A) rolled a die first and reported the outcome. Then, the second participant (role B) was informed of A's reported number, rolled a die as well, and reported the outcome. If their reported outcomes were identical, both of them received a reward. We also included an individual condition in which an individual subject rolled a die twice and received a reward if he/she reported the same die-roll outcome. We found that B lied significantly more than participants in the individual condition, whereas A lied as much as participants in the individual condition. Furthermore, when collaborating, more and more participants (both A and B) became dishonest as the game progressed, whereas there was no such trend among participants in the individual condition. These findings provide evidence indicating that collaborative settings increase dishonesty and that this effect becomes more evident as the collaboration progress.

Keywords: collaborative settings, dishonesty, die-rolling task, cooperation, deception

OPEN ACCESS

Edited by:
Tarek Jaber-Lopez, Université Paris Nanterre, France

Reviewed by:
Ori Weisel, Tel Aviv University, Israel
Xiuyan Guo, East China Normal University, China

****Correspondence:***
Liyang Sai
liyangsai@hznu.edu.cn
Qingzhou Sun
sunqingzhou2008@163.com

[†]These authors have contributed equally to this work and share first authorship

Specialty section:
This article was submitted to Personality and Social Psychology, a section of the journal Frontiers in Psychology

Received: *06 January 2021*
Accepted: *21 April 2021*
Published: *14 May 2021*

Citation:
Du Y, Ma W, Sun Q and Sai L (2021) Collaborative Settings Increase Dishonesty. Front. Psychol. 12:650032. doi: 10.3389/fpsyg.2021.650032

INTRODUCTION

Cooperation is essential to humans because it allows them to perform tasks more effectively and to develop trust (Kramer, 1999; Rempel et al., 2001). It also helps them to build relationships with one another (Bazerman et al., 2000; Kameda et al., 2005). For these reasons, individuals tend to prefer cooperation over working alone (Rand, 2017). However, in some situations, cooperation can involve violating certain moral rules. For example, corruption typically arises when people work together to obtain profits illegally (Gross et al., 2018). In a moral dilemma like this, will individuals be more inclined to cooperate and break moral rules or not to cooperate and obey them (e.g., honesty)?

Weisel and Shalvi (2015) were the first to examine the dishonesty of individuals in collaborative situations. They conducted an experiment involving sequential dyadic die-rolling, in which two participants were paid according to whether they reported the same number after rolling dice sequentially. Since the rolls were private, participants could misreport their actual outcomes. They found that the proportion of reported doubles was significantly

higher than was to be expected if the participants had been honest. It was also higher than the number of doubles reported when individuals rolled and reported alone. These findings suggest that individuals are more dishonest in collaborative situations than they are in individual situations. Wouda and his colleagues replicated the experiments of Weisel and Shalvi (2015) and verified their findings (Wouda et al., 2017). Researchers argued that collaborative situations provide individuals with a good reason to justify their immoral behavior, leaving them more likely to be dishonest (Weisel and Shalvi, 2015; Soraperra et al., 2017).

Although the above findings suggest that collaborative situations increase dishonesty, it is worth noting that they ignored the fact that collaboration also typically involves increased observability and accountability. As such, reputational concerns may limit people's willingness to break moral rules (Weisel and Shalvi, 2015). Weisel and Shalvi (2015), for example, asked their participants to sit in separate cubicles. The participants never met each other during the experiment, so the findings may be limited to such cases, where reputation plays a minor role. Therefore, it remains unknown whether individuals are as dishonest in face-to-face collaborative situations where they have concerns about their reputation. One primary aim of the current study is to address this issue.

Furthermore, while previous studies have demonstrated that individuals are more dishonest when collaborating with others, it is still unclear how this dishonesty unfolds over time. The previous evidence suggests that individuals are more likely to cooperate in multiple interactions because they may develop trust with each other over time (Levine and Schweitzer, 2015). In this case, it is expected that individuals will be more likely to become more dishonest as they collaborate over time. The previous research also suggests that individuals are likely to commit minor acts of unethical behavior but not major acts of unethical behavior because they can easily justify these minor acts. Furthermore, over time, individuals become more likely to engage in major forms of unethical behavior as it becomes easier for them to justify their conduct (Welsh et al., 2015). In this case, it is also expected that individuals will become more dishonest over time as they find it easier to justify their immoral behavior. The second aim of this study is to test whether individuals become more likely to collaborate through dishonesty. This will help us to gain a greater understanding of how dishonesty develops in collaborative situations and suggest ways of reducing it.

To address both of these research aims, the present study used a modified sequential dyadic die-rolling paradigm created by Weisel and Shalvi (2015). To increase observability and accountability, we asked participants to sit across from one another at a table. Also, a hidden camera was used to record the outcome of each die roll so that we could identify whether or not a participant lied in a specific trial by comparing the real outcome of each die roll with the outcome of a participant reported.

Based on the previous findings that have shown that collaboration can make it more likely that people will behave unethically (Weisel and Shalvi, 2015), we expected that participants would lie more when operating collaboratively than when operating alone. Moreover, since it becomes easier for participants to justify their immoral behavior over time (Welsh et al., 2015), and because the two participants would be able to develop trust with each other as the game progressed (Levine and Schweitzer, 2015), we expected that more participants would exhibit dishonest behavior over time as they collaborated more.

MATERIALS AND METHODS

Participants

We conducted a prior power analysis using G*Power version 3.1.9.2. The parameters used in this calculation were alpha = 0.05 and power = 0.8. The effect size was derived from the previous study conducted by Weisel and Shalvi (2015). The analysis indicated that 53 participants would be needed for the collaborative condition and 27 participants would be needed for the individual condition. This would provide enough data to test the difference between them. Thus, 88 students who were not psychology majors were recruited from Hangzhou Normal University. Participants were paired with another person of the same gender with whom they were unacquainted. Participants were then randomly assigned to either the collaborative condition or the individual condition. One participant in the individual condition was excluded because the camera was broken and could not record data completely. This left a final sample of 30 dyads (three male dyads, M = 20.2 years; SD = 2.02) in the collaborative condition and 27 participants (one male, M = 19.74 years; SD = 1.58) in the individual condition. Written informed consent was obtained from each participant in accordance with the Declaration of Helsinki. The study was approved by the Ethics Committee of the Center for Cognition and Brain Disorder at Hangzhou Normal University.

Experimental Procedure

A pair of participants came into the laboratory and were randomly assigned to the collaborative condition or the individual condition. Participants in the collaborative condition were then randomly assigned to role A (A) or role B (B). In the individual condition, the same person acted in both roles. In both conditions, the pair of participants were sat across from each other and each had a computer screen and a device for rolling a die (**Figure 1A**).

Participants were told to play the dice-rolling game. While the experiment was going on, the participants were not allowed to talk to each other. In the collaborative condition, A rolled first and reported the outcome by typing a number on the computer. B was then informed of A's outcome. Finally, B rolled and reported their outcome in turn. If their reported outcomes were identical, they both received a reward (see **Figure 2A**). The amount paid was the equivalent in RMB to the number reported on the dice. For example, if both A and B reported a roll of two, they would each

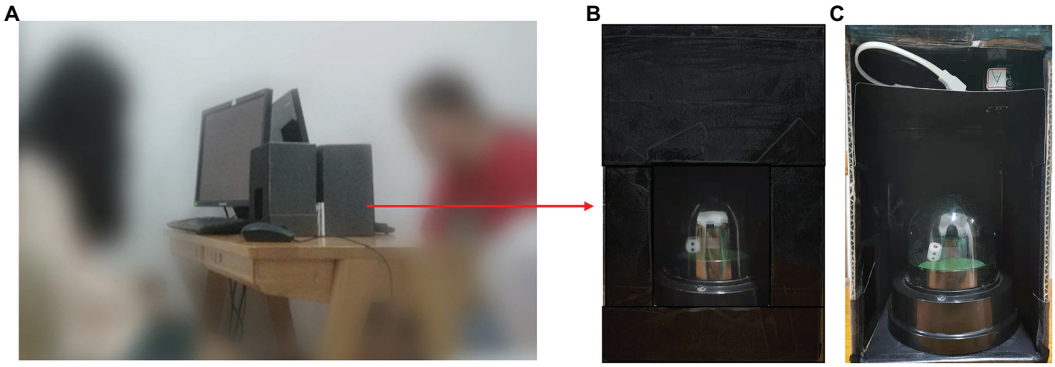

FIGURE 1 | Experimental environment and materials. **(A)** The experimental environment. **(B)** The appearance of the automatic rolling dice device. **(C)** The internal structure of the automatic rolling dice device.

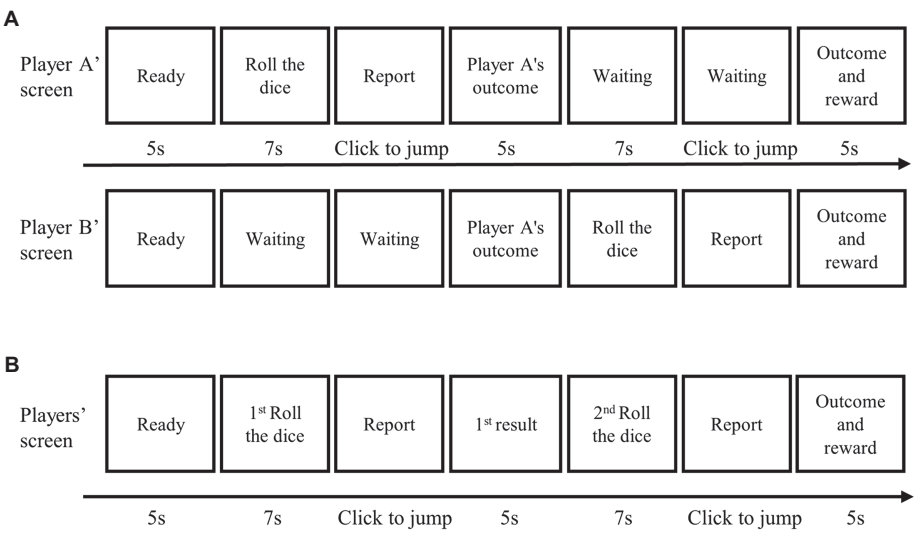

FIGURE 2 | Experimental procedure. **(A)** The procedure in the collaborative condition. **(B)** The procedure in the individual condition.

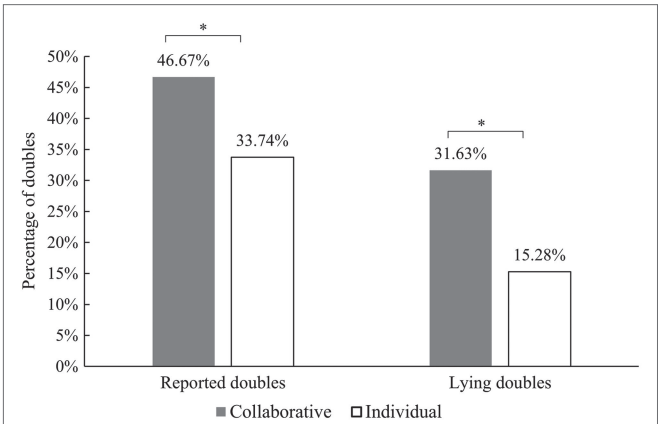

FIGURE 3 | The percentage of reported doubles and the percentage of reports that were lies in both conditions. *$p < 0.05$.

earn ¥0.2; if both A and B reported six, they would each earn ¥0.6. In contrast to the previous studies, the participants were given an electronic rolling device so they could roll their die simply by pressing a button. They could then observe the die through a small window, which was only visible to them alone (**Figure 1B**). There was also an electronic light inside the box to allow participants to see the result of each roll clearly. However, unbeknownst to the participants, a mini camera was hidden at the top of the box to record the outcome of each roll (**Figure 1C**). This allowed us to know whether a participant had misreported their result in each trial.

The experimental procedure for participants in the individual condition (**Figure 2B**) was identical to that in the collaborative condition, except that the participants were told to play the game on their own. Specifically, participants were told to roll the die twice in each trial and were told that if the same number was reported twice, they would receive a reward.

There were 45 trials of dice rolling in total. The participants received a break after every 15 trials. The entire experiment lasted about 30 min. Before the experiment formally began, the participants were allowed to three practice trials. Each participant was also paid ¥15 for showing up to the experiment.

RESULTS

Frequency of Lying

On average, participants in the collaborative condition reported 21 doubles (46.67%), and participants in the individual condition reported 15.18 doubles (33.74%). We also calculated the number of doubles participants reported by lying. As shown in **Figure 3**, the participants in the collaborative condition lied about 14.23 doubles (31.63%) on average, which was significantly more than the participants in the individual condition, who lied about 6.88 doubles (15.28%; Mann–Whitney U test: $U_{lying} = 276.50$, $p = 0.027$, and effect size $r = 0.317$).

We also examined whether both A and B lied more in the collaborative condition than the participants in the individual condition. Results showed that B lied 14.23 times ($SD = 15.53$) in the collaborative condition. This means that they lied significantly more frequently than the participants did in the individual condition when playing role B ($M = 7.67$; $SD = 13.88$), $U = 281.50$, $p = 0.034$, and $r = 0.305$. In the collaborative condition, A lied 10.1 times ($SD = 15.75$), but there was no significant difference in the frequency at which A lied in the collaborative condition compared with the A in the individual condition ($M = 6.59$; $SD = 13.72$), $U = 345.50$, $p = 0.257$, and $r = 0.112$ (see **Figure 4**). The reports for the first and second dice rolls in the individual condition were labeled as role A and role B in the figures.

The Number of People Who Lied

We also calculated the number of people who lied in each condition. In the collaborative condition, 12 people playing as A (40%) lied, and 18 people playing as B (60%) lied. In the individual condition, there were seven participants (25.93%) who lied about their first roll and 9 (33.33%) who lied about their second roll (**Figure 5**). Cross-Tabs analysis showed that, when playing as B, more participants lied when collaborating than when operating individually ($\chi^2(1) = 4.05$, $p = 0.044$, and effect size $\varphi = 0.267$). However, there was no significant difference between the number of participants who lied when playing as A in the collaborative condition compared to those who lied on their first roll in the individual condition ($\chi^2(1) = 1.267$, $p = 0.260$, and $\varphi = 0.149$).

Were Participants More Likely to Lie as the Game Progressed?

To investigate whether participants lied more as the game progressed, Pearson correlation analysis was used to analyze the correlation between the 45 trials and the number of participants (A and B) who lied in each trial. The results showed that the number of participants who lied increased in the collaborative condition for both A and B ($r_A = 0.392$, $p = 0.008$; $r_B = 0.655$, $p < 0.001$). By contrast, there were no significant correlations in the individual condition ($r_A = 0.165$, $p = 0.286$; $r_B = 0.109$, $p = 0.477$; see **Figure 6**).

DISCUSSION

The present study examined whether people are more dishonest in collaborative settings where there are concerns about reputation. It also examined how people's dishonesty unfolds as they continue to collaborate. The results showed that participants told more lies in a collaborative setting than an

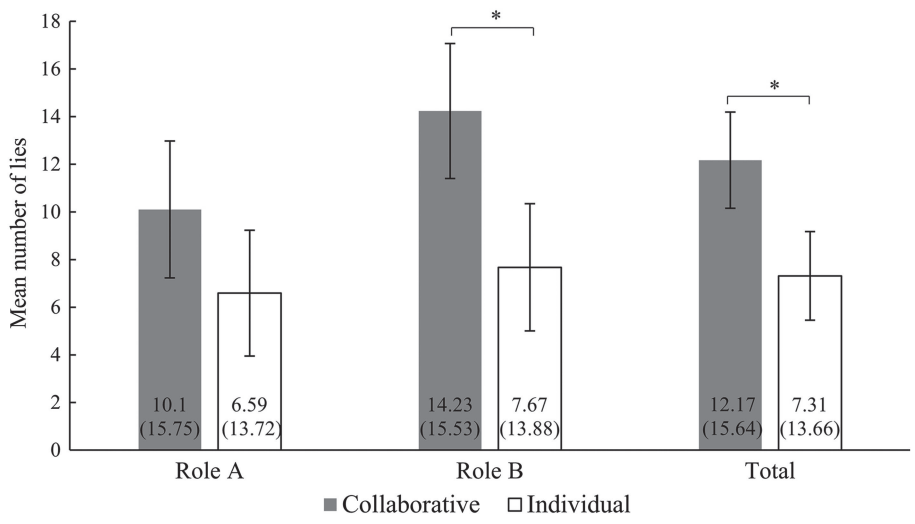

FIGURE 4 | The mean number of lies for different roles in the two conditions. Error bars are ±1 SE; mean and SD are at the bottom of each bar; significance indicators: $*p < 0.05$.

individual setting, even though they had more reason to be concerned about their reputation in the collaborative setting. Results also showed that more participants become dishonest as they continued to collaborate.

Our results are consistent with the previous findings that indicate that participants tend to lie more in collaborative settings than in individual settings (Weisel and Shalvi, 2015; Wouda et al., 2017). Our results also show that this increased effect exists even in a face-to-face situation where there are concerns about reputation, which we introduced by asking participants to sit opposite each other at a table in this study. These findings together demonstrate that collaborative settings do indeed increase the likelihood of dishonest behavior. However, it should be noted that the participants in our study did not lie as much as the participants in the study of Weisel and Shalvi. One possible reason for this is that Weisel and Shalvi (2015) overestimated the size of the effect because the sizes of effects tend to be greater in pioneering studies that are the first to report them (Wouda et al., 2017). Another possible reason is that the participants' concerns about their reputation in our study discouraged their dishonest behavior to a certain extent (Koch and Schmidt, 2010; Kimbrough and Rubin, 2015; Behnk et al., 2019).

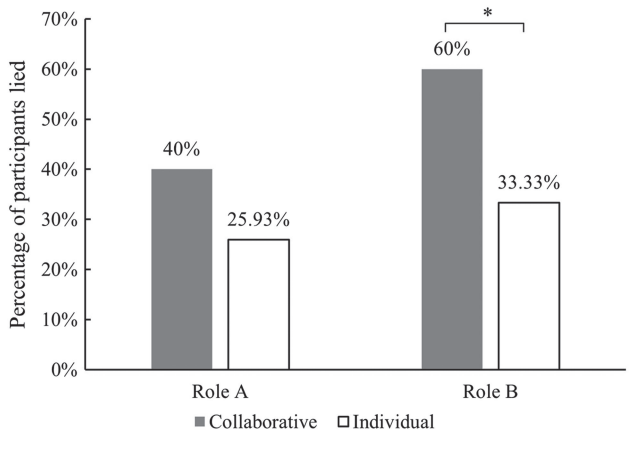

FIGURE 5 | The percentage of dishonest participants in both conditions. *$p < 0.05$.

Furthermore, our results for the collaborative condition showed that while B lied more than the individual participants on their second roll, A lied as frequently as the individual participants on their first roll. This finding suggests that collaboration only makes participants more dishonest when they can determine in advance whether they will be rewarded for lying. One possible reason is that participants playing as A were able to exploit the moral wiggle room provided by their partners, taking advantage of their partners' lies without feeling morally culpable (Gross et al., 2018).

We also found that more and more participants (both A and B) lied as the game progressed. This result suggests that more participants become dishonest after they cooperate for longer. As Gachter and Falk (2002) argue, repetitive play can increase reciprocal

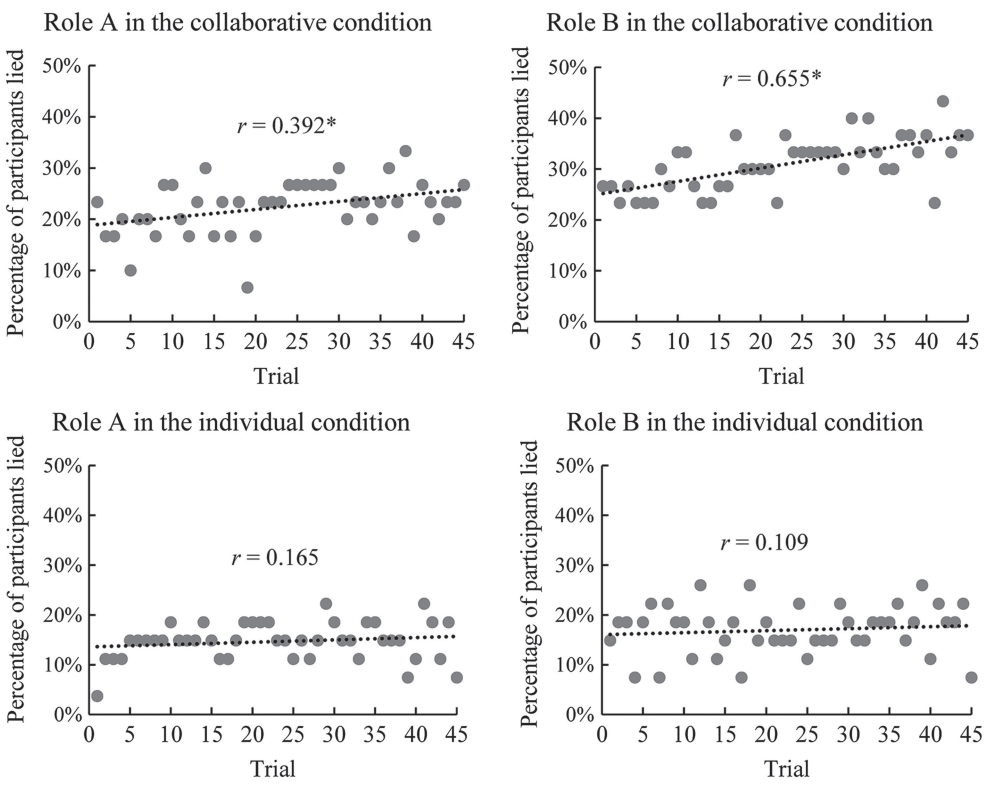

FIGURE 6 | Scatter plots of the percentage of participants who lied in each trial. *$p < 0.05$.

collaboration because it is an appropriate device for re-enforcing contact. Therefore, A and B may learn to cooperate more as the task progresses, resulting in more dishonesty. Furthermore, lying in the collaborative condition benefits not just one participant but both participants, and the previous studies have revealed that prosocial lies promote trust (Levine and Schweitzer, 2015). Also, studies have shown that people who work with the same partner over time are more likely to take bribes from them as they come to trust them more (Abbink, 2004). Therefore, as participants' interactions increase over time, they learn to trust each other more, causing them to lie more frequently when they collaborate. This finding has important implications for attempts to reduce dishonest behavior in collaborative situations. For example, Abbink (2004) suggests that rotating the players in two-player bribery games significantly reduces the amount of bribery. In socio-political spheres, many countries engage in regular staff rotation in public administration as a precautionary measure against corruption. This is the case with the Chinese civil service and the German federal government. Thus, it is reasonable to assume that it would be possible to reduce dishonest behavior in collaborative situations further by increasing staff rotation. Further research would be needed to test this hypothesis.

There are several limitations to the present study. First, the present findings suggest that individuals are more dishonest in face-to-face collaborative situations than in face-to-face individual situations. However, it should be noted that our findings may be due to the interaction between the face-to-face setting and the collaborative situation. Further studies should also include two conditions in which participants work collaboratively or individually but do not see each other's face. This would help to examine the effect of a possible interaction between working face-to-face and working in collaboration. Second, our study indicates that individuals may not be as dishonest as Weisel and Shalvi (2015) found. This finding should be interpreted carefully because the two studies were conducted in different countries, years apart, and with different subject pools. Third, there were very few male participants, which may limit the external validity of this study. Future studies should include more male participants to examine the gender effect of collaborative dishonesty.

CONCLUSION

The present study examined whether collaborative situations make individuals more dishonest in face-to-face settings and how this dishonesty unfolds over time. It found that participants whose decisions determine the final payoff (in other words, those playing as B) lied more in collaborative situations than in individual situations. Those participants whose decisions did not determine the final payoff (those playing as A) lied equally in collaborative and individual situations. Furthermore, we found that more and more participants lied as they collaborated more with their partner. These findings suggest that in face-to-face settings, collaborative situations lead to more dishonesty than individual situations.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Ethics Committee of the Center for Cognition and Brain Disorder at Hangzhou Normal University. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

YD searched the literature, recruited the subjects, collected the data, performed the data analysis, and wrote the original draft. WM contributed to the experimental design and revised the article. QS revised the article. LS generated the research concept, provided funds, and revised the article. All authors contributed to the article and approved the submitted version.

FUNDING

This study was supported in part by grants from the National Science Foundation of China (U1736125), from the Cultivation Project of Provincial Characteristic Key Discipline in the College of Education of Hangzhou Normal University (20JYXK003), and from the Hangzhou Social Science Foundation of China (2018RC2X17) to LS.

REFERENCES

Abbink, K. (2004). Staff rotation as an anti-corruption policy: an experimental study. *Eur. J. Polit. Econ.* 20, 887–906. doi: 10.1016/j.ejpoleco.2003.10.008

Bazerman, M. H., Curhan, J. R., Moore, D. A., and Valley, K. L. (2000). Negotiation. *Annu. Rev. Psychol.* 51, 279–314. doi: 10.1146/annurev.psych.51.1.279

Behnk, S., Barreda-Tarrazona, I., and García-Gallego, A. (2019). Deception and reputation – An experimental test of reporting systems. *J. Econ. Psychol.* 71, 37–58. doi: 10.1016/j.joep.2018.10.001

Gachter, S., and Falk, A. (2002). Reputation and reciprocity: consequences for the labour relation. *Scand. J. Econ.* 104, 1–26. doi: 10.1111/1467-9442.00269

Gross, J., Leib, M., Offerman, T., and Shalvi, S. J. P. S. (2018). Ethical free riding: when honest people find dishonest partners. *Psychol. Sci.* 29, 1956–1968. doi: 10.1177/0956797618796480

Kameda, T., Takezawa, M., and Hastie, R. (2005). Where do social norms come From?: The example of communal sharing. *Curr. Dir. Psychol. Sci.* 14, 331–334. doi: 10.1111/j.0963-7214.2005.00392.x

Kimbrough, E. O., and Rubin, J. (2015). Sustaining group reputation. *J. Law Econ. Organ.* 31, 599–628. doi: 10.1093/jleo/ewu019

Koch, C., and Schmidt, C. (2010). Disclosing conflicts of interest – do experience and reputation matter? *Accounting Organ. Soc.* 35, 95–107. doi: 10.1016/j.aos.2009.05.001

Kramer, R. M. (1999). Trust and distrust in organizations: emerging perspectives. *Enduring Questions* 50, 569–598. doi: 10.1146/annurev.psych.50.1.569

Levine, E. E., and Schweitzer, M. E. (2015). Prosocial lies: when deception breeds trust. *Organ. Behav. Hum. Decis. Process.* 126, 88–106. doi: 10.1016/j.obhdp.2014.10.007

Rand, D. G. (2017). Social dilemma cooperation (unlike Dictator Game giving) is intuitive for men as well as women. *J. Exp. Social Psychol.* 73, 164–168. doi: 10.1016/j.jesp.2017.06.013

Rempel, J. K., Ross, M., and Holmes, J. G. (2001). Trust and communicated attributions in close relationships. *J. Pers. Soc. Psychol.* 81, 57–64. doi: 10.1037/0022-3514.81.1.57

Soraperra, I., Weisel, O., Zultan, R. I., Kochavi, S., Leib, M., Shalev, H., et al. (2017). The bad consequences of teamwork. *Econ. Lett.* 160, 12–15. doi: 10.1016/j.econlet.2017.08.011

Weisel, O., and Shalvi, S. (2015). The collaborative roots of corruption. *Proc. Natl. Acad. Sci. U. S. A.* 112, 10651–10656. doi: 10.1073/pnas.1423035112

Welsh, D. T., Ordóñez, L. D., Snyder, D. G., and Christian, M. S. (2015). The slippery slope: how small ethical transgressions pave the way for larger future transgressions. *J. Appl. Psychol.* 100, 114–130. doi: 10.1037/a0036950

Wouda, J., Bijlstra, G., Frankenhuis, W. E., and Wigboldus, D. H. (2017). The collaborative roots of corruption? A replication of Weisel & Shalvi (2015). *Collabra: Psychol.* 3, 1–3. doi: 10.1525/collabra.97

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychology

ORIGINAL RESEARCH
published: 01 July 2021
doi: 10.3389/fpsyg.2021.694841

Positive Emotion and Honesty

Evelyn Medai[1†] *and Charles N. Noussair*[2*†]

[1] School of Law, New York University, New York, NY, United States, [2] Eller College of Management, University of Arizona, Tucson, AZ, United States

We report an experiment that considers the impact of emotional state on honesty. Using the die-rolling task created by Fischbacher and Föllmi-Heusi to detect the level of dishonesty in a sample of individuals, we study the effects of induced happiness on the incidence of self-interested lying. The experiment uses 360-degree videos to induce emotional state. We find that people behave more honestly in a state of happiness than they do in a neutral state.

Keywords: honesty, happiness, virtual reality, emotion induction, experiment

OPEN ACCESS

Edited by:
Nikolaos Georgantzis, Burgundy School of Business, France

Reviewed by:
Iván Barreda Tarrazona, University of Jaume I, Spain
Aurora García-Gallego, University of Jaume I, Spain

***Correspondence:**
Charles N. Noussair
cnoussair@email.arizona.edu

[†] These authors have contributed equally to this work

Specialty section:
This article was submitted to Personality and Social Psychology, a section of the journal Frontiers in Psychology

Received: *13 April 2021*
Accepted: *31 May 2021*
Published: *01 July 2021*

Citation:
Medai E and Noussair CN (2021) Positive Emotion and Honesty. Front. Psychol. 12:694841. doi: 10.3389/fpsyg.2021.694841

INTRODUCTION

Individuals differ from each other in their propensity to behave dishonestly. Indeed, the same person may behave with more integrity on 1 day than another. What causes some people to behave more dishonestly than others? Factors such as background, personality, decision history, managerial philosophy, and reinforcement have all been shown to correlate with ethical behavior in business settings (Stead et al., 1990; see Ayal and Gino, 2011, for a survey). The payoffs at stake also exert an effect (Gneezy, 2005; Gneezy et al., 2020). But do *emotional states* also have an impact on honesty? If so, an organization might be able to use an insight in this regard to reduce unethical behavior by creating an environment conducive to particular emotional states. Since emotional states are malleable, interventions to do so may be feasible and cost-effective. The relationship between emotional state and honesty is the focus of the study reported here.

We report an experiment designed to explore the specific relationship between a positive emotional state and honesty. Our research question is whether individuals in a positive emotional state are more honest than those in a control treatment. To measure honesty, we utilize the die-rolling task created by Fischbacher and Föllmi-Heusi (2013). The task involves asking an individual to roll a die privately and then to report what was rolled. The individual receives a monetary payment based on the number that she reports. A player can typically earn more money if she makes a false report, and her actual roll can never be verified. Dishonest behavior can only be observed at the level of a sample, and not at the level of the individual[1]. In their original experiment, Fischbacher and Föllmi-Heusi (2013) found that individuals lie on average, but not to the maximum extent possible. Typically, a sample of individuals exploits on average ¼ of the potential monetary gains from lying (Abeler et al., 2019).

[1] As an alternative task, one could also employ the version of the dice-rolling task used by Pascual-Ezama et al. (2020). In their version, participants roll a die electronically on a website provided by the experimenter and are asked to enter the result on their computer. The result of the roll is recorded so that the experimenter can know if a roll actually took place, and can compare the result to the participant's report. This would allow the experimenter to identify different types of dishonest behavior: misreporting the roll, not rolling the die at all, and rolling multiple times until a favorable result is achieved.

The die-rolling task has become standard in experimental economics to measure honesty, and follow up studies confirm the existence of substantial, though less than ubiquitous, dishonesty. Abeler et al. (2019) have reviewed 90 studies using the Fischbacher and Föllmi-Heusi paradigm to identify the correlates of truth-telling. Among the patterns that they report, they find that women behave more honestly than men on average, and that among student participants, field of study has no effect on honesty. Rosenbaum et al. (2014), in an earlier survey of 63 experiments on ethical behavior, similarly report that there is evidence that women behave more ethically than men on average, and that the results on whether economics and business students differ from others are mixed. These findings are relevant to our work in that we test for and find no difference between the genders or between business/economics majors and those enrolled in other programs of study. None of the studies that Rosenbaum et al. (2014) or Abeler et al. (2019) reviewed studied the causal relationship between honesty and emotional state.

There has been some previous work on the connection between other emotions and unethical behavior. Motro et al. (2018) find that anger increases, while guilt reduces, deceptive behavior. Klygte et al. (2013) also report that anger leads to less ethical, while fear induces more ethical, decisions. Lim et al. (2015) find that subliminal priming with disgusted faces makes individuals slightly more honest in a mind-game die-rolling task (Jiang, 2013),[2] which is closely related to the task we employ. Kugler et al. (2021) find no relationship between disgust and honesty in three different tasks. Brain imaging studies have revealed a network of brain regions that exhibit greater activation when individuals are being deceptive, suggesting that lying is more demanding of the brain than honesty (see e.g., Greene and Paxton, 2009). We are unaware of any research studying the causal impact of happiness on ethical behavior.

The experiment reported in this paper has two treatments: one in which a positive emotional state, which we will refer to as *Happiness*,[3] is induced, and one control treatment, which we call *Neutral*. As a means of emotion induction, we employ a novel method. We conduct the experiment using Oculus Rift virtual reality headsets, which participants use to view a 360-degree video that induces happiness or one that does not have an effect on emotional state.[4] After subjects watch the video, they are sent into another room where they are read a set of instructions that describe how they are to roll the die, and are informed that the die roll would be completely private. They are then sent out of the room one at a time to roll the die privately out of the view of any other person, and to report their roll to an experimenter in another room.

We find that the Happiness treatment results in lower levels of dishonesty than the Neutral treatment. The effect is significant at conventional or borderline levels, depending on the statistical analysis that is employed. We observe no significant difference in lying between women and men. Section Experimental Design describes the experiment and section Results reports the results. We offer some concluding remarks in section Conclusion.

EXPERIMENTAL DESIGN

Procedures Common to All Treatments

This experiment is an individual decision-making task, with subjects acting completely independently of each other. The study was conducted with 106 University of Arizona students between November 2017 and May 2018, with 53 participants assigned to each of two treatments. Between three and six subjects participated in each session. All sessions were conducted at the Economic Science Laboratory at the University of Arizona, located in Tucson, Arizona, USA. There were no other tasks conducted in the session, either before or after those described here. The subjects were recruited from the laboratory's subject pool and were all undergraduates from a variety of programs at the university. Of the 106 total subjects, 48 were male and 58 were female. Sixty-five were studying economics or business and the remaining 41 were pursuing other studies.

At the beginning of a session, subjects reported to Room A, one of the rooms in the Economic Science Laboratory facility. The experiment began with individuals watching a video using an Oculus Rift virtual reality headset in Room A for ∼5–6 min. These videos were played using a program called Virtual Desktop and induced either a state of Happiness or one of Neutrality. The videos are filmed from a perspective of someone inside the video and displayed in 360 degrees. This means that the subject sees the video no matter in which direction she is looking and feels like an active participant in the video. The experience is highly immersive.

After the video, subjects were led to Room B, another room in the laboratory facility adjacent to room A, where an experimenter read the instructions for the die-rolling task. The subjects also had a written copy of the instructions they could use to follow along [a copy of the instructions can be found in the Appendix (**Supplementary Material**)]. These instructions were very short so that the effect of the induced emotion did not have time to dissipate. They explained to participants how they were to roll a six-sided die and report their roll to the experimenter. The instructions also explained that the subject would be the only one to observe the die roll. They also indicated how subjects would be paid. In addition to a $2 fee paid to all participants for viewing a video, a subject was given $2 times the number of the die roll that she reported. Therefore, in addition to the $2 payment for watching the video, subjects received $2 if they claimed that they rolled a 1, $4 if they reported a 2, $6 if they indicated a 3, and

[2]In the mind game task studied by Jiang (2013), individuals are told to role a die "in their minds," that is, to imagine a die roll that does not actually physically take place, and then to report the outcome of the roll.

[3]We recognize that "Happiness" is a broad term, with a variety of uses and meanings in the scientific literature. We use the term here to describe our treatment condition as a concise term to describe the positive emotional state that is induced by our video. We recognize that positive emotional state is only one component of subjective well-being. For example, Seligman (2011, 2018) considers positive emotional state as one of the five dimensions in his well-known PERMA model of well-being. Our pretest results show that individuals do describe themselves as "happier" after viewing the video than they were before viewing it.

[4]There is a long tradition of using videos to induce emotional state. Using 360 videos shown in virtual reality, in our view, constitutes a more intensive implementation of this established method. See Kugler et al. (2020) for an example of the use of virtual reality to induce emotional state for participants in a trust game.

so on. The higher the reported roll, the higher the payoff that the subject received. Since the roll was entirely private, with no other participant or experimenter ever knowing the true result of an individual roll, subjects had a material incentive to lie.

While it was impossible to know which individual subjects were lying, collecting many observations of data reveals the average level of dishonesty in a group. In any group, if all subjects are honest, the result would be an approximately uniform distribution of the frequency of reports of each number on the die. With a six-sided die, each number should make up ∼16.667% of the total number of observations. Lying causes the distribution of frequencies to shift, and if the lying is self-interested, it would shift toward higher numbers. The average report, and the percentage of individuals submitting the highest-paying report, can be interpreted as measures of the extent of self-interested lying among participants in a given treatment.

Once the instructions were read, subjects were sent out of room B, one-by-one, to roll their die. The die roll was completely private, with no experimenters or other participants witnessing the roll. The subjects were instructed that they could go anywhere in the building to perform the task, and that they should make sure to roll the die privately. They then returned to either Room A or another available, empty room (depending on the session) where an experimenter was present. The subject reported the roll to the experimenter with the room door closed and was paid accordingly. They were then asked to immediately leave the building. Each subject was sent out of room B to roll the die only after the previous participant had completed reporting her roll in the other room and had left the area. These procedures ensured that no other participants were within view or earshot when a report was made, and that subjects could not discuss their reports with each other before submitting them. In addition to the number rolled on the die, subjects were asked about their major (program of study) after they reported their roll. The experimenter also recorded their gender.

Treatments

Subjects were shown one of two 360-degree videos in virtual reality, depending on the treatment. The experimental design had a between-subject structure, in that each subject was only shown one video and performed the die-rolling task only once. No subject participated in the experiment more than once. Of the male participants, 25 were in the Neutral treatment and 23 were in the Happiness condition. Twenty-eight females were in the Neutral, and 30 took part in the Happiness, treatment respectively.

Neutral

the video for the control treatment was a simple video of a tulip field on a sunny day. The video is taken from the perspective of an individual sitting in the field. There was no music, but there were soft noises, such as birds chirping and distant chatter. The video lasted for ∼5 min and subjects were shown the video once before proceeding with the rest of the experiment. The video can be found at https://www.youtube.com/watch?v=SmhuzTzUKQY.

Happiness

The video inducing a state of happiness was a video shown from the point of view of surfers in a tropical beach setting. Viewers would get a first-person viewpoint of surfing on waves, paddling out to sea, and swimming in the ocean. Accompanying the visual component of the video, there was upbeat, positive music playing, further promoting a pleasant experience. The video was approximately two and a half minutes long and subjects were shown the video twice; the video was immediately played again once it finished playing for the first time. This video was played twice to maintain consistency among video lengths across treatments. This video can be viewed at https://www.youtube.com/watch?v=MKWWhf8RAV8[5].

We conducted a manipulation check during several earlier sessions, with different individuals than those who participated in the experiment, to verify that the videos increased the level of the targeted emotion while not increasing any others. In the manipulation check, we asked individuals to report the levels of five emotions: Happiness, Fear, Sadness, Anger, and Disgust, that they were currently experiencing. We asked these participants to indicate, on a scale of 1–7, the strength with which they felt each emotion. They did so both before and after viewing one the videos. The data are given in the **Table 1**.

The table shows that the Neutral video did not increase the strength of any of the emotions (other than an insignificant increase in happiness). The Happiness video raised average reported happiness while not increasing any of the other emotions.[6] The average level of self-reported happiness, 5.36, was significantly greater after viewing the Happiness video than before viewing a video (4.21). A pooled variance t-test rejects the hypothesis that the two means are equal ($t = 1.89$, $p < 0.05$). The Neutral video did not yield a level of self-reported happiness significantly different from that recorded prior to the viewing of a video ($t = 0.515$, $p > 0.25$). Those who viewed the Happiness video reported a greater degree of happiness afterward than those who had viewed the Neutral video ($t = 1.31$, $p < 0.1$), though the effect is only borderline significant. The average level of each of the other four emotions after viewing a video is not different between the two treatments.

[5]There are obviously many types of positive emotions, and among these are a number of states that are referred to with the term "Happiness." Sports tend to create positive emotional states with high arousal. See Hills and Argyle (1998) for a discussion of this point. It is quite possible that different types of positive emotional state may exert different effects on honesty, and this is an agenda of questions that can be addressed in follow-up research. Hills and Argyle show that participating in sports increases positive feelings toward others and toward life, improves body image and self-esteem, and creates feelings of achievement and excitement. Any of these could serve as channels whereby the surfing video, which simulates a sporting activity in virtual reality, might increase or decrease honest behavior. Nevertheless, Hills and Argyle found no correlation between participation in sport and scores on a social conformity index that they interpret as a "lie scale".

[6]The Happiness video lowered the average level of sadness ($t = 2.46, p < 0.05$) and fear ($t = 2.77$, $p < 0.01$) significantly, but did not significantly affect the average level of disgust or anger. The Neutral video reduced the levels of sadness ($t = 3.19$, $p < 0.01$), fear ($t = 2.47, p < 0.05$), and anger ($t = 3.23, p < 0.01$) significantly.

TABLE 1 | Manipulation check: average self-reported emotional states on a scale of 1–7, before and after viewing the videos.

	Average self-reported emotion				
Emotion condition	**Disgust**	**Sadness**	**Happiness**	**Fear**	**Anger**
Before video ($n = 47$)	1.33	2.14	4.21	2.32	2.01
After Neutral video ($n = 22$)	1.15	1.33	4.55	1.4	1.14
After Happiness video ($n = 25$)	1.21	1.14	5.36	1.46	1.55

Hypothesis

Before conducting the experiment, we formulated the following hypothesis regarding our treatment differences. Since there are no prior results, to our knowledge, to guide our a priori beliefs, we have no basis to hypothesize a sign for a treatment effect. Thus, our hypothesis is a two-sided claim that there would be no treatment effect.

> *Hypothesis: People behave equally honestly in the Happiness and Neutral treatments. The average reports, as well as the percentage of individuals reporting a roll of 6, are not different between the two treatments.*

For a two-sided t-test of the hypothesis that there is no difference in average report between treatments, our sample size yields a power of 73% to detect a medium sized treatment effect of 0.5 standard deviations at a significance level of 0.05. In terms of the proportion of individuals claiming a roll of 6, our sample size yields a power of 60% of detecting a difference between treatments at a significance level of 0.05, if the true means are 0.17 and 0.35 in the two treatments. Although the experiment was not designed specifically to do so, in our analysis of the data, reported in section Results, we also consider whether there are differences in the level of honesty between women and men, and between economics/business majors and those pursuing other programs of study.

RESULTS

Summary of Data

The distribution of reported dice rolls in each treatment can be seen in **Figure 1**. In the figure, the vertical axis represents the number of individuals who reported a particular roll, while the horizontal axis indicates the roll reported.

In the Neutral treatment, the average report was 4.83 (std. dev = 1.369), significantly greater than the average under honest reporting of 3.5 ($t = 7.07$, $p < 0.001$). There was also a greater than random incidence of the reporting of 5 or 6. Thirty-eight of the 53 subjects reported rolling either a 5 or a 6. If people had been honest, we would expect 17.67 out of 53, one-third of, subjects to report either a 5 or a 6. We reject the hypothesis that the proportion reporting 5 or 6 is equal to 1/3, using a binomial test ($z = 5.92$, $p < 0.001$). On the other hand, if all participants were selfish and willing to be as dishonest as needed to maximize their monetary payment, all players would report a six. This is also clearly not observed, with only a minority of participants reporting a 6. Gender differences are small and insignificant, with 40% of men and 42.8% of women reporting a 6, and the average reports being 4.89 and 4.76 for women and men, respectively.

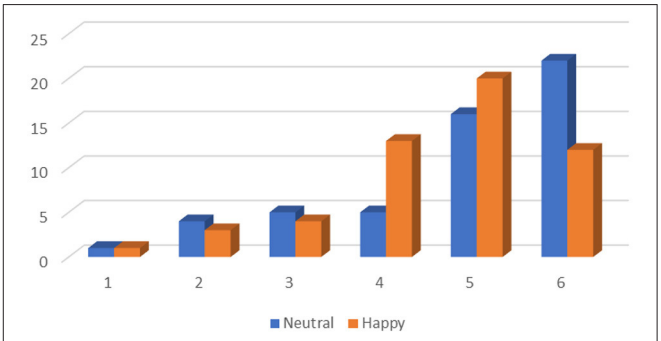

FIGURE 1 | Reported rolls, both treatments.

In the Happiness treatment, the average reported roll was 4.58 (std. dev = 1.20), 0.25 lower than in the Neutral treatment. This is also significantly different from the average under honesty of 3.5 ($t = 6.55$, $p < 0.01$). **Figure 1** shows that 32 of 53 subjects reported a 5 or a 6, significantly greater than under honest reporting ($z = 4.17$, $p < 0.01$). However, only 12 subjects in the Happiness treatment reported rolling a 6 (22.6%). In the Happiness treatment, 20% of women and 26.6% of men claimed a six, and the average report was 4.5 and 4.69 for women and men, respectively. The difference in the average report between the two treatments is therefore 0.39 for women and 0.06 for men. There is a 18.9% point difference between treatments in the incidence of claiming a roll of 6, with almost twice as many claims of 6 in the Neutral treatment. Because women make slightly lower average reports than men under the Happiness treatment, while making slightly higher reports than men under Neutral, there is no overall gender effect.

Formal Comparison Between Treatments

We conducted a number of formal statistical tests to compare the average report, the incidence of extreme lying, and the distribution of reports, between the two treatments. A t-test comparing the difference in means between the Neutral and Happiness treatments results in a t-statistic of 2.16, significant at $p < 0.05$ (two-tailed test). The Neutral treatment generates a higher average report than the Happiness condition.

TABLE 2 | Determinants of claiming a six and of overall roll claimed.

	Prob. claim 6 (Probit) (1)	Prob. claim 6 (Logit) (2)	Claim (OLS) (3)	Claim (OLS) women only (4)	Claim (OLS) men only (5)
Constant	−0.354 (0.349)	−0.513 (0.577)	5.144*** (0.346)	5.352*** (0.401)	4.741*** (0.416)
Happiness	−0.549* (0.360)	−0.916* (0.607)	−0.458* (0.346)	−0.480* (0.336)	−0.015 (0.403)
Gender	0.032 (0.351)	0.005 (0.567)	−0.173 (0.358)		
Major	0.096 (0.297)	0.161 (0.497)	−0.293 (0.285)	−0.587* (0.381)	−0.047 (0.432)
Gender × Happiness	0.168 (0.523)	0.295 (0.869)	0.374 (0.508)		
n	106	106	106	58	48

*The dependent variable in columns (1) and (2) is a dummy variable that equals 1 if an individual reports a roll of 6, and 0 otherwise. The dependent variable in columns (3)–(5) is the die roll an individual reports. Column (1) is a Probit specification, (2) is a Logit, and (3)–(5) are OLS specifications. Each individual is an observation. n = 106. *Means $p < 0.1$, ***Refers to $p < 0.01$, standard errors in parentheses.*

To compare the amount of extreme lying between treatments, we conduct a binomial test of the hypothesis that the proportion of 6s is equal in the two treatments. The test yields a *p*-value of 0.038. The probability of claiming 6 is significantly lower in the Happiness than in the Neutral treatment.

Finally, we conducted a chi squared test to determine whether there were significant differences in the distribution of reported rolls between treatments. This test results in a statistic of 18.795. At five degrees of freedom, this is significant at 1%. Thus, the distribution of reports differs between the two conditions.

Regression Analysis

To evaluate the hypothesis, while controlling for influences that might affect the comparison between treatments, we conducted regressions with two different dependent variables. The first is a dummy variable, which takes on a value of 1 if the participant rolls a 6 and 0 otherwise. These regressions consider the determinants of extreme lying. The second is the actual reported roll, a measure of the general tendency to lie. The dummy variable *Happiness* was coded as a 1 for the Happiness treatment and 0 otherwise. To create the variable "*Major,*" business and economics majors were coded as a 1 and all other majors were coded as a 0. For the "Gender" variable, all males were coded as a 1 while all females were coded as a 0.

In **Table 2**, the estimates for the variable Happiness reveal an effect of treatment that is significant at $p < 0.1$, and that is robust to the specification. It confirms, albeit at a marginal significance level, that controlling for gender and major, there is more honest behavior in the Happiness than in the Neutral treatment. Splitting the sample between women and men, however, reveals that the treatment effect is specific to women. The variable Happiness is significant in equation (4) though not in (5). In specifications (1)–(3) in which both women and men are included, the coefficient for Happiness is marginally significant, indicating that there is a treatment effect for women (the base category for gender). However, the sum of the coefficients for Happiness and Gender*Happiness is not significant, indicating that there is no treatment effect for men.

The regressions also show that there is no overall effect of gender on honesty. There is also no effect of program of study for the sample as a whole. However, there is an effect of major if only women are considered. Women who are studying business or economics submit lower reported rolls than those pursuing other majors.

CONCLUSION

We observe some evidence that people are more honest in a state of happiness than in a state of neutrality. In the laboratory, emotions can have an effect on the extent of ethical behavior. We do not know, for now, how general this relationship is. However, if the effect transfers to a workplace environment, it would indicate that a creating a more positive workplace environment would lead to more honest behavior on the part of employees. Research on how to create a positive workplace culture is well-developed. Some of these strategies include caring for colleagues on a personal level, providing support and compassion when others are struggling, avoiding blame, forgiving mistakes, and emphasizing the meaning of the work being done (Seppala and Cameron, 2015). Using such techniques to create a positive work environment may lead to a decrease in dishonesty in the workplace. Similarly, if schools and universities are able to improve overall levels of positive emotion in students, particularly when they are in the classroom, it could lead to a reduction in academic dishonesty.[7]

We observed no significant difference in honesty by gender. The conclusion that there was not a significant effect of gender on honesty provides yet another rationale for the equal treatment of the genders in the workplace. There is no reason to believe, based on what we have observed in this study, that an employee or a student of one gender would be more or less ethical than an individual of another gender.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

[7]It may be that such relatively early inducement of changes in habits might have long-term effects; studies have found that college students who engage in academic dishonesty are more likely to act dishonestly in the workplace (Nonis and Swift, 2001).

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Institutional Review Board of the University of Arizona. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

Both authors were involved in designing and conducting the experiment, analyzing the data, and writing the manuscript.

ACKNOWLEDGMENTS

We thank two referees and participants at the ANZWEE 2019 conference at Monash University, as well as the 2019 Workshop on Emotions, Stress and Incentives at the Labex Cortex in Lyon, France, for helpful comments.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2021.694841/full#supplementary-material

REFERENCES

Abeler, J., Nosenzo, D., and Raymond, C. (2019). Preferences for truth-telling. *Econometrica* 87, 1115–1153. doi: 10.3982/ECTA14673

Ayal, S., and Gino, F. (2011). "Honest rationales for dishonest behavior," in *The Social Psychology of Morality: Exploring the Causes of Good and Evil*, eds M. Mikulincer, and P. R. Shaver (Washington, DC: American Psychological Association), 149–166. doi: 10.1037/13091-008

Fischbacher, U., and Föllmi-Heusi, F. (2013). Lies In disguise - an experimental study on cheating. *J. Eur. Econ. Assoc.* 11, 525–547. doi: 10.1111/jeea.12014

Gneezy, U. (2005). Deception: the role of consequences. *Am. Econ. Rev.* 95, 384–394. doi: 10.1257/0002828053828662

Gneezy, U., Kajakaite, A., and Sobel, J. (2020). Lying aversion and the size of the lie. *Am. Econ. Rev.* 108, 419–453. doi: 10.1257/aer.20161553

Greene, J., and Paxton, J. (2009). Patterns of neural activity associated with honest and dishonest moral decisions. *Proc. Natl. Acad. Sci. U. S. A.* 106, 12506–12511. doi: 10.1073/pnas.0900152106

Hills, M., and Argyle, P. (1998). Positive moods derived from leisure and their relationship to happiness and personality. *Pers. Individ. Dif.* 25, 523–535. doi: 10.1016/S0191-8869(98)00082-8

Jiang, T. (2013). Cheating in mind games, the subtlety of rules matters. *J. Econ. Behav. Organ.* 93, 328–336. doi: 10.1016/j.jebo.2013.04.003

Klygte, V., Connolly, S., Thiel, C., and Davenport, L. (2013). The influence of anger, fear, and emotion regulation on ethical decision making. *Hum. Perf.* 26, 297–326. doi: 10.1080/08959285.2013.814655

Kugler, T., Noussair, C. N., and Hatch, D. (2021). Does disgust increase unethical behavior? A replication of winterich, mittal, and morales (2014). *Soc. Psychol. Personal. Sci.* doi: 10.1177/1948550620944083. [Epub ahead of print].

Kugler, T., Ye, B., Motro, D., and Noussair, C. N. (2020). On trust and disgust: evidence from face reading and virtual reality. *Soc. Psychol. Personal. Sci.* 11, 317–325. doi: 10.1177/1948550619856302

Lim, J., Ho, P., and Mullette-Gillman, O. (2015). Modulation of incentivized dishonesty by disgusted facial expressions. *Front. Neurosci.* 9:250. doi: 10.3389/fnins.2015.00250

Motro, D., Ordóñez, L. D., Pittarello, A., and Welsh, D. T. (2018). Investigating the effects of anger and guilt on unethical behavior: a dual-process approach. *J. Bus. Ethics* 152, 133–148. doi: 10.1007/s10551-016-3337-x

Nonis, S., and Swift, C. O. (2001). An examination of the relationship between academic dishonesty and workplace dishonesty: a multicampus investigation. *J. Educ. Bus.* 77, 69–77. doi: 10.1080/08832320109599052

Pascual-Ezama, D., Prelec, D., Munoz, A., and Gil-Gomez de Liano, B. (2020). Cheaters, liars, or both? A new classification of dishonesty profiles. *Psychol. Sci.* 31, 1097–1106. doi: 10.1177/0956797620929634

Rosenbaum, S., Billinger, S., and Steiglitz, M. (2014). Let's be honest: a review of experimental evidence of honesty and truth-telling. *J. Econ. Psychol.* 45, 181–196. doi: 10.1016/j.joep.2014.10.002

Seligman, M. (2011). *Flourish*. New York, NY: Free Press.

Seligman,. M. (2018). PERMA and the building blocks of well-being. *J. Posit. Psychol.* 13, 333–335. doi: 10.1080/17439760.2018.1437466

Seppala, E., and Cameron, K. (2015). Proof that positive work cultures are more productive. *Harv. Bus. Rev.* 1:2015.

Stead, W. E., Worrell, D. L., and Stead, J. G. (1990). An integrative model for understanding and managing ethical behavior in business organizations. *J. Bus. Ethics* 9, 233–242. doi: 10.1007/BF00382649

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychology

ORIGINAL RESEARCH
published: 27 August 2021
doi: 10.3389/fpsyg.2021.693942

Do Not Tell Me More; You Are Honest: A Preconceived Honesty Bias

David Pascual-Ezama[1,2,3]*, *Adrián Muñoz*[4] and *Drazen Prelec*[2,5,6]

[1] Accounting and Financial Administration Department, Universidad Complutense de Madrid, Madrid, Spain, [2] Sloan School of Management, Massachusetts Institute of Technology, Boston, MA, United States, [3] RCC Fellow - Harvard Business School, Harvard University, Boston, MA, United States, [4] Methodology and Social Psychology, Universidad Autónoma de Madrid, Madrid, Spain, [5] Department of Economics, Massachusetts Institute of Technology, Boston, MA, United States, [6] Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Boston, MA, United States

According to the previous literature, only a few papers found better accuracy than a chance to detect dishonesty, even when more information and verbal cues (VCs) improve precision in detecting dishonesty. A new classification of dishonesty profiles has recently been published, allowing us to study if this low success rate happens for all people or if some people have higher predictive ability. This paper aims to examine if (dis)honest people can detect better/worse (un)ethical behavior of others. With this in mind, we designed one experiment using videos from one of the most popular TV shows in the UK where contestants make a (dis)honesty decision upon gaining or sharing a certain amount of money. Our participants from an online MTurk sample ($N = 1{,}582$) had to determine under different conditions whether the contestants would act in an (dis)honest way. Three significant results emerged from these two experiments. First, accuracy in detecting (dis)honesty is not different than chance, but submaximizers (compared to maximizers) and radical dishonest people (compare to non-radicals) are better at detecting honesty, while there is no difference in detecting dishonesty. Second, more information and VCs improve precision in detecting dishonesty, but honesty is better detected using only non-verbal cues (NVCs). Finally, a preconceived honesty bias improves specificity (honesty detection accuracy) and worsens sensitivity (dishonesty detection accuracy).

Keywords: dishonesty, cheating, lying, behavioral profiles, detection accuracy

OPEN ACCESS

Edited by:
Ismael Rodriguez-Lara, University of Granada, Spain

Reviewed by:
Lara Ezquerra, University of the Balearic Islands, Spain
Yossef Tobol, Tel-Hai College, Israel

***Correspondence:**
David Pascual-Ezama
dp_ezama@mit.edu;
david.pascual@ccee.ucm.es

Specialty section:
This article was submitted to Personality and Social Psychology, a section of the journal Frontiers in Psychology

Received: *13 April 2021*
Accepted: *30 July 2021*
Published: *27 August 2021*

Citation:
Pascual-Ezama D, Muñoz A and Prelec D (2021) Do Not Tell Me More; You Are Honest: A Preconceived Honesty Bias. Front. Psychol. 12:693942. doi: 10.3389/fpsyg.2021.693942

INTRODUCTION

Being able to detect when someone is (dis)honest has always been a social goal. A lot of work has been done to identify when people lie and when they tell the truth. In areas like criminology, politics, negotiation, or even playing poker, detecting when someone is lying gives you a competitive advantage over your opponent. It has long been evident in literature that dishonest behavior, both lies (DePaulo et al., 1996) and deception (Weiss and Feldman, 2006), is everyday and frequent occurrences. Therefore, detecting it without the help of technology is essential for everybody in our day-to-day life.

The study of detecting dishonest behavior has come a long way with technology. Truth serums, polygraphs, eye movements, facial analysis, body temperature changes, MRIs, and many other techniques have been used to detect such unethical behavior in the past. More recently, individual physiological responses can offer clues to see dishonest behavior according to contactless non-invasive automatic technologies (also known as automatic deception detection in the literature). Among the different technologies, facial expressions have become one of the most studied features due to their high exposure (e.g., easy to record by a simple camera) and

the relevant information of micro-expressions associated with dishonest behavior (e.g., Ekman, 2009). To detect dishonesty, researchers have investigated the potential of automatic physiological approaches, such as a database of facial microexpressions (Pfister et al., 2011) or a method based on dynamic geometric features obtained from facial microexpressions (Owayjan et al., 2012). These earliest approaches demonstrated the capability of automatic systems to detect markers associated with dishonest misconduct. During the last decade, multimodal systems and new machine learning technologies have improved mechanical dishonesty detection performance. Multimodal systems exploit the complementarity of features obtained by a combination of different modalities, such as previously mentioned facial microexpressions, thermal imaging (Rajoub and Zwiggelaar, 2014; Abouelenien et al., 2017), voice (Mendels et al., 2017), and hand gestures (Maricchiolo et al., 2012). In conjunction with available data sets and machine learning algorithms, these multimodal approaches have boosted the performance of automatic systems of wicked recognition accuracy in some scenarios (Krishnamurthy et al., 2018).

However, when technology is not available, no other mechanisms guide us other than our intuition based on our experience to detect the behavior of the person in front of us. Sometimes, when we directly face our opponents, we have environmental or additional information that can help us: something a person has done, something a person has said, or some corporal gesture can give us information and help us have a criterion. It is also possible to ask questions that raise the cognitive load more in liars than in truth-tellers (Vrij et al., 2011). The receptor may likewise become aware of the lie if there are inconsistencies in the message, through verbal (VCs) or non-verbal cues (NVCs), or an investigation after the statement (Ekman, 2009; Vrij et al., 2010). However, many other times, when we only can see the face of the opponent or listen without interaction in the communication, we are able (or we think we are) to detect whether they are honest or dishonest at the time. It is with respect to this situation that we would like to contribute. We want to provide new data on how we are able to detect (dis)honesty when we only see faces of our opponents or when we hear them speak without further environmental interference. With this objective in mind, we put forward the following hypothesis:

H1: Our ability to detect (dis)honest behavior is directly related to the way we behave (dis)honestly.

To justify this hypothesis, we will use the existing literature about (dis)honesty detection. There have been two marked trends in the literature, one for and one against, about whether we can detect unethical behavior. There are few studies where we can observe indications that noticing the behavior of others is an elementary, innate ability (e.g., Willis and Todorov, 2006; Fiske et al., 2007; Miller, 2007). Nevertheless, a substantial finding in the deception detection literature indicates that people are not better than casually able to detect a liar (Bond and DePaulo, 2006). So, according to the literature, we should hypothesize that general accuracy will also be no better than chance in our research. However, in addition to analyzing general accuracy, we also want to analyze specificity (honesty detection) and sensitivity (dishonesty detection) since we believe that the ability to detect dishonest people does not necessarily have to be directly related to the ability to detect honest people. To fulfill our purpose, we will sort the literature by answering three fundamental questions: Who can detect dishonesty? How can dishonesty be detected? What information is necessary for detecting dishonesty?

Regarding who can detect (dis)honesty, there is hardly any literature analyzing whether profiles of people who are better able to detect (dis)honesty than others exist. Moreover, there is also no literature dealing with whether those people who are more (dis)honest are better able to detect (dis)honesty. Are dishonest people better at detecting dishonesty than honest people? Are honest people better at detecting honesty than dishonest people? Getting an answer to these questions is the first contribution we wish to make in this research article. With respect to the different profiles of dishonest people, we have the classification proposed by Fischbacher and Föllmi-Heusi (2008, 2013), which offers three types of profiles: honest, liars, and partial liars. In addition, Shalvi et al. (2011) found that when people were allowed to repeat a task more than once but only the first result was valid for reporting purposes, the highest outcome was sometimes reported (even if it was not the first one). Pascual-Ezama et al. (2020) found some additional profiles. In addition to the liars, they found cheater non-liars and radicals. The cheater non-liars did not lie: they reported the result they really obtained, but they obtained the result by repeating the task several times, thus breaking the rules. Even when the rules were strict with respect to doing the task only once (contrary to Shalvi et al., 2011, who permitted the task to be repeated), participants repeated it until they obtained the expected result. On the other hand, radicals reported the result without running the task. They simply reported a result and collected a reward without doing anything. Finally, and in line with Fischbacher and Föllmi-Heusi (2013), Pascual-Ezama et al. (2020) found non-maximizer (partial) profiles for all liars, cheater non-liars, and radicals. Both the strategic behavior of cheater non-liars and the drastic behavior of radicals show two very different patterns of behavior from that of liars. In addition, Pascual-Ezama et al.'s (2020) classification allows us to analyze the data according to four different classifications: first, we consider only whether people are honest or dishonest (simple classification); second, we take into account the different behaviors/strategies of liars, cheater non-liars, and radicals (by nature); third, we consider whether the participants have maximized their dishonesty (by gradient); and finally, we analyze the data according to the eight profiles, two of which are honesty profiles, and six are dishonesty profiles (full classification).

With respect to how dishonesty can be detected, the literature offers evidence for how we can better detect dishonest behavior indirectly, unconsciously (e.g., Reinhard et al., 2013; Brinke et al., 2014), whereas other articles deny this evidence and find the opposite results (see Bond and DePaulo, 2006 for a meta-analysis). Brinke et al. (2014) found some evidence for unconscious lie detection (done without one realizing how),

although Franz and von Luxburg (2015), in a critique of the results of the previous study, found evidence for unconscious lie detection but concluded that a significant difference does not imply accurate classification. Moreover, the literature shows that honest behavior (HB) detection is better done with indirect predictions than direct judgments (Vrij et al., 2001; Ulatowska, 2014). It has also been observed that quick, automatic, and subjective decisions make it possible to differentiate between honest and dishonest people much better than premeditated, thoughtful, and objective judgments (DePaulo et al., 2003; Albrechtsen et al., 2009). Taking this information into account and based on the "by nature" classification, we could assert that liars and cheaters are more strategic. They have to think about how they lie or cheat and what strategy they will follow and their decision-making will be more thoughtful and meditated. However, the behavior of radicals will be more automatic, as they do not have to think about their strategy and have clarity regarding what they want to report. Along the same line and based on "by gradient" classification, non-maximizers have a higher self-concept and are less strategic than maximizers, who act in a more meditated manner. Maximizers set their strategy in order to obtain the most money possible, their decision-making being completely objective. However, those who do not maximize due to their self-concept (Mazar et al., 2008) will make their decision-making in an automatic and more subjective way, being an emotional and not very meditated decision. Therefore,

H1(a): Radicals should be better than cheater non-liars at detecting (dis)honesty.
H1(b): Submaximizers should be better than maximizers at detecting (dis)honesty.

Finally, with respect to what information is necessary to detect dishonesty, there is an extensive literature that analyzes the ability to detect dishonesty in terms of the different cues available, mainly VCs and NVCs. There are widespread beliefs about how people behave when they act dishonesty: stereotypes about gender, ethnicities, or races and about whether dishonest people get nervous and act in a different way. It is also possible to discern information about status, dominance, romantic involvement, and relationship potential (Ambady et al., 2000). There is a general consensus that there has been an overemphasis on NVCs and that VCs are very relevant. One of the most contrasting results in the literature is that the combination of NVCs and VCs is the best way to detect dishonesty. However, the literature is focused on detecting dishonesty but not on detecting honesty. Our second contribution in this paper is to analyze not only dishonesty accuracy detection but also honesty accuracy detection. There is a consensus that VCs facilitate the detection of dishonesty (e.g., Vrij et al., 2010) in two ways: VCs in addition to NVCs (e.g., Ekman and O'Sullivan, 1991; Vrij et al., 2004) and a higher amount of VCs improve accuracy in detecting dishonesty (e.g., Anderson et al., 1999; Feeley and Young, 2000). So, we can affirm that dishonesty is better detected with more information and using VCs. Therefore, we hypothesize that

H2: Honesty should also be better detected using more information and verbal cues.

To confirm these hypotheses, we conducted a pilot study with 276 participants, in which we obtained very satisfactory preliminary data, and an experiment with more than 2,000 participants, in which they performed two tasks. The first task consisted of the adaptation of Pascual-Ezama et al. (2020) for the die-under-the-cup task of Fischbacher and Föllmi-Heusi (2013). We decided to use this task, as it is one of the most popular literatures (e.g., Abeler et al., 2019; Charness et al., 2019), With this task, we managed to classify participants according to different profiles of (dis)honesty. The second task consisted of watching a series of TV shows for which participants had to decide whether the contestants were honest or dishonest (other papers used videos: Belot et al., 2012; Serra-García and Gneezy, 2021). In this research paper, we aim to bring more evidence to the literature on detecting dishonest behavior in two ways. On the one hand, we want to examine if different (dis)honest people can detect better/worse (un)ethical behavior of others. We have focused our attention on general accuracy and sensitivity (dishonesty detection accuracy—DDA) and specificity (honesty detection accuracy—HDA) (Baratloo et al., 2015) to determine if different profiles can detect better honesty or dishonesty. On the other hand, we want to analyze if more information and different cues improve not only dishonesty detection but also honesty. Finally, we have detected a bias that makes us overestimate honesty and facilitates the detection of honesty and hinders the detection of dishonesty.

EXPERIMENT

Materials and Methods

Participants

To guarantee enough power in the analyses, we decided to run the experiment with a significant sample of about 2,000 participants. They were finally 2,050 individuals recruited by Amazon Mechanical Turk, who got $1.50 as a show-up fee and the opportunity to earn a $0.50 performance-based bonus in the first part of the experiment. Eighty-seven participants did not complete the task appropriately (did not complete the MTurk process with the MTurk code), so they were eliminated. Another 381 participants were not considered for the analysis, according to the exclusion criterion of Pascual-Ezama et al. (2020)[1] because they did not follow the rules of the experiment, and therefore we were unable to obtain sufficient information from these participants. The final sample to analyze (dis)honesty detection accuracy consisted of 1,582 participants: 43% were women, and the average age was 37 (SD = 11).

Materials and Procedure

Participants ran the experiment on the MTurk platform out of the lab, and they were paid according to their report on the

[1]Individuals who gave an immediate response (<5 s after receiving computerized instructions) without using www.rollandflip.com (see the following section) were classified as "radically dishonest" because they claimed the high money outcome without flipping a coin or rolling a die on some other website (5 s is insufficient time to go to an alternative website and/or initiate a coin toss or a die roll). Individuals who gave a report that took more than 5 s and also did not go to our website were eliminated from the final sample.

platform. They ran the experiment using the website http://www.behavioralexperiments.com and conducted the experiment in two completely distinct parts. Behavioralexperiments.com is a platform where any researcher can perform experiments. It offers the advantage that it automatically classifies participants according to the classification of Pascual-Ezama et al. (2020) based on their (dis)honesty profiles.

The first part of the experiment consisted of an adaptation of the die-under-the-cup task proposed by Fischbacher and Föllmi-Heusi (2013), using the new paradigm proposed by Pascual-Ezama et al. (2020). Participants were asked to roll the die in http://www.rollandflip.com or a similar website using their cell phone. They would only get no bonus if they got 6, following Fischbacher and Föllmi-Heusi's (2013) rewards system. So, using this task, participants could choose not only to be (dis)honest, but they could adapt it to different levels, from maximum to minimum reward. Every participant received the same message with simple and short instructions: "First, ensure you have a smartphone, a tablet, or another electronic device with internet access. You have to roll a die, and you can earn money depending on your roll result: if you roll a 1, you will receive 0.10$. If you roll a 2, you will receive 0.20$. If you roll a 3, you will receive 0.30$. If you roll a 4, you will receive 0.40$. If you roll a 5, you will receive 0.50$. If you roll a 6, you will receive nothing. Take your cell phone, go to the following website http://www.rollandflip.com/ (or another similar site), select "roll the die" option, and roll the die once." The critical manipulation here was to link the real outcome and the reported one for a given person. We had access to the rollandflip.com database to match the rolls individually, controlling the exact moment every participant performed the task. Therefore, we were able to determine the precise number of rolls and the real outcome distribution to link with the reported one for each participant. Although not all participants in the study chose to use the rollandflip.com website, most of them did so, allowing us to connect their real and reported outcomes to study honest and dishonest behavior in detail. The website www.rollanflip.com is a website created by researchers to record the real outcome, with the versions "flip the coin" or "roll the die." We were able to record the real results, IP, timestamp, the reported results, and the time participants took to complete the task. Therefore, we were able to link data from http://www.rollandflip.com with http://www.behavioralexperiments.com to classify real behavior of participants.

In the second part of the experiment, participants had to watch five different videos extracted from the popular TV show in the UK called "golden balls." In the last part of this program, two contestants have to select between two options. They have two golden balls, one of them has the word "split," and the other has the word "steal." If both contestants select split, they share the accumulated money (this varies depending on the evolution of each program). If one contestant selects split, and the other one selects steal, those who select steal obtain all the economic rewards, and the other gets nothing. But, if both contestants choose to steal, both get nothing. This objective of the experiment was to detect whether contestants were honest or dishonest in two different moments. The first moment was before talking (our participants could only see the faces of the contestants, whereas the presenter explained the rules without VCs). In this first moment, participants were asked to give their general opinion on whether they considered the contestants (both) to be honest or dishonest as a general concept. The second moment was after talking; each contestant tried to convince the other to split to open the golden ball with the split/steal option (NVCs + VCs). In this second moment, after hearing the contestants say that they would share the prize (they all do), the participants had to decide whether the contestants were really honest, that means, did they intend to share the prize as they had said (and choose the ball with the word split) or, on the contrary, would they be dishonest, and therefore, despite promising to share the prize, would they choose the steal ball to keep all the money. If the participants decide that a contestant is honest (honesty prediction; HP), and the contestant is honest (HB), the honesty detection (HDA) is considered to be correct. Otherwise, it would be incorrect. Therefore, honesty detection will be the percentage of times a participant detects an honest contestant divided by the total number of contestants who behave honestly (HDA = HP/HB). For example, since the number of dishonest contestants is controlled at 50%, there will be five honest contestants and five dishonest contestants. If a participant detects three of the five honest contestants, they will have an HDA = 3/5 = 60%. Similarly, if the participants decide that a contestant is dishonest (DP) and the contestant behaves dishonestly (DB), the dishonesty prediction (DDA = DP/DB) is considered to be correct. Otherwise, it would be incorrect. Participants also had to answer questions about the two contestants, and they were asked their gender and approximate age before the first question to make sure they did not confuse contestant one and contestant two. We controlled the videos in three ways: the duration of all videos was about 1 min; all participants watched the same videos—five videos with 10 contestants; the contestants were 50% honest and 50% dishonest[2]. We also controlled the race and gender of the contestants to avoid stereotypes. We decided not to financially incentivize this second part of the experiment because it has not been demonstrated that an increase in motivation due to a financial incentive can improve the ability to detect dishonesty. However, we did consider that the pressure to receive an economic incentive could increase anxiety and provoke unnatural decision-making.

Results

Before presenting the results, we had to be sure to replicate the gray-scale (dis)honesty classification of Pascual-Ezama et al. (2020) with six different dishonesty profiles. We used these profiles to analyze if any profile could detect (dis)honesty better than the others. In **Table 1**, we can see the profiles found. We used four different models established according to the following classifications: simple classification—taking into account only if people are honest or dishonest; full classification—taking into account the eight profiles, two of which are honesty profiles, and six are dishonesty profiles; by nature, considering the different behaviors/strategies of liars, cheater non-liars, and radicals;

[2]We repeated the procedure with random selection (70% honest and 30% dishonest contestants) with similar results.

TABLE 1 | Classification of participants according to their reported/actual results.

			MTurk	
			(n = 1,582)	(n = 1,389)
Roll the die–obtain 5–report 5		Lucky	12.2%	–
Roll the die–obtain x different than 5–report x	Lucky honest	Honest	36.6%	41.7%
Roll the die–obtain 6–report 6	Unlucky honest		8.8%	10%
Roll the die–obtain x different than 5–repeat until x < 5–report x	Submaximizing cheater non-liars	Cheater non-liars	7.8%	8.9%
Roll the die–obtain x different than 5–repeat until 5–report 5	Maximizing cheater non-liars		7.7%	8.8%
Roll the die–obtain x–report > x but < 5	Submaximizing liars	Liars	3.0%	3.4%
Roll the die–obtain x different than 5–report 5	Maximizing liars		5.6%	6.4%
Do not roll the die at all–report < 5	Submaximizing radical dishonest	Radical dishonest	10.3%	11.7%
Do not roll the die at all–report 5	Maximizing radical dishonest		8.0%	9.1%

*Again, gray rows show percentage results, including "Lucky" people. White rows show percentages of the total sample excluding "Lucky" people.

TABLE 2 | (Dis)honesty detection statistics.

Classification	F	p	η^2	Power
Honesty detection accuracy				
Simple	$F_{(1,\ 1387)} = 6.544$	0.011	0.005	0.725
By nature	$F_{(1,\ 1387)} = 6.887$	0.001	0.010	0.923
By gradient	$F_{(1,\ 1387)} = 10.389$	<0.001	0.022	0.999
Full	$F_{(1,\ 1387)} = 5.458$	<0.001	0.027	0.999
Dishonesty detection accuracy				
Simple	$F_{(1,\ 1387)} = 0.370$	0.847	0.001	0.054
By nature	$F_{(1,\ 1387)} = 0.272$	0.762	0.001	0.093
By gradient	$F_{(1,\ 1387)} = 0.120$	0.948	0.001	0.072
Full	$F_{(1,\ 1387)} = 0.732$	0.645	0.004	0.321

and by gradient—taking into account whether the participants maximized their dishonesty. We found all the profiles in this experiment, thus replicating the profiles of Pascual-Ezama et al. (2020).

Result 1: Submaximizers and Radicals Detect Honesty Better

General accuracy was not different than chance. Participants only guessed correctly about the behavior of the contestants 47% of the time, taking into account its 10 predictions ($p = 0.5$). No differences were found when we repeated the analyses with the simple classification (46 and 47% for honest and dishonest, respectively); when we analyzed by nature, we found 46, 41, 43, and 42%, for honest, liars, cheater non-liars, and radicals, respectively, and by the gradient, the results were 46, 47, and 47%, for honest, submaximizers, and maximizers, respectively. Similar results were found for the full classification. Therefore, and as we might expect according to the literature, the overall predictive ability was absent. We had similar results when we analyzed sensitivity (dishonesty detection) as shown in **Table 2**.

However, when we analyzed specificity (honesty detection), clear differences appeared in the different classifications (see ANOVA in **Table 2**). In the simple classification (t-test), we can see how dishonest people were better at detecting honesty than honest people (64 vs. 60%; $p = 0.01$). By nature, we can observe how radicals were better at detecting honesty than honest people (71 vs. 61%; $p < 0.001$), liars (71 vs. 59%; $p < 0.001$), and cheater non-liars (71 vs. 62%; $p = 0.013$). There were no differences among the rest of the groups. So, this result partially confirms our first hypothesis. Radicals are not better at detecting dishonesty than the rest, but they detect honesty better than any other profile. When we analyzed the data by gradient, we found that submaximizer dishonest people were better at detecting honesty than maximizers (68 vs. 62%; $p = 0.022$) and honest people (68 vs. 61%; $p < 0.001$). This result confirms our second hypothesis. Submaximizers also detected honesty better than any other profile.

Result 2: Additional Information Is Not Always Better

Using the single classification, we ran a 2 × 2 ANOVA with level of information (low with NVCs and high with NVCs + VCs) and honesty (honest and dishonest people), and we had two dependent variables: HDA and DDA. In HDA, we found the main effects on level of information and significant interaction but no effects on honesty (see **Table 3** for statistics). In DDA, we found the main effects on level of information, but no effects on honesty or interaction. There were significant differences between NVCs and VCs both for dishonest and honest people (both $p < 0.001$), both in HDA and DDA. In HDA, the accuracy of honest people is 61% with NVC and 57% with VC ($p < 0.001$), a similar result to dishonest people (65% NVC vs. 56% VC; $p < 0.001$). Opposite results were found when we analyzed DDA both for honest people (27% NVC vs. 42% VC; $p < 0.001$) and dishonest people (25% NVC vs. 42% VC; $p < 0.001$). When we repeated the analyses using the "by gradient" classification with a 2 × 3 ANOVA with level of information (low with NVCs and high with NVCs + VCs) and honesty (honest, submaximizers, and maximizers), we found similar results. A similar situation arose when we repeated the analyses using the "by nature" classification with a 2 × 4 ANOVA with level of information (low with NVCs and high with NVCs + VCs) and honesty (honest, liars, cheater non-liars, and radicals) (see **Tables 3**, **4**). Therefore, our second hypothesis should be rejected. Honesty is better detected with

TABLE 3 | Information use statistics.

	F	*P*	η^2	Power
HONESTY DETECTION ACCURACY				
Simple classification				
Level of information	$F_{(1, 1387)} = 63.74$	<0.001	0.044	0.999
Honesty	$F_{(1, 1387)} = 1.739$	0.188	0.001	0.261
Interaction	$F_{(1, 1387)} = 8.47$	0.004	0.006	0.829
By gradient classification				
Level of information	$F_{(1, 1387)} = 70.76$	<0.001	0.049	0.999
Honesty	$F_{(1, 1387)} = 5.008$	0.007	0.007	0.815
Interaction	$F_{(1, 1387)} = 4.608$	0.010	0.007	0.780
By nature classification				
Level of information	$F_{(1, 1387)} = 57.413$	<0.001	0.004	0.999
Honesty	$F_{(1, 1387)} = 3.905$	0.009	0.008	0.829
Interaction	$F_{(1, 1387)} = 10.706$	<0.001	0.023	0.999
DISHONESTY DETECTION ACCURACY				
Simple classification				
Level of information	$F_{(1, 1387)} = 509.892$	<0.001	0.269	0.999
Honesty	$F_{(1, 1387)} = 1.624$	0.203	0.001	0.247
Interaction	$F_{(1, 1387)} = 2.745$	0.098	0.002	0.381
By gradient classification				
Level of information	$F_{(1, 1387)} = 471.874$	<0.001	0.254	0.999
Honesty	$F_{(1, 1387)} = 1.488$	0.226	0.002	0.319
Interaction	$F_{(1, 1387)} = 1.504$	0.223	0.002	0.322
By nature classification				
Level of information	$[F_{(1, 1387)} = 386.796$	<0.001	0.218	0.999
Honesty	$F_{(1, 1387)} = 1.082$	0.355	0.022	0.295
Interaction	$F_{(1, 1387)} = 3.030$	0.028	0.007	0.715

low levels of information (NVC), whereas dishonesty is better detected with high information levels (NVC + VC).

Result 3: A "Preconceived Honesty Bias" Is Detected

Specificity (honesty detection) was better than chance both for honest people (58%; $p < 0.001$) and dishonest people (60%; $p < 0.001$), with no difference between submaximizers and maximizers or among liars, cheater non-liars, and radicals. On the other hand, sensitivity (dishonesty detection) was abnormally low again both for honest people (34%; $p < 0.001$) and dishonest people (34%; $p < 0.001$) with no difference between submaximizers and maximizers or among liars, cheater non-liars, and radicals. When we try to detect the behavior of others (dis)honest, we tend to think that honesty prevails, which leads us to have good accuracy in detecting honesty, thinking people are honest. However, we also think that dishonest people are honest, and this leads us to have an extremely poor success rate, much lower than random chance because of a "preconceived honesty bias."

More evidence to support the "preconceived honesty bias" arose from the difference, both in sensitivity and specificity, with the different levels of information. Having a preconceived bias toward honesty, participants detected honesty very well and dishonesty very poorly with low information. However, as people got more information, they became increasingly hesitant and more likely to think of dishonest behavior, thereby improving sensitivity (26–42%; $p < 0.001$) but significantly worsening specificity (62–56%; $p < 0.001$). Similar results were found for the "by nature" or "by gradient" classifications (see **Table 3**; $p < 0.001$ for all cases). There was a very pronounced tendency to assume honesty *a priori* when participants only had the visual information of the face of a person (between 22 and 27% in dishonesty detection; $p < 0.001$ for all). This could be a good explanation for why general accuracy is not different than chance at detecting dishonesty, as we can show in our first result and can be found in the literature.

DISCUSSION AND CONCLUSIONS

Dishonesty detection is complicated. Even professionals, who work to detect criminal behaviors, perform no better than chance when it comes to detecting dishonesty (e.g., Bond and DePaulo, 2006; Granhag et al., 2015; Serra-García and Gneezy, 2021). The results presented here provide similar results. In line with the literature, our results show how, first, when we try to detect dishonesty, general accuracy is not different than chance, and second, when we increase the amount of information, and VCs, the detection of dishonesty rises considerably although it is still far below chance. We can explain these results from two different points of view. On the one hand, deception could be better detected from multiple cues as has been suggested in many papers that processing a large number of cues could be more efficient (Hartwig and Bond, 2014). On the other hand, in the dishonesty detection literature, people display better performance when using VCs instead of NVCs (e.g., Bond and DePaulo, 2006; Reinhard and Schwarz, 2012). So, results found concerning dishonesty detection are in line with previous results in the literature: we can improve the detection of dishonesty even though it is still far inferior to randomness. However, in analyzing not only general accuracy but also sensitivity (DDA) and specificity (HDA), we discovered a "preconceived honesty bias" to explain these results in the literature.

We have found that results when trying to detect dishonesty just by looking at the face of a person without any other interaction are much lower than those which would correspond to a random outcome. Therefore, the use of basic NVCs not only does not facilitate the detection of dishonesty but also harms it. The literature regarding NVCs and VCs to deception is extensive. There is a consensus that VCs facilitate the detection of dishonesty (e.g., Vrij et al., 2010). Our results show that the natural tendency and predisposition to judge people just by looking at their faces leads us to decide that they are honest. The results repeatedly show that the rate of detection of dishonesty in these circumstances is about 25% when the capacity of random hitting would be double. Therefore, there is a clear "preconceived honesty bias" here that negatively affects the ability of a person to judge our contemporaries at the first glance correctly. However, the vast majority of work has focused on analyzing the ability to detect dishonesty. Still, it has not paid as much attention to the ability (or lack thereof) to detect honesty. In our paper, there are relevant results regarding the

TABLE 4 | Information use descriptive.

HONESTY DETECTION ACCURACY									
Simple model									
Honest ($N = 718$)				Dishonest ($N = 671$)					
NVC	VC			NVC	VC				
61%	57%			65%	56%				
$p < 0.001$	$p < 0.001$								
By gradient									
Honest ($N = 718$)						Submaximizer ($N = 334$)		Maximizer ($N = 337$)	
NVC	VC					NVC	VC	NVC	VC
61%	57%					**68%**	58%	**62%**	54%
$p < 0.001$						$p < 0.001$		$p < 0.001$	
By nature									
Honest ($N = 718$)				Liars ($N = 246$)		Cheater non-liars ($N = 136$)		Maximizer ($N = 289$)	
NVC	VC			NVC	VC	NVC	VC	NVC	VC
61%	57%			**59%**	57%	**62%**	54%	**71%**	56%
$p < 0.001$				$p < 0.089$		$p < 0.001$		$p < 0.001$	
DISHONESTY DETECTION ACCURACY									
Simple model									
Honest ($N = 718$)				Dishonest ($N = 671$)					
NVC	VC			NVC	VC				
27%	**42%**			25%	**42%**				
$p < 0.001$	$p < 0.001$								
By gradient									
Honest ($N = 718$)						Submaximizer ($N = 334$)		Maximizer ($N = 337$)	
NVC	VC					NVC	VC	NVC	VC
27%	**42%**					24%	**41%**	26%	**43%**
$p < 0.001$						$p < 0.001$		$p < 0.001$	
By nature									
Honest ($N = 718$)				Liars ($N = 246$)		Cheaters non-liars ($N = 136$)		Maximizer ($N = 289$)	
NVC	VC			NVC	VC	NVC	VC	NVC	VC
27%	**42%**			27%	**41%**	25%	**43%**	22%	**42%**
$p < 0.001$				$p < 0.001$		$p < 0.001$		$p < 0.001$	

Bold values indicate the highest between NVC and VC.

cues used for honesty detection. We offer innovative results demonstrating how honesty is well-detected using only NVCs. Again, we can observe the "preconceived honesty bias" in the predisposition to judge people as honest just by looking at their faces. However, more information, in this case, VCs, not only does not improve the ability to detect honesty but significantly worsens it.

The implications of these results are very relevant because if we use only NVCs, we detect honesty better than dishonesty, but with VCs, the contrary occurs. A famous saying is that there is no second chance to make a first impression. In terms of dishonesty detection, our results suggest that to have a correct opinion of our opponent, we should not be guided by that first impression, and we should accumulate more information by combining NVCs and VCs. However, in terms of detecting honesty, the first impression is the correct one. In terms of criminology, a guilty person should remain free in a guaranteed legal system than an innocent person should go to prison. Therefore, we could understand that it would be better to have less information and detect honest people correctly than to stop catching dishonest people. But this logic is not necessarily the right one to apply in the business environment. If we accept that the detection of (dis)honesty is unconscious (done without one realizing how), we have a threshold between detecting more honest or dishonest

people. It will depend on the process and on what is of interest at each moment. If we accept that the process is conscious, our results suggest that more research is necessary to understand what information makes our process of detecting honest people worse when we include VCs.

Finally, we found significant results indicating that corrupt people who do not maximize their unethical behavior can detect honesty much better than honest people or dishonest people who maximize their unethical behavior. Submaximizers and radicals are less strategic and act in a more emotional and less meditated manner, so they have a greater critical capacity when establishing their pros and cons of decisions. This situation may mean that they can interpret better the decision-making of the people they observe. They can only do so for honest behavior since dishonest behavior is harder to detect, but they do it much better than the rest. It could also happen that there are hidden variables that we still have not taken into account. For instance, they may be more intelligent either at the level of general intelligence or emotional intelligence, making it easier for them to detect honesty, which is easier to detect than dishonest behavior. This research will be one of the future lines that we will follow. In addition, the perception of contestants of what the counterpart is going to do could be irrelevant in their decision-making. In this case, whatever the reason for their dishonesty, the objective of our participants was to detect whether they would be honest or not, but it is interesting to analyze this situation in another future line of research. But independently of the cause for why submaximizers and radicals can detect better honesty, the fact that they can do it has important implications. In selecting jobs in which honesty is fundamental (casinos, nightlife, security, etc.), submaximizers should conduct interviews. Indeed, they are not honest; still, they are not extremely dishonest either, and their capacity for the correct selection of honest people (above the rest) would imply significant economic benefits. Likewise, they would be much more suitable to carry out negotiation processes since they would regulate the strategies for the profit of company better and better detect their honest behavior of opponents.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Universidad Complutense de Madrid. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

DP-E developed the study concept. Testing and data collection were performed by DP-E and AM. DP-E drafted the manuscript and DP provided critical revisions. All authors contributed to the study design, data analysis and interpretation, and approved the final version of the document for submission.

FUNDING

This study was made possible thanks to funding received from the Fulbright Commission Award FMECD-ST-2017, the RCC Harvard University 2018 Research Fellowship granted to DP-E, and the Santander-UCM Research Project PR108/20-22.

ACKNOWLEDGMENTS

The authors would like to thank Leslie John and Shannon Sciarappa for their help in the data collection.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2021.693942/full#supplementary-material

REFERENCES

Abeler, J., Nosenzo, D., and Raymond, C. (2019). Preferences for truth telling. *Econometrica* 87, 1115–1153. doi: 10.3982/ECTA14673

Abouelenien, M., Pérez-Rosas, V., Mihalcea, R., and Burzo, M. (2017). Detecting deceptive behavior via integration of discriminative features from multiple modalities. *IEEE Trans. Inf. Forensics Secur.* 12, 1042–1055. doi: 10.1109/TIFS.2016.2639344

Albrechtsen, J. S., Meissner, C. A., and Susa, K. J. (2009). Can intuition improve deception detection performance? *J. Exp. Soc. Psychol.* 45, 1052–1055. doi: 10.1016/j.jesp.2009.05.017

Ambady, N., Bernieri, F. J., and Richeson, J. A. (2000). Toward a histology of social behavior: judgmental accuracy from thin slices of the behavioral stream. *Advan. Exp. Soc. Psychol.* 32, 201–271. doi: 10.1016/S0065-2601(00)80006-4

Anderson, D. E., Ansfield, M. E., and DePaulo, B. M. (1999). "Love's best habit: deception in the context of relationships," in *The Social Context of Nonverbal Behavior*, eds P. Philippot, R. S. Feldman, and E. J. Coats (Cambridge: Cambridge University Press), 372–409.

Baratloo, A., Hosseini, M., Negida, A., and El Ashal, G. (2015). Simple definition and calculation of accuracy, sensitivity and specificity. *Emergency* 3, 48–49.

Belot, M., Bhaskar, V., and Van De Ven, J. (2012). Can observers predict trustworthiness? *Rev. Econ. Stat.* 94, 246–259. doi: 10.1162/REST_a_00146

Bond, C. F., and DePaulo, B. M. (2006). Accuracy of deception judgments. *Pers. Soc. Psychol. Rev.* 10, 214–234. doi: 10.1207/s15327957pspr1003_2

Brinke, L. T., Stimson, D., and Carney, D. R. (2014). Some evidence for unconscious lie detection. *Psychol. Sci.* 25, 1098–1105. doi: 10.1177/0956797614524421

Charness, G., Blanco-Jimenez, C., Ezquerra, L., and Rodriguez-Lara, I. (2019). Cheating, incentives, and money manipulation. *Exp. Econ.* 22, 155–177. doi: 10.1007/s10683-018-9584-1

DePaulo, B. M., Kashy, D. A., Kirkendol, S. E., Wyer, M. M., and Epstein, J. A. (1996). Lying in everyday life. *J. Pers. Soc. Psychol.* 70, 979–995. doi: 10.1037/0022-3514.70.5.979

DePaulo, B. M., Lindsay, J. J., Malone, B. E., Muhlenbruck, L., Charlton, K., and Cooper, H. (2003). Cues to deception. *Psychol. Bull.* 129, 74–118. doi: 10.1037/0033-2909.129.1.74

Ekman, P. (2009). *Telling Lies: Clues to Deceit in the Marketplace, Politics, and Marriage (revised edition)*. London: WW Norton and Company.

Ekman, P., and O'Sullivan, M. (1991). Who can catch a liar? *Am. Psychol.* 46, 913–920. doi: 10.1037/0003-066X.46.9.913

Feeley, T. H., and Young, M. J. (2000). The effects of cognitive capacity on beliefs about deceptive communication. *Commun. Q.* 48, 101–119. doi: 10.1080/01463370009385585

Fischbacher, U., and Föllmi-Heusi, F. (2013). Lies in disguise—an experimental study on cheating. *J. Eur. Econ. Assoc.* 11, 525–547. doi: 10.1111/jeea.12014

Fiske, S. T., Cuddy, A. J. C., and Glick, P. (2007). Universal dimensions of social cognition: warmth and competence. *Trends Cogn. Sci.* 11, 77–83. doi: 10.1016/j.tics.2006.11.005

Franz, V. H., and von Luxburg, U. (2015). No evidence for unconscious lie detection: a significant difference does not imply accurate classification. *Psychol. Sci.* 26 1646–1648. doi: 10.1177/0956797615597333

Granhag, P. A., Rangmar, J., and Strömwall, L. A. (2015). Small cells of suspects: eliciting cues to deception by strategic interviewing. *J. Investig. Psychol. Offender Profiling* 12, 127–141. doi: 10.1002/jip.1413

Hartwig, M., and Bond, C. F. Jr. (2014). Lie detection from multiple cues: a meta-analysis. *Appl. Cogn. Psychol.* 28, 661–676. doi: 10.1002/acp.3052

Krishnamurthy, G., Majumder, N., Poria, S., and Cambria, E. (2018). An in-depth learning approach for multimodal deception detection. *arXiv preprint arXiv*

Maricchiolo, F., Gnisci,A., and Bonaiuto, M. (2012). Coding hand gestures: a reliable taxonomy and a multi-media support. *Cogn. Behav. Syst.* 7403, 405–416. doi: 10.1007/978-3-642-34584-5_36

Mazar, N., Amir, O., and Ariely, D. (2008). The dishonesty of honest people: a theory of self-concept maintenance. *J. Mark. Res.* 45, 633–644. doi: 10.1509/jmkr.45.6.633

Mendels, G., Levitan, S. I., Lee, K. Z., and Hirschberg, J. (2017). Hybrid acoustic-lexical in-depth learning approach for deception detection. *Proc. Inter. Speech* 2017, 1472–1476. doi: 10.21437/Interspeech.2017-1723

Miller, G. F. (2007). Sexual selection for moral virtues. *Q. Rev. Biol.* 82, 97–125. doi: 10.1086/517857

Owayjan, M., Kashour, A., Al Haddad, N., Fadel, M., and Al Souki, G. (2012). "The design and development of a lie detection system using facial micro-expressions," in *Advances in Computational Tools for Engineering Applications (ACTEA), 2012 2nd International Conference on Computational Tools for Engineering Applications* (Beirut: IEEE), 33–38.

Pascual-Ezama, D., Prelec, D., Muñoz-García, A., and Gil-Gómez de Liaño, B. (2020). Cheaters, liars, or both? a new classification of dishonesty profiles. *Psychol. Sci.* 31, 1097–1106. doi: 10.1177/0956797620929634

Pfister, T., Li, X., Zhao, G., and Pietikäinen, M. (2011). "Recognizing spontaneous facial micro-expressions," in *Computer Vision (ICCV), 2011 IEEE International Conference on Computer Vision* (Barcelona: IEEE), 1449–1456.

Rajoub, B. A., and Zwiggelaar, R. (2014). Thermal facial analysis for deception detection. *IEEE Trans. Inf. Forensics Secur.* 9, 1015–1023. doi: 10.1109/TIFS.2014.2317309

Reinhard, M., and Schwarz, N. (2012). The influence of affective states on the process of lie detection. *J. Exp. Psychol. Appl.* 18, 377–389. doi: 10.1037/a0030466

Reinhard, M.-A., Greifeneder, R., and Scharmach, M. (2013). Unconscious processes improve lie detection. *J. Pers. Soc. Psychol.* 105, 721–739. doi: 10.1037/a0034352

Serra-García, M., and Gneezy, U. (2021). Mistakes, Overconfidence, and the Effect of Sharing on Detecting Lies. *Am. Econ. Rev.* (in press). Available online at: https://www.aeaweb.org/articles?id=10.1257/aer.20191295

Shalvi, S., Dana, J., Handgraaf, M. J. J., and De Dreu, C. K. W. (2011). Justified ethicality: observing desired counterfactuals modifies ethical perceptions and behavior. *Organ. Behav. Hum. Decis. Process.* 115, 181–190. doi: 10.1016/j.obhdp.2011.02.001

Ulatowska, J. (2014). Different questions—different accuracy? the accuracy of various indirect question types in deception detection. *Psychiatry Psychol. Law* 21, 231–240. doi: 10.1080/13218719.2013.803278

Vrij, A., Edward, K., and Bull, R. (2001). Police officers' ability to detect deceit: the benefit of indirect deception detection measures. *Legal Criminol. Psychol.* 6, 185–196. doi: 10.1348/135532501168271

Vrij, A., Evans, H., Akehurst, L., and Mann, S. (2004). Rapid judgements in assessing verbal and nonverbal cues: their potential for deception researchers and lie detection. *Appl. Cogn. Psychol.* 18, 283–296. doi: 10.1002/acp.964

Vrij, A., Granhag, P., and Porter, S. (2010). Pitfalls and opportunities in nonverbal and verbal lie detection. *Psychol. Sci. Public Interest* 11, 89–121. doi: 10.1177/1529100610390861

Vrij, A., Granhag, P. A., Mann, S., and Leal, S. (2011). Outsmarting the liars: toward a cognitive lie detection approach. *Curr. Dir. Psychol. Sci.* 20, 28–32. doi: 10.1177/0963721410391245

Weiss, B., and Feldman, R. S. (2006). Looking good and lying to do it: deception as an impression management strategy in job interviews. *J. Appl. Soc. Psychol.* 36, 1070–1086. doi: 10.1111/j.0021-9029.2006.00055.x

Willis, J., and Todorov, A. (2006). First impressions: making up your mind after a 100-ms exposure to a face. *Psychol. Sci.* 17, 592–598. doi: 10.1111/j.1467-9280.2006.01750.x

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychology

ORIGINAL RESEARCH
published: 20 September 2021
doi: 10.3389/fpsyg.2021.732184

Psychopathy and Economic Behavior Among Prison Inmates: An Experiment

Loukas Balafoutas[1], *Aurora García-Gallego*[2], *Nikolaos Georgantzis*[2,3], *Tarek Jaber-Lopez*[4]* *and Evangelos Mitrokostas*[5]

[1] Department of Public Finance, University of Innsbruck, Innsbruck, Austria, [2] Department of Economics, Universitat Jaume I, Castellón de la Plana, Spain, [3] Burgundy School of Business-School of Wine & Spirits Business, Dijon, France, [4] Economix, Université Paris Lumière, Univ Paris Nanterre, Centre National Recherche Scientifique, Nanterre, France, [5] University of Portsmouth, Portsmouth, United Kingdom

This paper investigates whether there is a connection between psychopathy and certain manifestations of social and economic behavior, measured in a lab-in-the-field experiment with prison inmates. In order to test this main hypothesis, we let inmates play four games that have often been used to measure prosocial and antisocial behavior in previous experimental economics literature. Specifically, they play a prisoner's dilemma, a trust game, the equality equivalence test that elicits distributional preferences, and a corruption game. Psychopathy is measured by means of the Levenson Self-Report Psychopathy Scale (LSRP) questionnaire, which inmates filled out after having made their decisions in the four games. We find that higher scores in the LSRP are significantly correlated with anti-social behavior in the form of weaker reciprocity, lower cooperation, lower benevolence and more bribe-oriented decisions in the corruption game. In particular, not cooperating and bribe-maximizing decisions are associated with significantly higher LSRP primary and LSRP secondary scores. Not reciprocating is associated with higher LSRP primary and being spiteful with higher LSRP secondary scores.

Keywords: psychopathy, pro-social behavior, prison inmates, lab-in-the-field, experiment

OPEN ACCESS

Edited by:
Bojana M. Dinic,
University of Novi Sad, Serbia

Reviewed by:
Valerio Capraro,
Middlesex University, United Kingdom
Cesar Mantilla,
Rosario University, Colombia

***Correspondence:**
Tarek Jaber-Lopez
tarekjaberlopez@gmail.com

Specialty section:
This article was submitted to Personality and Social Psychology, a section of the journal Frontiers in Psychology

Received: *28 June 2021*
Accepted: *23 August 2021*
Published: *20 September 2021*

Citation:
Balafoutas L, García-Gallego A, Georgantzis N, Jaber-Lopez T and Mitrokostas E (2021) Psychopathy and Economic Behavior Among Prison Inmates: An Experiment. Front. Psychol. 12:732184. doi: 10.3389/fpsyg.2021.732184

INTRODUCTION

The World Prison Population List[1] gives its readers information on the number of prisoners held in the territories of 222 countries worldwide. Although comparability of imprisonment rates across countries must be subject to caution, data show that the overall prison population has been increasing in the last four decades. The United States of America currently hold over 2.3 million people in prison (Sawyer and Wagner, 2020), which represents the highest prison population rate in the world. The costs for correctional spending and crime combat are the fastest growing budget item after Medicaid (Henrichson and Delaney, 2012). Calculating the costs of criminal activity is quite difficult, since they vary widely among various offense categories. For instance, estimates place the total cost of crime in England and Wales at £60 billion in the year 2000 (Brand and Price, 2020). The vast amount of costs generated by criminal incidents and the attempt to administrate its consequences make it necessary to better understand the underlying nature of criminal behavior.

[1] http://www.prisonstudies.org/sites/default/files/resources/downloads/world_prison_population_list_11th_edition_0.pdf

Anti-social and criminal behavior can partly be explained by various personality disorders. One of them is psychopathy, a personality disorder defined by a lack of empathy for others. This disorder is related to antisocial disposition and characterized by having impaired empathy or lack of remorse, egotistical personality traits and sometimes even expressing cold blooded behavior toward others. Brandt et al. (1997) estimate that the base rate of psychopathy among prison population is as high as 37%. This high prevalence of psychopathic traits means that examining the relationship between such traits and behavior among criminals is of particular value.

The objective of this study is to investigate whether there is a connection between psychopathy and certain manifestations of social and economic behavior, measured in a lab-in-the-field experiment with inmates. To the best of our knowledge, this is the first study connecting psychopathic traits with social and economic behavior in a prison environment. Hence, our main research question is: Can psychopathic traits explain social and economic behavior among inmates? Many dilemmas in economic situations involve a conflict between selfish monetary reward maximization and devotion to ethical, pro-social norms connected to inferior economic benefit. The existence of ethical behavioral patterns among institutionalized subjects is of great interest as the starting point of rehabilitation and social inclusion strategies based on the principle that everyone is ethical to some extent.

To answer our research question, we use four games that have often been used to measure prosocial and antisocial behavior in the experimental economics literature: a prisoner's dilemma (henceforth PD), a trust game (henceforth TG), the equality equivalence test that elicits distributional preferences (henceforth EET), and a corruption game (henceforth CG). This choice of games is motivated by the fact that trust, reciprocity, cooperativeness, and distributional preferences are behavioral traits of essential importance for a successful rehabilitation of inmates into social and professional life after their release from prison (see Balafoutas et al., 2020, for a discussion). In addition, our study is the first to collect data on inmates' actions in a game meant to capture essential aspects of a corruption setting. Data from the corruption game allow us to study inmates' decisions when facing a social dilemma that includes an ethical component. We correlate behavior in all these games to a measure of psychopathy based on the Levenson Self-Report Psychopathy Scale (henceforth LSRP).

The data collection took place as part of a lab-in-the-field experiment conducted with 176 inmates in two prisons in Chania, Greece. Inmates played the games described above, in a number of sessions conducted within prison and following standard experimental protocol (regarding randomization, anonymity, and the use of monetary incentives). It is important to note that not all inmates played the four games. Out of the 176 inmates, 71 were recruited in the 2015 sessions and decided only on the CG, and 105 were recruited in 2016–2017 and decided on the TG, PD, and EET. The behavioral data from the economic games are complemented by administrative and survey data, including the LSRP.

Our results reveal that psychopathy as measured in the LSRP explains several aspects of inmates' social behavior. Higher scores in the LSRP are significantly correlated with anti-social behavior in the form of weaker reciprocity, lower cooperation, lower benevolence (implying a higher likelihood that a person is classified as having spiteful distributional preferences), and more bribe-oriented decisions in the corruption game.

LITERATURE REVIEW

Economic Experiments in Prisons

Despite the large economic and social costs of crime and the importance attributed by society and policymakers on the rehabilitation of criminal offenders, there is a relative scarcity of economic research on the behavior of prison inmates. Recently, a few studies using experimental economic methods have successfully overcome the practical and administrative challenges linked to this kind of research, yielding valuable insights on several aspects of the social and economic behavior of prison populations. One lesson that can be drawn from this literature is that differences in pro-social behavior (mainly altruism and cooperativeness) between prison inmates and samples of non-criminals are not systematic or consistent. Some studies find either very small, or negligible differences (Birkeland et al., 2014; Chmura et al., 2016), while others suggest that prison inmates are less pro-social than other groups of participants (Clark et al., 2015), or even more pro-social in some cases (Khadjavi and Lange, 2013; Nese et al., 2016).

Besides documenting patterns of behavior among inmates and comparing them to different samples, a few recent studies in prisons have considered topics such as the deterrence effect of punishment on antisocial behavior (Khadjavi, 2015), criminal identity and ethical behavior (Cohn et al., 2015), and the existence of in-group bias within a stigmatized group such as prison inmates (Balafoutas et al., 2020). Guo et al. (2020) differentiate between inmates' behavior toward an in-prison and out-of-prison sample and show that a simple priming intervention can promote rehabilitation by strengthening inmates' pro-social behavior toward the out-group. The current study uses, in part, the same sample as Balafoutas et al. (2020), but it studies an entirely different and hitherto unanswered question on the relationship between psychopathy and behavior among prisoners.

Psychopathy and Economic Behavior

Cleckley (1956) defines psychopathy as being manipulative, egocentric, impulsive, deceitful, and exhibiting antisocial behavior. The partial overlap between this definition and the purely self-interest notion of *homo oeconomicus* initiated a series of studies investigating whether psychopathy or psychopathic personality traits are linked to entrepreneurial abilities and success. Babiak et al. (2010) estimated that the general psychopathy prevalence is three times higher among the business workforce compared to the general population. Akhtar et al. (2013) argued that a certain degree of manipulativeness and callousness, both psychopathic characteristics, can be necessary for high achievements in a respective business field. Walters

(2004) remarks that the primary psychopathic personality traits such as "superficial charm, deceit, lack of guilt read like the job description of a good car salesman or a politician" (page 144). Akhtar et al. (2013) report moderate correlation between entrepreneurial activities and psychopathy but provide only weak support for the stereotype of a "corporate psychopath." They find that primary psychopathy is negatively correlated to "social entrepreneurship," i.e., initiate social activities such as improving the community, enhance education, or create student organizations. Similar conclusions have been obtained by Hassall et al. (2015) who measure academic success and psychopathic personality traits among business and psychology students and find that business students score significantly higher on psychopathy scores—albeit without a significant effect on academic success. One lesson that emerges from this strand of the literature is that we need to better understand psychopathic traits.

The literature in experimental economics that relates psychopathy to behavior in economic games is rather scarce, and at the same time highly relevant for our work. In a lab experiment, Ibáñez et al. (2016) study the relationship between emotions and trust. As a sign of the manipulative stage of a psychopath's behavior, they find that higher psychopathy scores are correlated with non-reciprocal decisions. A similar branch of the literature has examined the relationship between psychopathy and cooperative behavior in economic games. Mokros et al. (2008) find that psychopaths in a high-security psychiatric hospital behave in a non-cooperative manner in a prisoner's dilemma. Montañes et al. (2003) use various modifications of the prisoner's dilemma and show that Antisocial Personality Disorder correlates with non-cooperative behavior. Rilling et al. (2007) use a sample of 30 non-clinical subjects whose psychopathy is assessed using LSRP scores and Psychopathic Personality Inventory (PPI). In the repeated version of the prisoner's dilemma, they find a high correlation between non-cooperative behavior and higher LSRP scores among the male participants of their sample. On the contrary, they find no effect of psychopathy measured by the PPI. Curry et al. (2011) report that individuals with higher scores in the Machiavellian Egocentricity subscale of the PPI are less likely to behave cooperatively. Hence, considering cooperative behavior as a metric of empathy and a pro-social inclination, research so far indicates that psychopathy relates negatively to cooperation in social dilemma situations[2,3]. Our paper contributes to this branch of the literature by being the first to examine the relationship between psychopathy and various measures of prosocial behavior in a sample of imprisoned subjects with a verified criminal record.

Psychopathy and Criminal Behavior

The definition of psychopathy by Cleckley (1956) suggests that a number of negatively perceived personality traits should be considered core characteristics of psychopaths. On the other hand, observing antisocial behavior clearly is not sufficient to categorize someone as a psychopath. Most prison inmates, for instance, would be considered as antisocial to a certain degree, while only a minority of them expresses psychopathic personality disorders (Levenson et al., 1995). Nevertheless, antisocial behavior, impulsivity, lack of remorse, and the proneness toward violence is often seen as an explanation for why psychopaths tend to show more aggressive behavior among the institutionalized population. Vaughn et al. (2009) examine the potential subtypes of psychopathy among incarcerated juveniles and find that offenders scoring high in psychopathic measures indicate a greater likelihood of participating in self and other-destructive behavior than non-psychopathic juveniles. Compared to other criminals, psychopaths commit a significantly higher number of crimes and more violent ones (Hare and McPherson, 1984). Although some researchers have opposed these findings and conceded psychopathy only a limited role in crime forecasting, they nevertheless acknowledge the need for further research into psychopathic personality traits and the behavior of criminals (Walters, 2004).

Forecasting the likelihood of criminal acts would be without doubt a useful tool for police resource allocation and a potential way to reduce costs caused by the imprisoned population. Hence, improving crime prediction using models based on regularity and space clusters combined with psychological risk assessments that predict antisocial behavior should be considered (Brinkley et al., 2008; Johnson, 2010). Models developed for crime forecasts presently concentrate on social status, locality and crime opportunity, but individual characteristics are growing in importance, especially since crimes committed in affect are hard to account for (Miller et al., 2008; Johnson, 2010). Current research appears to regard psychopathy as a promising indicator for violence even among the female population (Levenson et al., 1995)[4].

It is worth noting that Miller et al. (2008) and other researchers (Porter et al., 2001; Skeem et al., 2007) assume that primary psychopaths are born with such a predisposition, whereas secondary psychopaths are believed to be shaped by their environment. The first to introduce such a distinction was Karpman (1948) who proposed a re-orientation of the concept of psychopathic personality. Specifically, he suggested to divide it into two main groups: the symptomatic or secondary psychopathy, and the primary, essential, or idiopathic psychopathy. Under the heading of secondary psychopathy are included the psychoses and neuroses that have a strong antisocial or delinquent aspect. Individuals of the other, primary group, suffer from a disease of its own designated as anethopathy. This is a mental disease, characterized by a personality organization having in particular a virtual absence of any redeeming social reaction (conscience, guilt, binding and generous emotions, etc.), while purely egoistic, uninhibited instinctive trends are

[2]However, pro-social behavior may also relate to selfish inter-temporal cooperation (collusion), or even a subject's risk attitudes. See for instance Sabater-Grande and Georgantzis (2002).

[3]Gillespie et al. (2013) and Spitzer et al. (2007) consider the connection between psychopathy and ultimatum games.

[4]Regarding the role of gender, we note the existence of strong evidence that psychopathic personality traits as egocentrism, manipulativeness, etc. manifest quite differently among the genders (Brinkley et al., 2008; Croson and Gneezy, 2009).

predominant. These are as close to the constitutional as can be found. Despite the clarity of past literature, there are still many empirical studies that investigate the frontier between primary and secondary psychopathy[5]. Further research will shape the view of psychopathy as either being an inborn or a molded personality disorder.

Measuring Psychopathy

There exists a diverse selection of measurement tools to assess psychopathy. Two popular ones are the Hare Psychopathy Checklist-Revised (PCL-R) also sometimes called Psychopathy Checklist—revised (Hare and Neumann, 2006) and the Levenson Self-Report Psychopathy Scale (Levenson et al., 1995). The PCL-R is constructed as a semi-structured interview and additionally uses official records (Hare and Neumann, 2006). It is capable of addressing both primary and secondary, psychopathic subgroups (Hare and Neumann, 2006). From a cost effectiveness perspective, the PCL-R has some notable disadvantages. For example, it is necessary that a trained clinical expert executes the interview, which is relatively time consuming (Lynam et al., 1999; Brinkley et al., 2001). Moreover, its development is primarily based on male offenders and requires historical records (Lynam et al., 1999; Brinkley et al., 2001).

Based on Karpman's (1948) initial distinction, Levenson et al. (1995) studied antisocial dispositions among non-institutionalized populations and developed the well-known Levenson Self-Report Psychopathy scale (LSRP). The LSRP is a self-report measure which was designed to assess primary and secondary psychopathic features in non-institutionalized populations. It is an advantage that it does not require historical crime records. Lynam et al. (1999) regard, based on their findings, the LSRP-Scale as a reasonable measure for psychopathy in context of variant measurements. Miller et al. (2008) conclude that the LSRP is significantly related to personality traits commonly seen in psychopathic individuals such as agreeableness and narcissistic behavior. Furthermore, the LSRP is strongly correlated with negative emotionality and other personality disorder symptoms.

Hence, both self-report tests (PCL-R as the LSRP) are capable of measuring psychopathic tendencies reliably to various degrees (Zolondek et al., 2006; Brinkley et al., 2008; Miller et al., 2008; Becker et al., 2012). Given that the LSRP is less time consuming and does not require historical crime records, we selected it for the present study.

EXPERIMENTAL DESIGN

Our experimental design is based on four simple economic games[6] and the Levenson Self-Report Psychopathy (LSRP) scale, supplemented by a collection of socio-demographic data, questions related to inmates' experience inside the prison and data provided by the prison administration.

TABLE 1 | The prisoner's dilemma.

		Player 2	
		Defect	**Cooperate**
Player 1	Defect	3, 3	9, 1
	Cooperate	1, 9	7, 7

The Games

Trust Game

We use a discrete version of the trust game (Berg et al., 1995). Subjects are matched in groups of two and are randomly assigned one of two roles in a between-subjects design: player 1 (sender), or player 2 (receiver). The sender has two strategies, to trust or not to trust the receiver. If he does not trust, both players earn an outside option of €10 each. If he trusts, the total available surplus is doubled (€40) and the receiver is then asked to take one of two actions: she can either reciprocate the sender's trust by implementing an equal split of €20 for each player, or choose the non-reciprocal action and keep €35 for herself, leaving the sender with only €5[7]. While trust and reciprocity lead to an improvement and a doubling of payoffs for both players, the subgame perfect equilibrium prediction for this game is that receivers never reciprocate trust, and anticipating this, senders never trust[8].

Prisoner's Dilemma

We use the same version of the simultaneous prisoner's dilemma as Balafoutas et al. (2020) and Khadjavi and Lange (2013), depicted in **Table 1**. Two players simultaneously decide either to cooperate with the other player or to defect. The dominant strategy for both players—and hence the Nash equilibrium—is defection, while choosing to cooperate is the pro-social action that leads to a Pareto improvement in payoffs if it is chosen by both players.

Equality Equivalence Test

In contrast to all other games, the Equality Equivalence Test (Kerschbamer, 2015) entails no strategic interaction. This test elicits distributional preference types by asking each subject to make ten binary choices between an equal and an unequal allocation, involving an own payoff and a payoff for a randomly matched subject. The ten choices are shown in **Table 2**,

[5]For instance, Vaughn et al. (2009) demonstrate that young offenders who have been identified with strong expressions of the secondary subtype were more likely to have experienced trauma and abuses in their past, thus supporting the assumption that secondary psychopathy is possibly caused by environmental factors.

[6]Sessions were conducted in different years and although the LSRP test was filled in all sessions, not all four games were applied for all the sample. See subsection 3.3 for details on this aspect of our experimental procedures.

[7]We implemented the strategy method for collecting data on receivers' choices, which means that they were asked to make a choice between the two possible allocations for the event that the sender they were matched with decided to trust them.

[8]It should be noticed that pro-social choices by senders and receivers in the trust game can arise from several motivations, the identification of which is beyond the scope of this work (Cox, 2004; Isoni and Sugden, 2019). The literature commonly refers to such choices as trust (in the case of senders) and reciprocity or trustworthiness (in the case of receivers).

TABLE 2 | The equality equivalence test (EET).

Left				Right	
You	Another person gets			You	Another person gets
Disadvantageous Inequality Block					
3.2	5.2	LEFT	RIGHT	4	4
3.6	5.2	LEFT	RIGHT	4	4
4	5.2	LEFT	RIGHT	4	4
4.4	5.2	LEFT	RIGHT	4	4
4.8	5.2	LEFT	RIGHT	4	4
Advantageous inequality block					
3.2	2.8	LEFT	RIGHT	4	4
3.6	2.8	LEFT	RIGHT	4	4
4	2.8	LEFT	RIGHT	4	4
4.4	2.8	LEFT	RIGHT	4	4
4.8	2.8	LEFT	RIGHT	4	4

broken down into a disadvantageous inequality block and an advantageous inequality block, referring to the direction of inequality as seen from the perspective of the decision maker. The ten choices, and in particular the row at which the subject switches from the equal to the unequal allocation, allow us to classify all subjects into one of four basic distributional preference types: altruistic (or efficiency loving), inequality averse, spiteful, and inequality loving[9].

CG

The Corruption Game (CG) framework studied here is based on Jaber-López et al. (2014). In a framed interaction protocol, two subjects in the role of "firms" bid in quality (Q) and bribe (B) levels (both in integers ranging between 0 and 10, $Q + B = 10$), for the procurement of a "public project," the quality of which is beneficial to all players within a group and individually profitable to the winning firm. A third subject in the role of a "public official" chooses the winning proposal having full information on the two firms' bids. Payoffs in the CG are determined as follows:

$$\Pi_{official} = 10 + \frac{1}{2}Q_{winner} + B_{winner}$$

$$\Pi_{winner} = 10 + \frac{1}{2}Q_{winner} - 2B_{winner} + 10$$

$$\Pi_{loser} = 10 + \frac{1}{2}Q_{winner}$$

Assuming rational and selfish subjects, there are three pure strategy Nash equilibria for firms' behavior with a discrete strategy space in this game: $(Q, B) = (7, 3)$, $(Q, B) = (6, 4)$, and $(Q, B) = (5, 5)$. Rational and selfish public officials maximize earnings therefore the subgame perfect Nash equilibrium predicts that they will choose the firm that offers the highest bribe. This framework represents a social dilemma, in the form of a tradeoff between quality and bribes. For firms, higher bribe payments indicate lower pro-sociality, since they imply sacrificing social welfare in the interest of increasing one's likelihood of winning the prize. For public officials, bribe-maximizing (as opposed to quality-maximizing) choices capture selfish, anti-social behavior, while officials driven by pro-social motives may sacrifice part of their own monetary earnings in favor of a higher quality project.

Levenson Self-Report Psychopathy Scale and Questionnaires

As already mention in section Measuring psychopathy, the evaluation of psychopathy in our paper is based on the LSRP (Levenson et al., 1995). Respondents state their degree of agreement with each of 26 statements, on a Likert-scale ranging from 1 ("totally disagree") to 4 ("totally agree")[10]. One attractive feature of this scale is that it elicits the level of psychopathic elements in a respondent's personality by offering three types of information, namely an aggregate measure of psychopathy (comprising all 26 questions) and two specific ones: *primary*, which refers to selfishness, lack of caring, manipulation of others and callous attitudes and is based on the first 16 questions; and the *secondary* psychopathy scale, associated with an impulsive, volatile or self-destructive personal style and is based on the last 10 questions (see Levenson et al., 1995; Lynam et al., 1999). All questions are shown in **Supplementary Material 3**.

Psychopathy by definition comprises manipulative and abusive behavior. In particular, individuals who display psychopathic traits are considered to be able to manipulate others in order to achieve personal benefits, without guilt or unfairness entering their moral considerations. Therefore, in our experiment, we expect higher scores on the psychopathy scale to be associated with less cooperative and pro-social behavior.

Inmates were asked to fill out the LSRP questionnaire after having made their decisions in the four games (TG, PD, EET, and CG). They were also asked to provide socio-demographic information (on their nationality, age, marital status, education level, and number of siblings). Additionally, we asked them to answer some questions regarding the conditions of their imprisonment: time spent in the current prison, number of times imprisoned, total time spent in prison during their life, type and length of sentence, attendance of religious activities in prison, number of cell mates, frequency of leaving the prison (for any reason) and number of working days per month. The prison administration provided us with this same information, allowing us to double check and correct for minor discrepancies.

Procedures

In January 2015 we ran one session in the low security agricultural prison facility of "Agia" and one session in the high security prison facility "Crete 1," in which subjects played only the CG. In November 2016 we ran one session in the high security prison and two simultaneous sessions in the low security agricultural prison, in which subjects played the PD, TG, and

[9]For more details on the classification of types, see Kerschbamer (2015). Note that selfish subjects are a subset of the four other categories and that including them as a separate category does not affect any of our findings.

[10]The Likert-scale items are phrased so as to minimize indication of disapproval for item endorsement.

EET. In April 2017 we conducted an additional session in the low security prison and again subjects played PD, TG, and EET. In all sessions, subjects also filled out the LSRP questionnaire. We recruited volunteer male inmates by posting announcements around the prison premises. Additionally, 2 days before each session, the experimenters went to the prison to answer possible questions and give a short explanation of what is an economic experiment. Once they decided to participate, inmates had to register through the prison administration.

All sessions took place either in the prison's gym or in the library. No guards were present and we insisted on and guaranteed subjects' anonymity, by giving them a random number so there was no way to associate a decision with a name. We were very cautious in minimizing any kind of audience effects. The experiment was conducted with pencil and paper. Subjects could choose among four different languages for their booklet of instructions: Greek, English, Arabic or French[11]. We enforced the usual experimental practice of not allowing for communication among subjects and ensuring anonymity in decision making. Once the session was ready to start, one of the experimenters explained aloud the general instructions of the experiment and answered possible questions. Subjects were told that one game would be chosen randomly by the social worker at the end of the session. Given that inmates are not allowed to receive money directly, we explained to them that their payment would be credited to their personal prison account, which can be used to buy goods inside the prison.

Afterwards, the experiment started and participants were asked to keep silent until the end of the session. In the sessions conducted in 2016 and 2017, we randomized the order in which the PD and TG were presented and played, although we kept the EET always as the third game. The instructions for each game were read in silent by each subject and they could go through the booklet at their own pace. Three experimenters were present in each session in order to answer any question in private and to assist participants. After making all decisions and filling out the questionnaires, participants left the session and received their payment one day later[12].

Our sample consists of 176 inmates in total. The mean age of inmates is 36.40 years old, they have 4.21 siblings and 1.09 children on average, and 52% of them are married. The mean sentence is 20.06 years and the remaining sentence is 10.46 years on average[13]. Out of the 176 inmates, 71 were recruited in the 2015 sessions and decided only on the CG, and 105 were recruited in 2016–2017 and decided on the TG, PD, and EET. The data collected in 2016 and 2017 ($N = 105$) are also used in Balafoutas et al. (2020), which we already referred to in section Literature review. For this reason, it is important to clarify the commonalities and differences between the two studies. Two key features in Balafoutas et al. (2020) were the administration of a priming intervention for part of the sample in a between-subjects design, as well as the distinction between an in-group and an out-group: inmates played each of the three games (TG, PD, EET) once with another inmate (in-group) and once with someone from outside prison (out-group), in a within-subjects design. The priming intervention consisted of a piece of text that inmates were asked to write, reflecting on the time they had spent in prison and on how it had affected their behavior (see Balafoutas et al., 2020, for more details).

In the present study, we pool the data from the priming and the control condition, since one can reasonably expect this intervention to be orthogonal to the relationship between psychopathy and economic behavior, which is the research question here[14]. Regarding the distinction between an in-group and an out-group, in this study we only use data on decisions affecting an inmate's in-group (i.e., other inmates). This is due to two reasons: first, in the 2015 sessions all inmates interact with their in-group only, and therefore we do not have out-group data for the corruption game. Second, our interest in this study lies in the nature of the relationship between psychopathy and behavior, without the additional dimension of group favoritism or bias.

RESULTS

We begin this section by presenting (in **Table 3**) summary statistics for behavior in the four games played by the inmates in our sample. The table reveals strong statistical variation in behavior across participating inmates, thus facilitating the examination of a relationship between behavior and elicited psychopathic traits. Rates of trusting in the TG (*Trust*), cooperating in the PD (*Cooperation*) and taking the bribe-maximizing decision in the CG (*Bribe max*) are all within a 3- to 6 percentage point distance from 50%, while reciprocal choices (*Reciprocity*) are rather frequent at about two-thirds of all choices. In line with most existing studies in experimental economics, behavior in these games is not in line with the Nash equilibrium for selfish subjects. Trust is observed in almost half of the cases, and it is rewarded by second movers in a majority of interactions. Similarly, cooperation rates in the PD lie (at 55%) between the Nash equilibrium of 0% and the social optimum of 100%. In the CG, mean bribes (of 1.75) are between the social optimum of 0 and any of the three pure strategy Nash equilibria, while officials choose the quality-maximizing instead of the bribe-maximizing in just over half of the cases. All of these points toward a considerable degree of pro-social orientation

[11]Instructions in languages other than English were translated from English by native speakers. The experimental instructions can be found in **Supplementary Material 2**.

[12]We note that there was no attrition during a session: all participants completed all parts of the experiment and none left a session before doing so. However, some participants did not fill out all information in the questionnaires, including a few who did not answer all questions in the LSRP, leading to a smaller effective sample size used in the data analysis.

[13]For more information on the sociodemographic characteristics of the inmates recruited during the sessions in 2016 and 2017 please refer to Balafoutas et al. (2020).

[14]This orthogonality assumption is something that we can test: for all regressions presented in the results section (see in particular **Supplementary Tables 1–9** in the Supplementary Material), we have estimated versions in which we add a dummy variable equal to 1 for all inmates in the priming group, as well as interactions between this variable and the LSRP scores. All these terms are insignificant, supporting the validity of pooling the data from the two groups in the analysis. These regressions are not shown in the paper in the interest of brevity but they are available upon request.

TABLE 3 | Summary statistics.

	% / Mean	*N*	St. Dv.
TG			
Trust	44.97%	59	0.50
Reciprocity	65.22%	46	0.48
PD			
Cooperation	55.24%	105	0.49
CG			
Bribe	1.75	48	1.63
Bribe max	47.83%	23	0.51
EET			
Spiteful	20%	21/105	0.40
Inequality averse	13.33%	14/105	0.34
Inequality loving	34.28%	36/105	0.48
Altruistic	32.38%	34/105	0.47
LSRP			
Primary	27.40	153	12.32
Secondary	17.50	158	7.57
Total	44.58	151	18.65

among inmates. Finally, each of the four distributional preference types (*Spiteful, Inequality Averse, Inequality Loving, Altruistic*) accounts for at least 13% and at most 34% of the sample (the exact number of subjects in each type is also shown in **Table 3**).

Turning to an examination of our main research question regarding the relationship between psychopathy and behavior, **Table 4** reports mean values of psychopathy as measured in the LSRP, differentiating between primary, secondary, and total psychopathy and linking it to behavior in each of the four games[15]. In particular, LSRP scores are compared across two sub-groups (yes vs. no) of subjects in each game. For each comparison, the table reports *p*-values from two-tailed *t*-tests.

The first two rows in **Table 4** relate psychopathy to behavior in the trust game. Neither primary nor secondary psychopathy differs significantly between inmates who displayed trusting behavior in this game. Reciprocity, on the other hand, is significantly linked to the LSRP scale: inmates who do not reciprocate trust score higher on primary (and, as a result, on total) psychopathy than those who reciprocate. Similarly, the third row of the table reveals that inmates who take the antisocial action in the prisoner's dilemma (i.e., those who do not cooperate) score significantly higher on both dimensions of psychopathy (primary and secondary) than those who cooperate.

The EET allows us to classify each experimental participant into one of four types of revealed distributional preferences. On aggregate, we find that secondary psychopathy and total psychopathy differ significantly across the four distributional preference types ($p = 0.03$ and $p = 0.05$, respectively; Kruskal-Wallis tests), while the same is not true for primary psychopathy ($p = 0.16$). Turning to each of the four types in isolation, most differences are insignificant. One observation that stands out, however, is that inmates classified as having spiteful preferences have a significantly higher level of LSRP secondary and LSRP total than the other types combined (see row "Spiteful" in **Table 4**).

In the corruption game, as in the trust game, the sample is split between inmates deciding in the role of "public officials" and inmates deciding in the role of "firms." We thus report two behavioral outcomes for this game. The main finding with respect to psychopathy is that public officials who take bribe maximizing decisions—i.e., those who behave antisocially by reducing total welfare—have a significantly higher level of LSRP primary, LSRP secondary and total than those who take quality-maximizing decisions. With respect to the decisions of firms, we split our sample of inmates between those who offer a bribe above vs. below the median and compare LSRP levels across the two. We find no significant differences in any of the LSRP dimensions.

Our results thus show that psychopathy, as elicited in the LSRP, significantly correlates with several behavioral measures in the sample of prison inmates who participated in our experiment. Inmates who do not reciprocate, do not cooperate, who are spiteful and who maximize bribe offers have higher levels of psychopathy than their counterparts, ceteris paribus. For these dimensions, a consistent pattern emerges: inmates with higher scores on the psychopathy scales have a higher tendency toward antisocial behavior[16].

To confirm the robustness of these findings, in **Supplementary Material 1** we also report the results of regressions analyses with trust (**Supplementary Table 1**), reciprocity (**Supplementary Table 2**), bribe maximizing decisions (**Supplementary Table 3**), bribe levels and bribe maximizing behavior in the corruption game (**Supplementary Tables 4**, **5**), and belonging to each of the four distributional preference types (**Supplementary Tables 6**–**9**) as dependent variables. The main independent variables are *LSRP Primary*, *LSRP Secondary*, and *LSRP Total*. In addition to parsimonious specifications that include only psychopathy scores, we estimate (in the Probit regressions for trust, reciprocity and cooperation) specifications that control for a number of inmate characteristics available to us through the prison administration and elicited in the post-experimental surveys. These controls are: time served in prison (*time served*) and total sentence (*total sentence*), in months; a dummy variable (*high security*) equal to one for all inmates in the high security prison; the number of other inmates that someone shares a cell with (*cell share*); education level (coded as 0: none; 1: elementary 2:

[15]We perform a Cronbach's alpha test with LSRP primary, secondary and total leading to a scale reliability coefficient of 0.87.

[16]Given the framing in the CG, an alternative interpretation of the antisocial behavior of inmates is that it reflects their beliefs about the prison personnel that they consider a public official. For instance, if they believe guards are corrupt, they are more likely to engage in bribe-maximizing behavior to express how they believe guards tend to act. In this case egocentric inmates may exhibit this antisocial behavior to show their discomfort in their relationship with guards.

TABLE 4 | Social behavior and psychopathy.

	LSRP Primary		LSRP Secondary		Total	
	Yes	**No**	**Yes**	**No**	**Yes**	**No**
Trust	32.59 (*N* = 17)	35.13 (*N* = 23)	23.16 (*N* = 19)	22.61 (*N* = 23)	56.88 (*N* = 16)	57.64 (*N* = 22)
	p = 0.43		*p = 0.79*		*p = 0.88*	
Reciprocity	33.19 (*N* = 26)	41.15 (*N* = 13)	22.04 (*N* = 28)	23.57 (*N* = 14)	54.64 (*N* = 26)	64.77 (*N* = 13)
	*p = 0.02***		*p = 0.34*		*p = 0.02***	
Cooperation	33.12 (*N* = 49)	37.9 (*N* = 30)	21.67 (*N* = 49)	24.14 (*N* = 35)	54.68 (*N* = 47)	62.37 (*N* = 30)
	*p = 0.04***		*p = 0.05***		*p = 0.02***	
Spiteful	37.92 (*N* = 12)	34.40 (*N* = 67)	25.53 (*N* = 13)	22.18 (*N* = 71)	64.82 (*N* = 11)	56.48 (*N* = 66)
	p = 0.26		*p = 0.05***		*p = 0.07**	
Altruistic	35.21 (*N* = 14)	34.88 (*N* = 65)	22 (*N* = 13)	22.83 (*N* = 71)	56.92 (*N* = 13)	57.83 (*N* = 64)
	p = 0.91		*p = 0.64*		*p = 0.83*	
Inequality Averse	33.14 (*N* = 28)	35.92 (*N* = 51)	22.81 (*N* = 31)	22.64 (*N* = 53)	55.93 (*N* = 28)	58.67 (*N* = 49)
	p = 0.24		*p = 0.90*		*p = 0.42*	
Inequality Loving	35.36 (*N* = 25)	34.74 (*N* = 54)	21.55 (*N* = 27)	23.25 (*N* = 57)	56.88 (*N* = 25)	58.06 (*N* = 52)
	p = 0.80		*p = 0.21*		*p = 0.74*	
Bribe Max	26.82 (*N* = 11)	15.72 (*N* = 12)	13.55 (*N* = 11)	10.04 (*N* = 12)	40.37 (*N* = 11)	25.75 (*N* = 12)
	*p = 0.02***		*p = 0.01****		*p = 0.01****	
Bribe	> mean	< mean	> mean	< mean	> mean	< mean
	20.02 (*N* = 51)	17.86 (*N* = 23)	11.21 (*N* = 51)	12.47 (*N* = 23)	31.23 (*N* = 51)	30.32 (*N* = 23)
	p = 0.35		*p = 0.25*		*p = 0.76*	

All variables defined in text. p-values correspond to t-tests comparing the two binary categories created within each variable.
****, **, *indicate p < 0.01, p < 0.05, and p < 0.1, respectively.*

secondary school; 3: high school; 4: university; 5: master); age, a dummy variable equal to one for married inmates, number of children, and number of siblings[17].

The regression results confirm all main findings obtained so far, both in the parsimonious and in the full specifications. We document a significant relationship between primary psychopathy and reciprocal behavior in the trust game, between both dimensions of psychopathy and cooperation in the prisoner's dilemma, and between both dimensions of psychopathy and bribe maximizing decisions by inmates in the role of public officials in the corruption game. Regarding distributional preference types, **Supplementary Table 6** confirms that higher levels of primary and secondary psychopathy are more likely to be encountered among spiteful types (thereby strengthening the non-parametric test results, which were significant only for secondary psychopathy). In addition, the regression analysis in **Supplementary Table 7** points toward a further negative association between higher levels of (primary) psychopathy and pro-social behavior, measured by the likelihood of being classified as an inequality averse type.

DISCUSSION

Psychopathic personality traits are related to lack of empathy and low inhibition, which would be expected to yield antisocial behavior. In this paper, a population of subjects was recruited among the inmates of two Greek prisons. They were asked to reply to the questions of a self-reported psychopathy scale, the LSRP. They were also faced with four incentivized decision-making experimental tasks which are appropriate to study pro-social (or antisocial) behavior. The four tasks involved decisions affecting oneself and others and were chosen to represent different types of interaction. First, a distributional task, the EET, involved binary dictator-type choices among scenarios regarding own and others' rewards. Second, strategic interaction was involved in a simultaneous (prisoner's dilemma) game involving strategic uncertainty on behalf of both players in each subject pair. Third, in a sequential (trust) game, strategic

[17] In the corruption game these control variables are available only for a small sub-sample of inmates, thus not yielding enough degrees of freedom to estimate the full specifications.

uncertainty was limited to the first mover, while the second player's decision involved no strategic uncertainty. Finally, a more complex, sequential three-player bribery game involved both asymmetric roles and strategic uncertainty. In all these contexts, psychopathy was found to predict antisocial behavior in more or less the expected way, with the exception of active bribers whose psychopathy scores did not predict their bribing behavior. Specifically, higher psychopathy scores relate to lower levels of reciprocity and cooperation, and a higher probability of passive bribery, in the sense of making bribe-maximizing choices.

From a methodological point of view, the robustness of our findings across different economic and game-theoretic experimental tasks can be seen as a confirmation of the validity of the methods used, including the task used for the measurement of subjects' psychopathy, LSRP. Furthermore, the association of psychopathic traits with antisocial behavior is confirmed in a relatively demanding design, in which a broadly used psychometric instrument is shown to reasonably predict behavior in a series of tasks that have the usual abstract framing of context-free decision-making. This framing has interfered in the way others are perceived as (un)trustworthy. Furthermore, the economic decision- making contexts used here have shown further ways of interpreting the difference between primary and secondary psychopathy. The latter is a good predictor of the lack of reciprocity toward people trusting the subject in the first place. Therefore, the results reported here can be seen as an encouraging sign of the benefits from interdisciplinary approaches in order to address the important issue of external and internal validity of the experimental paradigm in both economics and psychology and ultimately document the existence of behavioral spillovers, not only among different economic decision-making tasks, but also across the borders of the two main behavioral sciences.

Regarding the limitations of our study, there are several domains in which our experimental design could be improved or at least complemented by new experiments. First of all, our results come from male prisons. A natural extension would be to check with female institutionalized subjects whether these results are gender-specific. Similarly, a lot could be gained by studying the behavior of inmates in other countries, in order to identify possible prison-specific and country-specific effects. During the experiments we often got the impression that the volunteering inmates accepted to participate in our sessions out of curiosity regarding our "true" objectives. They seemed to hold suspicions regarding our independence from the prison authorities and the anonymity of our protocols. In that sense, highly psychopathic subjects may have adapted their responses in the LSRP test and even their behavior in the experiments in order to project a better self-image in the eyes of the researchers and, supposedly, the prison authority. Future experiments in prison should try to elicit subjects' trust in the researchers' independence and elicit their beliefs on the intentions of the researchers when running similar studies.

A general caveat of experiments with prisoners is that the sample recruited among volunteering inmates will never be comparable with a naturally occurring similar sample extracted from the general population. The span of ages and nationalities contrasts with the unique gender, and the clustering of education on the lowest levels. Relatedly, a limitation of our study is that it documents a relationship between psychopathic traits and inmates' behavior, but it cannot address the question whether psychopathic individuals are more likely to enter prison, or whether psychopathic trends develop inside prison. In general, the extent to which the experience of incarceration shapes behavior is an open and very interesting research question. The small size of the sample is another issue which makes it difficult for the researchers to consider sufficiently many groups in terms of prisoner typologies in order to account for the numerous individual factors which may underlie behavioral differences. Finally, it is difficult -if not impossible- to find a similar, naturally occurring, sample outside the prison to make behavioral comparisons between the inmates and a baseline with non-inmates.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Ethics Committee of the School of Agriculture Policy and Development at the University of Reading. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

NG and EM contributed with funding issues and logistics of the experimental sessions, including obtaining ethics approval and permission to the specific prisons. NG, EM, and TJ-L performed the experiments in the two prisons. LB and TJ-L performed most of the data analysis and produced the manuscript, helped by NG and AG-G. All authors reviewed, edited and approved the final version.

ACKNOWLEDGMENTS

We are thankful to the Hellenic *Ministry of Justice, Transparency and Human Rights*, administration and staff at the high security prison facility Crete 1 and the agricultural prison facility of Agia. We thankfully acknowledge financial support from the *British Academy* (grants SG152916 and SG141101), the Spanish *Ministerio de Economía y Competitividad* (grant ECO2015-68469-R AEI/FEDER), and Spanish *Ministerio de Innovación y Universidades* (grant RTI2018-096927-B-100).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2021.732184/full#supplementary-material

REFERENCES

Akhtar, R., Ahmetoglu, G., and Chamorro-Premuzic, T. (2013). Greed is good? Assessing the relationship between entrepreneurship and subclinical psychopathy. *Pers. Individ. Diff.* 54, 420–425. doi: 10.1016/j.paid.2012.10.013

Babiak, P., Neumann, C. S., and Hare, R. D. (2010). The evolution of terrorism from 1914 to 2014. *Behav. Sci. Law* 28, 211–223. doi: 10.1002/bsl.2124

Balafoutas, L., García-Gallego, A., Georgantzis, N., Jaber-Lopez, T., and Mitrokostas, E. (2020). Rehabilitation and social behavior: experiments in prison. *Games Econ. Behav.* 119, 148–171. doi: 10.1016/j.geb.2019.10.009

Becker, A., Deckers, T., Dohmen, T., Falk, A., and Kosse, F. (2012). The relationship between economic preferences and psychological personality measures. *Ann. Rev. Econ.* 4, 453–78. doi: 10.1146/annurev-economics-080511-110922

Berg, J., Dickhaut, J., and McCabe, K. (1995). Trust, reciprocity, and social history. *Games Econ. Behav.* 10, 122–142. doi: 10.1006/game.1995.1027

Birkeland, S., Cappelen, A. W., Sørensen, E. O., and Tungodden, B. (2014). An experimental study of prosocial motivation among criminals. *Exp. Econ.* 17, 501–511. doi: 10.1007/s10683-013-9380-x

Brand, S., and Price, R. (2020). *The Economic and Social Costs of Crime*. MPRA Paper No. 74968, October, 1–101.

Brandt, J. R., Kennedy, W. A., Patrick, C. J., and Curtin, J. J. (1997). Assessment of psychopathy in a population of incarcerated adolescent offenders. *Psychol. Assess.* 9, 429–435. doi: 10.1037/1040-3590.9.4.429

Brinkley, C. A., Diamond, P. M., Magaletta, P. R., and Heigel, C. P. (2008). Cross-validation of Levenson's psychopathy scale in a sample of federal female inmates. *Assessment* 15, 464–482. doi: 10.1177/1073191108319043

Brinkley, C. A., Schmitt, W. A., Smith, S. S., and Newman, J. P. (2001). Construct validation of a self-report psychopathy scale: does Levenson's self-report psychopathy scale measure the same constructs as Hare's psychopathy checklist-revised? *Pers. Individ. Diff.* 31, 1021–1038. doi: 10.1016/S0191-8869(00)00178-1

Chmura, T., Engel, C., and Englerth, M. (2016). At the mercy of a prisoner three dictator experiments. *Appl. Econ. Lett.* 24, 774–8. doi: 10.1080/13504851.2016.1226486

Clark, B. C., Thorne, C. B., Hendricks, P. S., Sharp, C., Clark, S. K., and Cropsey, K. L. (2015). Individuals in the criminal justice system show differences in cooperative behaviour: implications from cooperative games. *Crim. Behav. Ment. Health* 21, 299–306. doi: 10.1002/cbm.1920

Cleckley, H. (1956). Mask of sanity. *N. Engl. J. Med.* 255:54. doi: 10.1056/NEJM195607052550117

Cohn, A., Maréchal, M. A., and Noll, T. (2015). Bad boys: how criminal identity salience affects rule violation. *Rev. Econ. Stud.* 82, 1289–1308. doi: 10.1093/restud/rdv025

Cox, J. C. (2004). How to identify trust and reciprocity. *Games Econ. Behav.* 46, 260–281. doi: 10.1016/S0899-8256(03)00119-2

Croson, R., and Gneezy, U. (2009). Gender differences in preferences. *J. Econ. Lit.* 47, 448–474. doi: 10.1257/jel.47.2.448

Curry, O., Chesters, M. J., and Viding, E. (2011). The psychopath's dilemma: the effects of psychopathic personality traits in one-shot games. *Pers. Individ. Diff.* 50, 804–809. doi: 10.1016/j.paid.2010.12.036

Gillespie, S. M., Mitchell, I. J., Johnson, I., Dawson, E., and Beech, A. R. (2013). Exaggerated intergroup bias in economical decision making games: differential effects of primary and secondary psychopathic traits. *PLoS ONE* 8:69565. doi: 10.1371/journal.pone.0069565

Guo, S., Liang, P., and Xiao, E. (2020). In-group bias in prisons. *Games Econ. Behav.* 122, 328–340. doi: 10.1016/j.geb.2020.04.015

Hare, R. D., and McPherson, L. M. (1984). Violent and aggressive behavior by criminal psychopaths. *Int. J. Law Psychiatry* 7, 35–50. doi: 10.1016/0160-2527(84)90005-0

Hare, R. D., and Neumann, C. S. (2006). "The PCL-R assessment of psychopathy. Development, structural properties, and new directions," in *Handbook of Psychopathy*. The Guilford Press, 58–88.

Hassall, J., Boduszek, D., and Dhingra, K. (2015). Psychopathic traits of business and psychology students and their relationship to academic success. *Pers. Individ. Diff.* 82, 227–231. doi: 10.1016/j.paid.2015.03.017

Henrichson, C., and Delaney, R. (2012). The price of prisons: what incarceration costs taxpayers. *Fed. Sentencing Rep.* 25, 68–80. doi: 10.1525/fsr.2012.25.1.68

Ibáñez, M. I., Sabater-Grande, G., Barreda-Tarrazona, I., Mezquita, L., López-Ovejero, S., Villa, H., et al. (2016). Take the money and run: psychopathic behavior in the trust game. *Front. Psychol.* 7:1866. doi: 10.3389/fpsyg.2016.01866

Isoni, A. and Sugden R. (2019). Reciprocity and the Paradox of Trust in psychological game theory. *J. Econ. Behav. Organ.* 167, 219–227. doi: 10.1016/j.jebo.2018.04.015

Jaber-López, T., García-Gallego, A., Perakakis, P., and Georgantzis, N. (2014). Physiological and behavioral patterns of corruption. *Front. Behav. Neurosci.* 8:434. doi: 10.3389/fnbeh.2014.00434

Johnson, S. D. (2010). A brief history of the analysis of crime concentration. *Eur. J. Appl. Math.* 21, 349–370. doi: 10.1017/S0956792510000082

Karpman, B. (1948). The myth of the psychopathic personality. *Am. J. Psychiatry* 104, 523–534. doi: 10.1176/ajp.104.9.523

Kerschbamer, R. (2015). The geometry of distributional preferences and a non-parametric identification approach: the equality equivalence test. *Eur. Econ. Rev.* 76, 85–103. doi: 10.1016/j.euroecorev.2015.01.008

Khadjavi, M. (2015). Deterrence works for criminals. *Eur. J. Law Econ.* 31, 1–14. doi: 10.1007/s10657-015-9483-2

Khadjavi, M., and Lange, A. (2013). Prisoners and their dilemma. *J. Econ. Behav. Organ.* 92, 163–175. doi: 10.1016/j.jebo.2013.05.015

Levenson, M. R., Kiehl, K. A., and Fitzpatrick, C. M. (1995). Assessing psychopathic attributes in a noninstitutionalized population. *J. Pers. Soc. Psychol.* 68, 151–158. doi: 10.1037/0022-3514.68.1.151

Lynam, D. R., Whiteside, S., and Jones, S. (1999). Self-reported psychopathy: a validation study. *J. Pers. Assess.* 73, 110–132. doi: 10.1207/S15327752JPA730108

Miller, J. D., Gaughan, E. T., and Pryor, L. R. (2008). The Levenson self-report psychopathy scale: an examination of the personality traits and disorders associated with the LSRP factors. *Assessment* 15, 450–463. doi: 10.1177/1073191108316888

Mokros, A., Menner, B., Eisenbarth, H., Alpers, G. W., Lange, K. W., and Osterheider, M. (2008). Diminished cooperativeness of psychopaths in a prisoner's dilemma game yields higher rewards. *J. Abnorm. Psychol.* 117, 406–413. doi: 10.1037/0021-843X.117.2.406

Montañes, R. F., Taracena, D. L., and Rodríguez, M. (2003). Antisocial personality disorder evaluation with the prisoner's dilemma. *Actas Esp. Psiquiatr.* 31, 307–314.

Nese, A., Higgins, N. O., Sbriglia, P., and Scudiero, M. (2016). Cooperation, punishment and organized crime: a lab in the field experiment in Southern Italy. *Iza Discussion Paper Series* 9901, 1–28.

Porter, S., Birt, A. R., and Boer, D. P. (2001). Investigation of the criminal and conditional release profiles of Canadian federal offenders as a function of psychopathy and age. *Law Hum. Behav.* 25, 647–661. doi: 10.1023/A:1012710424821

Rilling, J. K., Glenn, A. L., Jairam, M. R., Pagnoni, G., Goldsmith, D. R., Elfenbein, H. A., et al. (2007). Neural correlates of social cooperation and non-cooperation as a function of psychopathy. *Biol. Psychiatry* 61, 1260–1271. doi: 10.1016/j.biopsych.2006.07.021

Sabater-Grande, G., and Georgantzis, N. (2002). Accounting for risk aversion in repeated prisoners' dilemma games: an experimental test. *J. Econ. Behav. Organ.* 48, 37–50. doi: 10.1016/S0167-2681(01)00223-2

Sawyer, W., and Wagner, P. (2020). *Mass Incarceration: The Whole Pie 2020*. Available online at: www.Prisonpolicy.Org (accessed February 24, 2020).

Skeem, J., Johansson, P., Andershed, H., Kerr, M., and Louden, J. E. (2007). Two subtypes of psychopathic violent offenders that parallel primary and secondary variants. *J. Abnorm. Psychol.* 116, 395–409. doi: 10.1037/0021-843X.116.2.395

Spitzer, M., Fischbacher, U., Herrnberger, B., Grön, G., and Fehr, E. (2007). The neural signature of social norm compliance. *Neuron* 56, 185–196. doi: 10.1016/j.neuron.2007.09.011

Vaughn, M. G., Edens, J. F., Howard, M. O., and Smith, S. T. (2009). An investigation of primary and secondary psychopathy in a statewide sample of incarcerated youth. *Youth Violence Juv. Justice* 7, 172–188. doi: 10.1177/1541204009333792

Walters, G. D. (2004). The trouble with psychopathy as a general theory of crime. *Int. J. Offender Ther. Comp. Criminol.* 48, 133–148. doi: 10.1177/0306624X03259472

Zolondek, S., Lilienfeld, S. O., Patrick, C. J., and Fowler, K. A. (2006). The interpersonal measure of psychopathy: construct and incremental validity in male prisoners. *Assessment* 13, 470–482. doi: 10.1177/1073191106289861

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychology

ORIGINAL RESEARCH
published: 27 September 2021
doi: 10.3389/fpsyg.2021.729083

The Relationship Between Social Class and Generalized Trust: The Mediating Role of Sense of Control

Ruichao Qiang[1,2,3], Xiang Li[1,2] and Qin Han[4]**

[1] Key Laboratory of Intelligent Education Technology and Application of Zhejiang Province, Zhejiang Normal University, Jinhua, China, [2] Department of Psychology, College of Teacher Education, Zhejiang Normal University, Jinhua, China, [3] Tin Ka Ping Moral Education Research Center, Zhejiang Normal University, Jinhua, China, [4] Department of Psychology, School of Educational Science, Shanxi Normal University, Taiyuan, China

The success and well-being theory of trust holds that higher social class is associated with higher generalized trust, and this association has been well documented in empirical research. However, few studies have examined the processes that might explain this link. This study extends this assumption to explore the mediating mechanism in the association. We hypothesized that social class would positively predict generalized trust, and the relationship would be mediated by people's sense of control. Self-report data were collected from 480 adults (160 males, 320 females; ages 18–61) who participated through an online crowdsourcing platform in China. The results of multiple regression and mediation analyses supported the hypothesized model. This research provides further support for the success and well-being theory of trust, and builds on it by identifying greater sense of control as a possible explanation for the link between high social class and generalized trust. Limitations and possible future research are discussed.

Keywords: generalized trust, sense of control, social class, socioeconomic status, social trust

OPEN ACCESS

Edited by:
Ismael Rodriguez-Lara, University of Granada, Spain

Reviewed by:
Ginés Navarro-Carrillo, University of Jaén, Spain
Simone Di Plinio, University of Studies G. d'Annunzio Chieti and Pescara, Italy

****Correspondence:***
Ruichao Qiang
qiangruichao@126.com
Qin Han
hbiner@126.com

Specialty section:
This article was submitted to Personality and Social Psychology, a section of the journal Frontiers in Psychology

Received: *22 June 2021*
Accepted: *01 September 2021*
Published: *27 September 2021*

Citation:
Qiang R, Li X and Han Q (2021) The Relationship Between Social Class and Generalized Trust: The Mediating Role of Sense of Control. Front. Psychol. 12:729083. doi: 10.3389/fpsyg.2021.729083

INTRODUCTION

Generalized trust is regarded as the core component of social capital and the building block of modern societies (Fukuyama, 1995; Delhey et al., 2011; Freitag and Bauer, 2013; Kim, 2018). It motivates a range of positive societal outcomes, including economic development (Tabellini, 2010), institutional quality (Robbins, 2012), civic engagement (Delhey et al., 2011) and democracy (Paxton, 2002; Zmerli and Newton, 2008). Without generalized trust, social disorder and conflict are commonplace (Putnam, 2000; Robbins, 2016; Jing, 2019).

One consistent correlate of generalized trust is social class or socio-economic status, with the rich and well-educated reporting more generalized trust than their lower social class counterparts (Putnam, 2000; Alesina and La Ferrara, 2002; Hamamura, 2012; Brandt et al., 2015; Navarro-Carrillo et al., 2018b). The link between social class and generalized trust has been postulated for decades. For instance, Simmel (1950) emphasized that there was the discrepancy of resources that are available to different social class to afford the risks of trust. A few prior studies have investigated potential psychological mechanisms (e.g., relative deprivation) in the social class-interpersonal (dis)trust relationship (Yu et al., 2020). However, there has been a little empirical research regarding the explanation of why social class and generalized trust are correlated. In the current research, we fill this gap by testing the role that sense of control may play in explaining this association.

Generalized trust refers to one's belief that most people can be trusted (Yamagishi and Yamagishi, 1994; Uslaner, 2002; Freitag and Traunmuller, 2009; Navarro-Carrillo et al., 2018b). People with

high generalized trust hold a general belief in human benevolence and they believe that the trustee has benign intentions in social interactions (Yamagishi and Yamagishi, 1994). As a result, they tend to trust strangers, passersby on the street, and other people whom they do not know well. Although generalized trust exposes people to the risk that the target of trust has harmful intentions, this risk may be outweighed by the benefits of trusting strangers. One of the benefits of generalized trust is its promotion on interactions among unfamiliar individuals. Interactions with unfamiliar people expose individuals to novel information and resources that are not available in acquainted relationships (Hamamura, 2012).

Previous research suggested that the risks and benefits of generalized trust are balanced differently across people from different groups, including different social classes (Hamamura, 2012; Brandt et al., 2015). We hold that a sense of personal control may contribute to people's perception of these risks and benefits of trust. Members of the lower social class are likely to have a lower sense of control, and thus a lower trust to other people. In the following sections we review the literature on the direct relationship between social class and generalized trust, and the literature relevant to our proposal that sense of control may mediate this link.

Direct Relationship Between Social Class and Generalized Trust

Social class is typically conceptualized as a reflection of multiple features of social life (Fiske and Markus, 2012; Kraus and Keltner, 2013; Daganzo and Bernardo, 2018). Social class is a context rooted in both the resources of social life (e.g., wealth, education, occupation) and the individual's perceived rank within the social hierarchy (Kraus et al., 2009, 2012). Traditionally, researchers measure social class in terms of objective indicators such as the individual's level of education, income, and occupation prestige (Kraus et al., 2009, 2012; Daganzo and Bernardo, 2018).

However, there are several inherent problems in assessing social class with objective variables. For instance, it is uncertain how objective indicators (e.g., education, income) combine to yield a composite score representing social class (Kraus et al., 2009; Oakes and Rossi, 2003). As a result, many researchers have questioned the validity of objective metrics of social class in capturing the essence of class. Moreover, many research suggested that subjective measures of social class, compared with the objective measures, more strongly predict the psychological outcomes and serve as a more consistent predictor of social explanation (Adler et al., 2000; Kraus et al., 2011, 2012). Thus, in this study we refer to social class using subjective measures.

According to the success and well-being theory of trust (Delhey and Newton, 2003), generalized trust is more likely to be expressed by people from the upper class than people from the lower class (Alesina and La Ferrara, 2002; Gheorghiu et al., 2009; Hamamura, 2012; Brandt et al., 2015; Navarro-Carrillo et al., 2018b). Trust always carries risks, and it is more risky for lower class individuals (Hamamura, 2012; Navarro-Carrillo et al., 2018b). Lower status individuals who commonly face resource scarcity cannot afford to lose even a little if their trust is betrayed. In contrast, upper class individuals have abundant properties to protect against the risks and vulnerabilities of trust (Brandt et al., 2015), and they can gain more benefits from trust (Delhey and Newton, 2003; Hamamura, 2012).

Moreover, from this perspective, social trust is the product of adult life experiences. Upper class people have been treated with more respect and kindness. Consequently, they are more trusting than lower class individuals who always suffer discrimination and social exclusion (Putnam, 2000). This theory is supported to some degree by survey data provided by the American General Social Survey (Alesina and La Ferrara, 2002) and the German Socio-Economic Panel (Korndörfer et al., 2015). These studies suggest that social class is consistently and positively related to generalized trust.

Sense of Control as a Potential Mediator

Although previous studies have indicated that social class has enduring association with generalized trust (Hamamura, 2012; Brandt et al., 2015; Navarro-Carrillo et al., 2018b), the specific mechanism involved in this association has been rarely examined. One reason that social class is linked with generalized trust may be that members of different social classes differ in their sense of control. Several studies have documented a disparity in sense of control felt by members of the upper and lower social classes, with upper class individuals typically reporting greater perceived control over their life (Lachman and Weaver, 1998; Kraus et al., 2009; Daganzo and Bernardo, 2018).

Sense of control or self-agency has been described as the experience of being the source of one' s own actions and their consequences (Dewey et al., 2010; Di Plinio et al., 2020). From an event-control approach (Jordan, 2003), one's sense of agency depends partly on contextual information about the degree of control an individual has over the environment (Dewey et al., 2010; Kumar and Srinivasan, 2012; Di Plinio et al., 2019). From the social cognition perspective on social class, social class contexts elicit a coherent set of social cognitive patterns of thought, feeling, and action with regard to oneself and other people (Kraus et al., 2012). Specially, people from upper class inhabit an environment with abundant resources, personal freedom, and social opportunities. This makes them perceive a greater sense of personal control over life (Kraus et al., 2009, 2012). Furthermore, upper class individuals are more likely to occupy positions of influence and elevated status, which strongly promote their perceived personal control (Kraus et al., 2012). In contrast, the social contexts of lower class are characterized by reduced resources, external threats, and vulnerability, which may make them feel powerless to exert control over their lives (Haushofer and Fehr, 2014; Piff, 2014).

Furtherly, the social class difference in sense of control may lead to the disparity in generalized trust (Navarro-Carrillo et al., 2018a; Samson and Zaleskiewicz, 2020). Uslaner (2002), for instance, views individuals' sense of control over their life as key to understanding their trust in people. Generalized trust always carries risks due to the possible betrayal by others (Delhey and Newton, 2003; Hamamura, 2012; Navarro-Carrillo et al., 2018b). Individuals with greater sense of control can afford to maintain an optimistic view of other people and be more

trusting in general (Samson and Zaleskiewicz, 2020). In contrast, individuals with lower perceived personal control over life are psychologically defensive and prefer to distrust others (Brandt and Henry, 2012; Samson and Zaleskiewicz, 2020). It makes sense to think that people lack of perceived control express diminished generalized trust.

CURRENT STUDY

The main purpose of the present research was to examine the relationships among social class, generalized trust and sense of control. We expected to find further evidence of the social class difference in generalized trust, as it has been documented in many other studies. More importantly, we tested a model in which sense of control mediated the relationship between social class and generalized trust. The four hypotheses below were derived from the theoretical assumptions and empirical evidence presented above.

H1 Social class will significantly and positively predict generalized trust.

H2 Social class will be positively related to sense of control.

H3 Sense of control will be positively associated with generalized trust.

H4 Social class will be associated with higher generalized trust through heightened sense of control.

MATERIALS AND METHODS

Participants

An online crowdsourcing platform in mainland China, which provides functions equivalent to Amazon Mechanical Turk, recruited 494 Chinese participants. Of these, 14 reported their age to be below 18. These participants were excluded from the following analyses, leaving a final sample of 480 individuals (160 males, 320 females). The age of participants ranged from 18 to 61 years of age (M = 27.77, SD = 8.21).

A sensitivity power analysis using G*Power (Faul et al., 2007) indicated that, the minimum effect size required to produce power at the 0.80 level in linear multiple regression with current sample size was 0.016. The effect size of regression coefficients in our study were all greater than it.

Procedure

Participants were instructed that they would participate in an online survey about their social attitudes. They were informed that their answers would be anonymous and that they could stop participating at any time. They signed an informed consent form prior to participating in the online surveys. Then, they filled out measures of social class, sense of control, and generalized trust. The participants also provided their gender and age. It took about four min to complete all the scales. If participants skipped an item, they were reminded to complete it when they clicked the submit button. The survey could not be submitted until all items were completed. This provided a data set with no missing values. The participants were thanked for participating in the study but received no other reward.

Measures

Social Class

We assessed social class using the MacArthur Ladder Scale (Adler et al., 2000). Participants were shown a picture of a 10-rung ladder and asked to imagine that the ladder represented where people stand in society. They were told that at the bottom (social class = 0) are the people who are the worst off—who have the least education, the least money, and the least respected jobs or no jobs; at the top of the ladder (social class = 10) are the people who are the best off—those who have the most education, the most money, and the most respected jobs. Then, they were asked to indicate their position at the ladder at this time of their life relative to other people in society (M = 4.55, SD = 1.69).

Sense of Control

Sense of control was assessed using the established measure from Lachman and Weaver (1998). The Sense of Control Scale is composed of 12 items—4 measuring personal mastery and 8 measuring perceived constraints. Sample items are: "*I can do just about anything that I really set my mind to*" and "*When I really want to do something, I usually find a way to succeed at it*" from the personal mastery dimension, and "*Other people determine most of what I can and cannot do*" and "*There is little I can do to change many of the important things in my life*" from the perceived constraints dimension. Items were rated on a 7-point Likert scale (1 = *strongly disagree,* 7 = *strongly agree*). The items belonging to perceived constraints dimension were reverse-scored, then all items were averaged to obtain a composite score for sense of control (α = 0.82). Confirmatory Factor Analysis showed that the scale had high construct validity in this study (CFI = 0.95, TLI = 0.94, $RMSEA$ = 0.06, 90% CI [0.04, 0.07], $SRMR$ = 0.05).

Generalized Trust

Generalized trust was assessed using an established three-item measure (Chen et al., 2011). To assess generalized trust, participants indicated their agreement with three statements. The first item is the classic binary trust question from the World Value Survey: "*Generally speaking, would you say that most people can be trusted or that you need to be very careful in dealing with people?*" Responses were coded as 1 = *need to be very careful,* 2 = *don't know,* 3 = *most people can be trusted.* The second item is "*Do you think that most people will take advantage of your weakness or that they will do you justice?*" with responses coded as 1 = *take advantage of me,* 2 = *a 50–50 chance,* 3 = *do me justice*. The third item is "*No matter known or not, most people are trustworthy.*" Responses were coded as 1 = *they aren't trustworthy,* 2 = *a 50–50 chance,* 3 = *they are trustworthy*. Scores on the three items were averaged to form the generalized trust scale. The Cronbach's alpha coefficient in this study was 0.62.

TABLE 1 | Relationships among social class, sense of control, and generalized trust.

	Class-trust relationship	Class-control relationship	Control-trust relationship
Without control variables	β = 0.23, $F(1, 478) = 26.10$	β = 0.33, $F(1, 478) = 56.31$	β = 0.30, $F(1, 478) = 47.05$
With control variables	β = 0.23, $F(2, 476) = 8.91$	β = 0.33, $F(2, 476) = 20.47$	β = 0.30, $F(2, 476) = 15.66$

Note: All effects are significant at $p < 0.001$. The control variables were gender and age.

RESULTS

Prior to the main analyses, we conducted a preliminary analysis among variables.[1] Correlations between primary variables of interest and demographic variables were all not significant. However, prior research has controlled for gender and age when analyzing the contribution of social class to social trust (Hamamura, 2012; Brandt et al., 2015; Kim et al., 2021). Therefore, and as they are common sociodemographic variables, we consider their effects on the hypothesized associations.

Then, we performed regression analyses predicting the links among social class, sense of control and generalized trust. We included gender and age as control variables to determine whether the associations held beyond the effects of demographic variables. The results of these analyses are summarized in **Table 1**.

Supporting H1, social class was significantly associated with generalized trust, $R^2 = 0.052$, $F(1, 478) = 26.10$, β = 0.23, $p < 0.001$, 95% *CI* [0.140, 0.315]. This relationship remained significant when controlling for gender and age, $R^2 = 0.053$, $F(2, 476) = 8.91$, β = 0.23, $p < 0.001$, 95% *CI* [0.140, 0.316]. Supporting H2, social class predicted greater sense of control, $R^2 = 0.105$, $F(1, 478) = 56.31$, β = 0.33, $p < 0.001$, 95% CI [0.240, 0.410]. This link remained significant when controlling age and gender, $R^2 = 0.114$, $F(2, 476) = 20.47$, β = 0.33, $p < 0.001$, 95% CI [0.241, 0.410]. H3 was also confirmed. Sense of control predicted significantly greater generalized trust, $R^2 = 0.09$, $F(1, 478) = 47.05$, β = 0.30, $p < 0.001$, 95% CI [0.214, 0.385]. This result was still significant after controlling for gender and age, $R^2 = 0.09$, $F(2, 476) = 15.66$, β = 0.30, $p < 0.001$, 95% CI [0.213, 0.385].

To determine whether sense of control acted as a mediator between social class and generalized trust, we tested the hypothesized mediation model in Amos 27. Structural equation modeling indicated that the hypothesized model (**Figure 1**) showed a good fit with the data (*CFI* = 0.86, *TLI* = 0.83, *RMSEA* = 0.07, 90% *CI* [0.06, 0.08]). We used a bootstrapping technique with 5,000 iterations to estimate the indirect effect of social class on generalized trust through perceived control. The size of the indirect effect was estimated by examining the 95% bootstrap confidence interval (*CI*) of the estimate; the effect is considered significant when the *CI* does not include zero. Supporting H4, the indirect effect was significant; that is, higher social class was associated with higher generalized trust via a process of greater sense of control (β = 0.10, *SE* = 0.03, $p < 0.001$, bias-corrected 95% *CI* [0.05, 0.18]). As can be seen in **Figure 1**, the direct effect of social class on generalized trust remained significant (β = 0.20, *SE* = 0.06, $p = 0.003$, bias-corrected 95% *CI* [0.07, 0.31]) after including the mediation component, suggesting partial mediation.

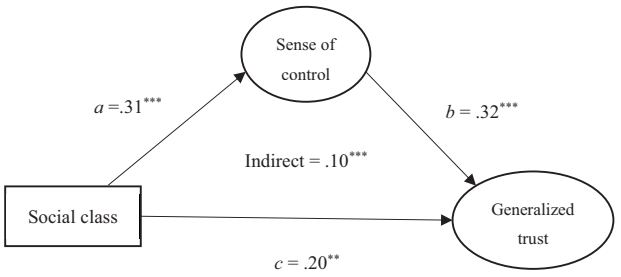

FIGURE 1 | Mediating model illustrating the relationships among social class, sense of control, and generalized trust. Paths a, b, and c represent direct effects. All numbers are standardized regression coefficients. $^{**}p < 0.01$, $^{***}p < 0.001$. The 95% confidence interval of the bias-corrected indicates a significant indirect effect.

DISCUSSION

It is well documented that members of the upper social class show more generalized trust than members of the lower social class, but a little research has examined the reason for this association. The present study tested whether sense of personal control plays a mediating role in the association between social class and social trust. We found evidence that supported four key hypotheses derived from the success and well-being theory of trust (Delhey and Newton, 2003). This study represents the first empirical demonstration of a mediator of sense of control between social class and generalized trust, and the new evidence that this is a process through which social cognition effect of social class (Kraus et al., 2012) can operate.

This study enriches the growing body of research on social class and trust. As expected, social class significantly and positively predicted generalized trust. This finding is consistent with the success and well-being theory of trust that asserts a positive association between social class and generalized trust (Delhey and Newton, 2003; Brandt et al., 2015; Edelman, 2017). The present study supports this long-held view and adds new evidence that helps explain why higher social class is associated with greater generalized trust.

The results showed that perceived control may be a mediating psychological mechanism in the association between social class and trust beliefs. As the social cognition perspective on social class suggests (Kraus et al., 2012), social class contexts elicit a coherent set of social cognitive patterns of thought, including the perception of personal control. Specifically, the upper social

[1]In our analyses, gender was coded as a dummy variable, male = 1, female = 2. All the variables were standardized before the analyses.

class context generates a stronger sense of control than the lower social class context (Kraus et al., 2009; Daganzo and Bernardo, 2018). This is consistent with the event-control approach, which asserts that context information can modulate individual's sense of control (Jordan, 2003; Di Plinio et al., 2019). Furthermore, several researchers have highlighted generalized trust as a direct consequence of sense of personal control (Uslaner, 2002; Navarro-Carrillo et al., 2018a). Samson and Zaleskiewicz (2020) declared that people who have a strong sense of control over one's own life may be more likely to maintain an optimistic view of other people and to be more trusting in general. Sense of control thus is a psychological mechanism that links social class to trust and a helpful focus of intervention for people of lower class who struggle with trusting others.

The current study not only supports the success and well-being theory of trust, but extends the theory by revealing that the cognitive factors work when social class may serve to structure social psychological functioning. Furthermore, a new question is raised and might need to be incorporated into the success and well-being theory of trust. That is, whether emotional factors act in the function process of social class, given that some negative emotion such as insecurity and anxiety are the powerful attenuators of trust (Patterson, 1999; Nguyen, 2017; Navarro-Carrillo et al., 2018a).

The present work is not without limitations. First, the cross-sectional design limits the causal conclusions that can be drawn from the data. Given that social class involves a long-term experience and is probably more stable than generalized trust, it may be that social class influences generalized trust in the association. However, it is also possible that a third variable affects both of them. For example, social class of parents may partially determine the social class of children and influence the trust belief of children through the socialization of social cognition in the family.

Secondly, the internal reliability of the generalized trust scale was acceptable but not high. This is a common weakness of short scales (Tavakol and Dennick, 2011). Cronbach's alpha, a commonly used measure of internal consistency, is affected by the length of the scale. If the scale length is too short, the value of alpha is reduced (Streiner, 2003; Tavakol and Dennick, 2011). However, this measure effectively exhibited social class tendency of generalized trust in our sample, and a similar measure has been used in other research on social class and generalized trust (Brandt et al., 2015). Nevertheless, a longer scale may be more useful in future research, such as the General Trust Scale (Yamagishi and Yamagishi, 1994) which has 6 items and showed high internal consistency (alpha values range from 0.71 to 0.74) in other studies (Navarro-Carrillo et al., 2018a,b).

Thirdly, we used only subjective measures to assess social class, and different results may be obtained using objective measures. Subjective measures have been found to be more potent predictors of psychological outcomes than the objective measures (Adler et al., 2000; Kraus et al., 2011, 2012). However, in some cases, objective measures of social class work better in predicting social explanations (Kim et al., 2021). As a consequence, including both subjective and objective measures of social class would allow a test of which aspects of social class are most predictive of generalized trust. It might also help in inspecting the inter-relations between objective vs. subjective social class.

Fourthly, the role of psychological defensiveness playing in the association between social class and generalized trust should be further explored. People from the lower class face long-term prejudice and psychological threats to the self, which make them psychologically defensive against these self-threats (Henry, 2009; Brandt et al., 2015). A manifestation of psychological defensiveness is in terms of distrust in other people (Brandt and Henry, 2012). Psychological defensiveness may be an individual difference that would explain some of the variability in generalized trust among people of the lower social class.

Lastly, cultural issues should be considered in interpreting the results. Culture exists as a socially shared reality that generates values, beliefs, and social interaction norms in social life (Barker, 2017). For instance, some cultures value interdependence and benevolence while others value independence and competitiveness. Consequently, social trust is generally affected by cultural elements (Gheorghiu et al., 2009; Berigan and Irwin, 2011; Steel et al., 2018). This may lead to diverse baselines of social trust across different cultures and, in turn, may impact the association between social class and trust. This effect may function through some general cultural factors (e.g., individualism vs. collectivism, multiculturalism, politics regarding immigration) or through attention to emotional and cognitive stimuli (Grossmann et al., 2012).

CONCLUSION

There is a well-documented link between social class and generalized trust (Hamamura, 2012; Brandt et al., 2015). However, a little research has examined the reason for this link. The key contribution of the present study is the finding that sense of control acts as a mediator between social class and generalized trust. Members of the upper social class were inclined to perceive high control over their outcomes, and they held a strong generalized trust in daily life. In contrast, members of the lower social class were more likely to feel a low sense of control, and in turn, low social trust.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The study involving human participants were reviewed and approved by the Department of psychology of the Zhejiang Normal University. Written informed consent to participate in the study was provided by the participants or where

applicable, the participants legal guardian/next of kin. The patients/participants provided their written informed consent to participate in this study. Written informed consent was obtained from the individual (s) for the publication of any potentially identifiable images or data included in this article.

AUTHOR CONTRIBUTIONS

RQ conceived the study, analyzed relevant literature, and wrote the manuscript in all its sections. XL collected the data, conducted the statistical analysis of the data, wrote the current version of the results section. QH conceived the study, analyzed relevant literature, structured the questionnaire, and wrote the results and conclusion sections. All authors contributed to the article and approved the submitted version.

FUNDING

This research was supported by the Open Research Fund of College of Teacher Education, Zhejiang Normal University (No. JYKF20067) and a project from the Education Department of Shanxi Province (No. J2017051).

ACKNOWLEDGMENTS

We thank the two reviewers who helped us improve the manuscript.

REFERENCES

Adler, N. E., Epel, E. S., Castellazzo, G., and Ickovics, J. R. (2000). Relationship of subjective and objective social status with psychological and physiological functioning: preliminary data in healthy white women. *Health Psychol.* 19, 586–592. doi: 10.1037/0278-6133.19.6.586

Alesina, A., and La Ferrara, E. (2002). Who trusts others? *J. Public Econ.* 85, 207–234. doi: 10.1016/S0047-2727(01)00084-6

Barker, G. G. (2017). Acculturation and bicultural integration in organizations: conditions, contexts, and challenges. *Int. J. Cross Cult. Manag.* 17, 281–304. doi: 10.1177/1470595817712741

Berigan, N., and Irwin, K. (2011). Culture, cooperation, and the general welfare. *Soc. Psychol. Q.* 74, 341–360. doi: 10.1177/0190272511422451

Brandt, M. J., and Henry, P. J. (2012). Psychological defensiveness as a mechanism explaining the relationship between low socioeconomic status and religiosity. *Int. J. Psychol. Relig.* 22, 321–332. doi: 10.1080/10508619.2011.646565

Brandt, M. J., Wetherell, G., and Henry, P. J. (2015). Changes in income predict change in social trust: a longitudinal analysis. *Polit. Psychol.* 36, 761–768. doi: 10.1111/pops.12228

Chen, J., Huhe, N., and Lu, C. (2011). Causal mechanisms between social trust and community governance. *Chin. J. Sociol.* 31, 22–40. doi: 10.15992/j.cnki.31-1123/c.2011.06.003

Daganzo, M. A. A., and Bernardo, A. B. (2018). Socioeconomic status and problem attributions: the mediating role of sense of control. *Cogent Psychol.* 5:1525149. doi: 10.1080/23311908.2018.1525149

Delhey, J., and Newton, K. (2003). Who trusts? The origins of social trust in seven nations. *Eur. Soc.* 5, 93–137. doi: 10.1080/1461669032000072256

Delhey, J., Newton, K., and Welzel, C. (2011). How general is trust in 'most people'? Solving the radius of trust problem. *Am. Sociol. Rev.* 76, 786–807. doi: 10.1177/0003122411420817

Dewey, J. A., Seiffert, A. E., and Carr, T. H. (2010). Taking credit for success: the phenomenology of control in a goal directed task. *Conscious. Cogn.* 19, 48–62. doi: 10.1016/j.concog.2009.09.007

Di Plinio, S., Arno, S., Perrucci, M. G., and Ebisch, S. J. (2019). Environmental control and psychosis-relevant traits modulate the prospective sense of agency in non-clinical individuals. *Conscious. Cogn.* 73:102776. doi: 10.1016/j.concog.2019.102776

Di Plinio, S., Perrucci, M. G., Aleman, A., and Ebisch, S. J. (2020). I am Me: brain systems integrate and segregate to establish a multidimensional sense of self. *Neuroimage* 205:116284. doi: 10.1016/j.neuroimage.2019.116284

Edelman (2017). *2017 Edelman Trust Barometer-Global report*. Available online at: https://www.slideshare.net/EdelmanInsights/2017-edelman-trust-barometer-global-results-71035413 (accessed March 27, 2019.

Faul, F., Erdfelder, E., Lang, A.-G., and Buchner, A. (2007). G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav. Res. Methods* 39, 175–191. doi: 10.3758/BF03193146

Fiske, S. T., and Markus, H. R. (eds) (2012). *Facing Social Class: How Societal Rank Influences Interaction*. New York, NY: Russell Sage Foundation.

Freitag, M., and Bauer, P. C. (2013). Testing for measurement equivalence in surveys. *Public Opin. Q.* 77, 24–44. doi: 10.1093/poq/nfs064

Freitag, M., and Traunmuller, R. (2009). Spheres of trust: an empirical analysis of the foundations of particularised and generalised trust. *Eur. J. Polit. Res.* 48, 782–803. doi: 10.1111/j.1475-6765.2009.00849.x

Fukuyama, F. (1995). *Trust: The Social Virtues and the Creation of Prosperity*. New York, NY: Free Press.

Gheorghiu, M., Vignoles, V., and Smith, P. (2009). Beyond the United States and Japan: testing Yamagishi's emancipation theory of trust across 31 nations. *Soc. Psychol. Q.* 72, 365–383. doi: 10.1177/019027250907200408

Grossmann, I., Ellsworth, P. C., and Hong, Y. Y. (2012). Culture, attention, and emotion. *J. Exp. Psychol. Gen.* 141, 31–36. doi: 10.1037/a0023817

Hamamura, T. (2012). Social class predicts generalized trust but only in wealthy societies. *J. Cross Cult. Psychol.* 43, 498–509. doi: 10.1177/0022022111399649

Haushofer, J., and Fehr, E. (2014). On the psychology of poverty. *Science* 344, 862–867. doi: 10.1126/science.1232491

Henry, P. J. (2009). Low-status compensation: a theory for understanding the role of status in cultures of honor. *J. Pers. Soc. Psychol.* 97, 451–466. doi: 10.1037/a0015476

Jing, S. (2019). "Conceptualizing and measuring sense of social trust," in *Social Mentality in Contemporary China*, ed. Y. Yang (Singapore: Springer), 87–109. doi: 10.1007/978-981-13-7812-6_7

Jordan, J. S. (2003). Emergence of self and other in perception and action: an event-control approach. *Conscious. Cogn.* 12, 633–646. doi: 10.1016/S1053-8100(03)00075-8

Kim, H. S. (2018). Particularized trust, generalized trust, and immigrant self-rated health: cross-national analysis of World Values Survey. *Public Health* 158, 93–101. doi: 10.1016/j.puhe.2018.01.039

Kim, Y., Sommet, N., Na, J., and Spini, D. (2021). Social class—not income inequality—predicts social and institutional trust. *Soc. Psychol. Personal. Sci.* doi: 10.1177/1948550621999272 [Epub ahead of print].

Korndörfer, M., Egloff, B., and Schmukle, S. C. (2015). A large scale test of the effect of social class on prosocial behavior. *PLoS One* 10:e0133193. doi: 10.1371/journal.pone.0133193

Kraus, M. W., Horberg, E. J., Goetz, J. L., and Keltner, D. (2011). Social class rank, threat vigilance, and hostile reactivity. *Pers. Soc. Psychol. Bull.* 37, 1376–1388. doi: 10.1177/0146167211410987

Kraus, M. W., and Keltner, D. (2013). Social class rank, essentialism, and punitive judgment. *J. Pers. Soc. Psychol.* 105, 247–261. doi: 10.1037/a0032895

Kraus, M. W., Piff, P. K., and Keltner, D. (2009). Social class, sense of control, and social explanation. *J. Pers. Soc. Psychol.* 97, 992–1004. doi: 10.1037/a0016357

Kraus, M. W., Piff, P. K., Mendoza-Denton, R., Rheinschmidt, M. L., and Keltner, D. (2012). Social class, solipsism, and contextualism: how the rich are different from the poor. *Psychol. Rev.* 119, 546–572. doi: 10.1037/a0028756

Kumar, D., and Srinivasan, N. (2012). Hierarchical event-control and subjective experience of agency. *Front. Psychol.* 3:410. doi: 10.3389/fpsyg.2012.00410

Lachman, M. E., and Weaver, S. L. (1998). The sense of control as a moderator of social class differences in health and well-being. *J. Pers. Soc. Psychol.* 74, 763–773. doi: 10.1037/0022-3514.74.3.763

Navarro-Carrillo, G., Valor-Segura, I., Lozano, L. M., and Moya, M. (2018a). Do economic crises always undermine trust in others? The case of generalized, interpersonal, and in-group trust. *Front. Psychol.* 9:1955. doi: 10.3389/fpsyg.2018.01955

Navarro-Carrillo, G., Valor-Segura, I., and Moya, M. (2018b). Do you trust strangers, close acquaintances, and members of your ingroup? Differences in trust based on social class in Spain. *Soc. Indic. Res.* 135, 585–597. doi: 10.1007/s11205-016-1527-7

Nguyen, C. (2017). Labour market insecurity and generalized trust in welfare state context. *Eur. Sociol. Rev.* 33, 225–239. doi: 10.1093/esr/jcw058

Oakes, J. M., and Rossi, R. H. (2003). The measurement of SES in health research: current practice and steps toward a new approach. *Soc. Sci. Med.* 56, 769–784. doi: 10.1016/S0277-9536(02)00073-4

Patterson, O. (1999). "Liberty against the democratic state: on the historical and contemporary sources of American distrust," in *Democracy and Trust*, ed. M. E. Warren (Cambridge: Cambridge University Press), 151–207. doi: 10.1017/CBO9780511659959.006

Paxton, P. (2002). Social capital and democracy: an independent relationship. *Am. Sociol. Rev.* 67, 254–277. doi: 10.2307/3088895

Piff, P. K. (2014). Wealth and the inflated self: class, entitlement, and narcissism. *Pers. Soc. Psychol. Bull.* 40, 34–43. doi: 10.1177/0146167213501699

Putnam, R. D. (2000). *Bowling Alone: The Collapse and Revival of American community*. New York, NY: Simon and Schuster.

Robbins, B. G. (2012). Institutional quality and generalized trust: a nonrecursive causal model. *Soc. Indic. Res.* 107, 235–258. doi: 10.1007/s11205-011-9838-1

Robbins, B. G. (2016). From the general to the specific: how social trust motivates relational trust. *Soc. Sci. Res.* 55, 16–30. doi: 10.1016/j.ssresearch.2015.09.004

Samson, K., and Zaleskiewicz, T. (2020). Social class and interpersonal trust: Partner's warmth, external threats and interpretations of trust betrayal. *Eur. J. Soc. Psychol.* 50, 634–645. doi: 10.1002/EJSP.2648

Simmel, G. (1950). *The Sociology of Georg Simmel*. Glencoe, IL: Free Press.

Steel, P., Taras, V., Uggerslev, K., and Bosco, F. (2018). The happy culture: a theoretical, meta-analytic, and empirical review of the relationship between culture and wealth and subjective well-being. *Pers. Soc. Psychol. Rev.* 22, 128–169. doi: 10.1177/1088868317721372

Streiner, D. (2003). Starting at the beginning: an introduction to coefficient alpha and internal consistency. *J. Pers. Assess.* 80, 99–103. doi: 10.1207/S15327752JPA8001_18

Tabellini, G. (2010). Culture and institutions: economic development in the regions of Europe. *J. Eur. Econ. Assoc.* 8, 677–716. doi: 10.1111/j.1542-4774.2010.tb00537.x

Tavakol, M., and Dennick, R. (2011). Making sense of Cronbach's alpha. *Int. J. Med. Educ.* 2, 53–55. doi: 10.5116/ijme.4dfb.8dfd

Uslaner, E. M. (2002). *The Moral Foundations of Trust*. Cambridge: Cambridge University Press.

Yamagishi, T., and Yamagishi, M. (1994). Trust and commitment in the United States and Japan. *Motiv. Emot.* 18, 129–166. doi: 10.1007/BF02249397

Yu, G., Zhao, F., Wang, H., and Li, S. (2020). Subjective social class and distrust among Chinese college students: the mediating roles of relative deprivation and belief in a just world. *Curr. Psychol.* 39, 2221–2230. doi: 10.1007/s12144-018-9908-5

Zmerli, S., and Newton, K. (2008). Social trust and attitudes toward democracy. *Public Opin. Q.* 72, 706–724. doi: 10.1093/poq/nfn054

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Psychology

ORIGINAL RESEARCH
published: 28 September 2021
doi: 10.3389/fpsyg.2021.727550

Predicting Trustworthiness Across Cultures: An Experiment

Adam Zylbersztejn[1,2]*, *Zakaria Babutsidze*[3,4] *and Nobuyuki Hanaki*[5]

[1] Univ Lyon 2, Université Lumière Lyon 2, GATE L-SE UMR 5824, Lyon, France, [2] Vistula University Warsaw (AFiBV), Warsaw, Poland, [3] SKEMA Business School, Université Côte d'Azur (GREDEG), Valbonne, France, [4] Observatoire Français des Conjonctures Economiques (OFCE), Sciences Po, Paris, France, [5] Institute of Social and Economic Research, Osaka University, Osaka, Japan

We contribute to the ongoing debate in the psychological literature on the role of "thin slices" of observable information in predicting others' social behavior, and its generalizability to cross-cultural interactions. We experimentally assess the degree to which subjects, drawn from culturally different populations (France and Japan), are able to predict strangers' trustworthiness based on a set of visual stimuli (mugshot pictures, neutral videos, loaded videos, all recorded in an additional French sample) under varying cultural distance to the target agent in the recording. Our main finding is that cultural distance is not detrimental for predicting trustworthiness in strangers, but that it may affect the perception of different components of communication in social interactions.

Keywords: trustworthiness, communication, hidden action game, cross-cultural comparison, laboratory experiment

Edited by:
Tarek Jaber-Lopez, Université Paris Nanterre, France

Reviewed by:
Jan B. Engelmann, University of Amsterdam, Netherlands
Noemí Herranz-Zarzoso, University of Jaume I, Spain

***Correspondence:**
Adam Zylbersztejn
zylbersztejn@gate.cnrs.fr

Specialty section:
This article was submitted to Personality and Social Psychology, a section of the journal Frontiers in Psychology

Received: *18 June 2020*
Accepted: *30 August 2021*
Published: *28 September 2021*

Citation:
Zylbersztejn A, Babutsidze Z and Hanaki N (2021) Predicting Trustworthiness Across Cultures: An Experiment. Front. Psychol. 12:727550. doi: 10.3389/fpsyg.2021.727550

1. INTRODUCTION

A common pattern in human strategic behavior is conditional cooperation, i.e., the willingness to sacrifice personal resources for the mutual benefit as long as others do the same (Fischbacher et al., 2001; Kocher et al., 2008). The extent to which individuals follow the notion of conditional cooperation determines their trustworthiness in social interactions that require mutual cooperation or involve economic exchange (Boone and Buck, 2003). Notwithstanding the standard economic prediction that communication in such contexts should be "cheap talk" and considered as irrelevant for final decisions (Farrell and Rabin, 1996), but in line with the "mind reading" hypothesis that communication may help uncover the motivational states of others (Sally, 2000), experimental evidence suggests that communication helps detect trustworthiness. Communication can thus contribute to creating successful partnerships, and help protect against potential exploitation (He et al., 2017).

Clearly, the verbal content of communication may provide valid signals for the receiver about the sender's intentions. A well-established finding is that making a voluntary promise (i.e., a free statement of intent) to cooperate is predictive of the sender's cooperative behavior (see Woike and Kanngiesser, 2019, for a recent and exhaustive review of this vast literature). In addition, Babutsidze et al. (2021) provide experimental evidence that this signal is correctly taken into account by the receivers across several communication protocols (ranging from plain text transcript to audio recording to video recording to face-to-face interaction) varying the amount of nonverbal content conveyed in the sender's message.

However, communication in social interactions is not only about words. Under the standard definition applied in animal studies, communication consists of any *behavior in [. . .] the sender [. . .] which evokes a response in [. . .] the receiver*; for humans, this definition may also encompass notions of conscious intent or volition (see Chapter 2 in Ekman, 2006, p. 21). Accordingly,

another important result in the experimental literature is that the role of communication as means of signaling trustworthiness is not restricted to its purely verbal content. The nonverbal components of communication—such as facial displays, body movements, tone of voice—also play a role in signaling trustworthiness. For instance, echoing the evolutionary argument by Boone and Buck (2003) that spontaneous emotional expressivity can act as a marker of pro-social motives like trustworthiness and cooperativeness, Brown et al. (2003) provide experimental evidence that altruists are perceived as more expressive than non-altruists. Oda et al. (2009b) highlight a particular dimension of human emotional expressivity: altruists are more likely to display genuine smiles. In the same vein, Centorrino et al. (2015) investigate the role of smiles in creating social exchange. Using an incentivized trust game with pre-play communication stage in which the trustee transmits to the trustor a pre-recorded video message with standardized verbal content, they find that the trustees conveying genuine smiles in their recordings also tend to be more trustworthy (i.e., generous toward their partners), and incite higher trust from others. An important line of experimental work also shows that information gathered through a brief, controlled and superficial access to physical characteristics of an unknown counterpart—their face, body gestures, way of expression (sometimes referred to as "thin slices" of observable information)—may help detect cooperativeness in various types of economic interactions (for a recent survey, see Bonnefon et al., 2017).

Our paper contributes to the growing experimental literature on detecting other-regarding preferences based on "thin slices" of observable information. We investigate the extent to which the recognition of trustworthiness in social interactions is a pancultural trait. We address the following question: Does cultural distance matter when it comes to detecting trustworthiness in social interactions? We build on a series of previous experiments by Oda et al. (2009a) and Tognetti et al. (2018) who offer a cross-cultural (Japan vs. France) comparison of the ability to detect the degree of altruism of Japanese subjects based on a short and muted video recording taken in a context which is unrelated to the target behavior. Tognetti et al. (2018) interpret the main finding—the general capacity (inability) of the Japanese (French) subjects to distinguish between altruistic and non-altruistic Japanese subjects based on the provided visual stimuli—as evidence that the nonverbal cues of prosociality are specific to one's culture rather than universally detectable. Our laboratory experiment is based on a variation of the trust game (Berg et al., 1995) with moral hazard, known as the hidden-action game (Charness and Dufwenberg, 2006). Our first set of stimuli comes from the previous experimental dataset reported by Babutsidze et al. (2021). It consists of video recordings of short, free-form pre-play statements delivered by the trustees to the trustors in direct face-to-face interactions happening in Nice, France. We provide the nonverbal content of those recordings as stimuli in an incentivized task in which subjects need to correctly predict the decisions previously made by the trustees. To allow for a cross-cultural comparison of prediction accuracy, this part of experiment relies on a different French sample (Lyon), as well as on a Japanese sample (Osaka).

As compared to the standard prediction tasks employing the "thin slice" paradigm, our methodological focus on nonverbal communication is novel and taps into the behavioral ecology of laboratory experimentation with social interactions. From the behavioral ecology perspective, facial displays are specific to intent and context, are issued in the service of social motives, and are interpretable in the context of interaction (see, e.g., Chapter 7 in Fridlund, 1994). In the words of Chovil and Fridlund (1991):

> *Facial displays are a means by which we communicate with others. Like words and utterances, they are more likely to be emitted when there is a potential recipient, when they are useful in conveying the particular information, and when that information is pertinent or appropriate to the social interaction.* (p. 163)

Clearly, this argument also applies to other components of nonverbal communication, such as gestures and body language. However, the previous studies—including those mentioned above (the study by Centorrino et al., 2015, is a notable exception), as well as the later contributions by, e.g., Van Leeuwen et al. (2018) and Oda et al. (2021)—are typically based on visual stimuli which are strongly dissociated from the social context in which the predicted target behavior (i.e., detection of certain facets of cooperativeness, such as altruism, trustworthiness, reciprocity) occurs. This is either because the visual stimuli used therein only consist of a neutral mugshot picture (like in our first control condition—PHOTO) or a neutral video recording with made-up content (like in our second control condition—neutral video, henceforth VIDNE)[1]. Thus, such standard design may only capture the extent to which certain morphological characteristics and general expressivity can be helpful in predicting human behavior. Our main condition (loaded video, henceforth VIDLO) extends this standard setup by providing the visual stimuli that belongs to the same social context as, and thus is intertwined with, the target behavior—the personal statement made by a trustee in front of the trustor prior to the decision-making stage of the trust game. Thus, the "thin slice" of observable information and the subsequent target behavior are both components of the same social interaction[2].

We find several consistent patterns of prediction-making in our two samples. For both samples, the overall rates of accurate

[1]These two sets of stimuli come from our previous experimental work reported in Zylbersztejn et al. (2020) and Babutsidze et al. (2021).

[2]For a similar approach based on non-experimental data see, e.g., Belot et al. (2010, 2012), Sylwester et al. (2012), Van den Assem et al. (2012), Turmunkh et al. (2019). They use data from a TV game show—*The Golden Balls*—which consists of a high stake prisoner's dilemma environment with a pre-play stage of natural face-to-face communication moderated by the host. Despite the clear virtues in terms of behavioral ecology, some features of these data fall short of the rigorous requirements of experimental control that is achieved in our laboratory setting. First, there is a continuous two-way communication between participants, so each subject acts both a sender and a receiver of messages. In our design, the players' roles in the process of communication are unique and reflect their respective tasks in the game. Second, in a TV game show the process of communication may be interrupted, and its content affected by a third party: the game host. For instance, often times the host talks one player into making a promise to cooperate with the other player. Our design rules our any possibility of such interference, allowing for a free and uninterrupted flow of communication from the trustee to the trustor.

detection of trustworthiness in strangers based on "thin slices" of observable information remain constant across the three types of stimuli. Moreover, we look at certain morphological traits of the target agents (facial masculinity, asymmetry, and weight-to-height ratio, as well as sex) and find that both the French and the Japanese subjects resort to the same heuristics (thus exhibiting similar biases) when making judgments about others' trustworthiness.

Nonetheless, some notable differences also arise across the two cultures. Overall, the VIDLO condition is the only instance where we observe predictions being made with a "better than chance" accuracy. However, this only happens for the Japanese subjects; despite cultural proximity with the target agents, the French subjects are not able to distinguish between the trustworthy and untrustworthy ones after observing the nonverbal content of communication. To shed more light on this (somewhat surprising) outcome, we then extend our empirical analysis with a new dataset containing the same set of recordings, this time with unmuted verbal content. The availability of this verbal content significantly improves prediction accuracy of the French subjects in the unmuted VIDLO condition. In line with the previous studies, we confirm a particular role of voluntary promises in signaling trustworthiness among strangers. This suggests that cultural distance (proximity) makes people relatively sensitive (insensitive) to the relevant components of nonverbal content of communication that go beyond basic morphological heuristics. Rather, within cultural proximity attention is attuned to the relevant aspects of the verbal content of communication. Hence, cultural distance *(i)* is not detrimental for the comprehension of the nonverbal content of communication (if anything, it is exactly the opposite), and *(ii)* it may affect the perception of the different components of communication in social interactions.

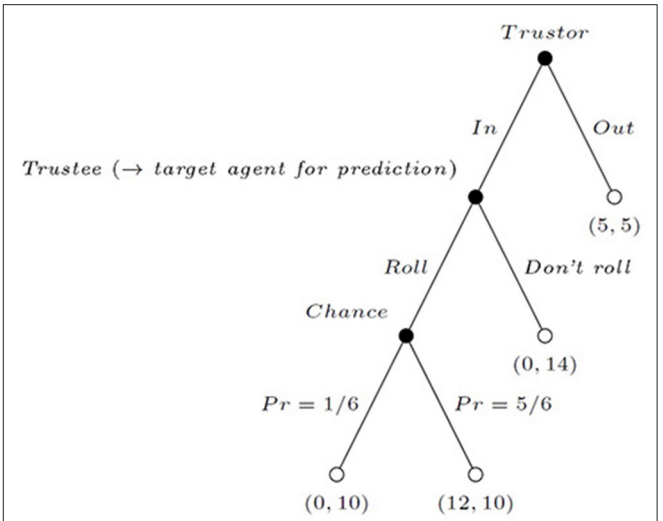

FIGURE 1 | Experimental hidden action game.

2. EXPERIMENTAL DESIGN

2.1. Experimental Stimuli for the Prediction Task

For implementing the prediction task, we exploit the dataset previously reported in Babutsidze et al. (2021). That study is based on the hidden action game by Charness and Dufwenberg (2006) presented in **Figure 1**. All payoffs are in Euros. The game is played between two parties: the trustor and the trustee. The trustor may either choose an outside option *Out* which yields 5 to both players and ends the interaction, or go *In*. Then, the trustee may either choose to *Roll* a die (which yields 12 to the trustor and 10 to the trustee with the probability of 5/6, and 0 to the trustor and 10 to the trustee with the probability of 1/6), or not to *Roll* (yielding 0 to the trustor and 14 to the trustee with certainty). This game provides a simple setting for studying voluntary cooperation under moral hazard: incentives are not aligned between the two parties, and earning 0 is not perfectly informative for the trustor about the trustee's action. For this reason, we believe that the hidden action game offers a conservative way of measuring trustworthiness compared to the classic trust game due to Berg et al. (1995).

Like Charness and Dufwenberg (2006), we simultaneously elicit both players' decisions. Namely, the trustee makes a decision without knowing the trustor's move, and that decision is only implemented had the trustor gone *In*. The game is preceded by a pre-play stage with face-to-face communication and is implemented as follows. In every experimental session, six trustors are seated in one room (in separate cubicles and without the possibility to communicate) where they make all their decisions in the game. Each of the six trustees, in turn, makes an individual decision in a separate room. Prior to the decision-making stage of the game, each trustee is given approximately two minutes to prepare a short statement for the trustors. At this point, we provide an additional set of instructions emphasizing the fact that the statement may affect the trustors' decisions and, consequently, the trustee's gain from the experiment[3]. Then, the trustee enters the trustors' room and delivers the statement in front of them. The trustors can clearly see and hear the trustee, and the trustee can also observe the trustors while delivering the statement. After that, the trustee leaves to a separate room to make a decision. Simultaneously, the six trustors privately make their decisions. At the end of the experiment, the trustees and the trustors are randomly and anonymously matched into six pairs for payments. Further implementation details, including the instructions used in that experiment, are provided in **Appendices A1, A2**.

In addition to the trustees' decisions in the experimental game (and, if relevant, the outcomes of die rolls), our dataset contains several recordings. Following Van Leeuwen et al. (2018), upon arrival to the laboratory and before learning about the rules of the hidden action game, each subject in the role of a trustee is invited to a separate room for a mugshot picture and a standardized video recording: the subjects are asked to read a

[3]This information is part of the summary of the hidden action game experiment provided in the instructions employed in the current study.

short extract from a printer instruction manual, while keeping a neutral facial expression. These two sources of information are used, respectively, in our PHOTO and VIDNE (neutral video) treatments. Finally, the trustees are also video recorded while making a statement in the pre-play communication stage of the hidden action game. We use this information in our VIDLO (loaded video) treatment.

The original database in Babutsidze et al. (2021) includes 41 trustees and has been collected at Laboratoire d'Economie Expérimentale de Nice (LEEN) of the University of Nice, France. These participants gave their explicit consent (*i*) for being recorded, and (*ii*) for those recordings being used for strictly scientific purposes in related experimental studies. For the sake of the present study, we restrict the set of stimuli to an ethnically homogeneous group of subjects classified as Caucasian by an independent coder ($N = 26$; 13 females; average age 22.58, SD = 3.18). Furthermore, we do not disclose the location in which this sample was collected. The purpose of these design choices is to minimize the role of ethnic and/or racial biases in reaction to each stimulus. These trustees are the target agents in the prediction tasks implemented in the main experiment. Among these 26 target agents, 16 chose to *Roll*. The 26 stimuli are presented in random order.

2.2. Main Experiment

Our main experiment is implemented through a between-subject design and involves a total of $N = 273$ participants (97% students; 53% Japanese; 40% females; average age 21.51, SD = 3.89). **Table 1** provides further information about the assignment of subjects in our 3×2 factorial design: across the three treatments (PHOTO, VIDNE, VIDLO) and two locations (Lyon, France and Osaka, Japan). For each of the six conditions, we run two experimental sessions that took part in May 2018 in the Experimental Economics Laboratory at the Institute of Social and Economic Research (ISER) at Osaka University in Japan, and in December 2019 in the GATE-Lab, an experimental laboratory at the GATE Lyon-Saint-Etienne research institute in France[4]. Experimental sessions were entirely computerized: subjects were recruited using ORSEE (Greiner, 2015), and all the experimental tasks were programmed in z-Tree (Fischbacher, 2007).

Participants make a series of twenty six predictions of trustees' behavior in an earlier hidden action game (i.e., whether the target person rolled a die or not). A correct (an incorrect) prediction is worth 10 (2) euros in the experiments run in France, and 1,200 (240) yen for those run in Japan. No feedback is provided from one prediction to the other, and two rounds out of twenty six are randomly drawn for payoff at the end of each experimental session. Unlike some previous studies using the "better than chance" paradigm, we do not constrain the base rate of "success" at the chance level of 50%[5]. Our experimental treatments progressively enrich the set of information about the trustee that is provided to the subject prior to making a prediction: either a mugshot picture (PHOTO), or one of muted video recording: either showing that person making a non-strategic statement that has been recorded before (and independently of) the experimental hidden action game (VIDNE), or a loaded one in which the trustee makes a strategic pre-play statement in front of the trustors (VIDLO)[6].

2.3. Experimental Procedures

Upon arriving to the lab, subjects are seated in individual cubicles and informed about the general rules of a lab experiment[7]. The preliminary part of the session consists of a basic socio-demographic questionnaire (age, sex, education, major, current occupation, score at the *baccalauréat* exam at the end of high school in the case of French subjects), as well as a set of (moderately) incentivized and non-incentivized computerized tasks designed to measure specific individual characteristics[8]. After that, subjects receive paper instructions describing the

[4]Since acquaintance between the experimental subjects in Lyon and the target agents recorded in Nice is unlikely, one may plausibly assume that performance in the prediction task actually measures the individual capacity to detect cooperativeness in strangers. See Centorrino et al. (2015) and Van Leeuwen et al. (2018) for a similar approach.

[5]Under the "better than chance" paradigm, subjects typically receive randomly generated pairs of stimuli—one coming from a person that exhibited certain behavior, and one from another person that did not (which is common knowledge; see, e.g., Bonnefon et al., 2013; Van Leeuwen et al., 2018). Another method is to show a series of individual stimuli and inform the subjects about the underlying base rate (50%) of a given behavioral outcome, but not about the length of the series (Vogt et al., 2013). Although the "better than chance" paradigm provides a clean and simple benchmark for measuring the extent to which observable information affects prediction accuracy, it has been criticized for the lack of external validity. As pointed out by Todorov et al. (2015a), this criterion seems weak when it comes to evaluating prediction performance in many real-world environments in which the different types of behavior are unequally prevalent. Following this argument, in our experiment the lack of information about the underlying base rate adds to the overall complexity of the prediction task. See Fetchenhauer et al. (2010) for a similar approach.

[6]The average duration of a recording in VIDNE (VIDLO) is 33.38 (25.85) s with SD 5.27 (13.31) and range 27–49 (11–60). Given that PHOTO only involves static content, in this treatment we adopted the following procedure. Each time, the picture of the target person is displayed on the computer screen. After 15 s, a button appears underneath the picture allowing the subject to move on to the prediction-making stage. This choice came about as the outcome of the pilot test of our experimental setup, and appears to be a remedy against the risk of "under-exposing"—the exposure to the displayed content being insufficient to fully grasp all the available information, as well as "over-exposing"—participants eventually getting inattentive due to factors such as boredom, impatience, or a decay in their interest in the displayed static content.

[7]The original instructions are in French for the experiments run in Lyon, and in Japanese for those run in Osaka. Their English version can be found in **Appendix A3**.

[8]This procedure closely follows Babutsidze et al. (2021), and its details can be found therein. The set of tasks includes standard measures of other-regarding preferences (Social Value Orientation, SVO, task by Murphy et al., 2011), cognitive skills (3-item Cognitive Reflection Test, CRT, Frederick, 2005), the theory of mind (The Reading the Mind in the Eyes Test, RMET, Baron-Cohen et al., 2001), risk preferences (Gneezy and Potters, 1997), and general trust attitudes (based on the German Socio-Economic Panel Study, SOEP). In most cases, we find no differences between the two samples—this applies to distributional preferences, cognitive skills, risk preferences, and general attitudinal trust toward other people. One notable exception, however, is the theory of mind: the French subjects attain a significantly higher score on RMET (mean scores of out 34: 27.28 vs. 21.71, $p < 0.001$ based on two-sided t-test). However, in neither experimental environment of our 3×2 experimental design we observe statistically significant (Spearman's rank) correlation between this measure of the theory of mind and individual prediction accuracy rates (ρ varies between 0.04 and 0.24, all $p > 0.117$). This result stands in line with the previous evidence reported by Sylwester et al. (2012).

TABLE 1 | Average prediction accuracy rates across countries and treatments: aggregate data.

	France	Japan	*p*
PHOTO	51.0% ($N = 43$)	50.9% ($N = 50$)	0.972
VIDNE	52.1% ($N = 37$)	51.6% ($N = 49$)	0.814
VIDLO	49.9% ($N = 48$)	52.3% ($N = 46$)	0.209
p	0.533	0.779	

p-values in the last column (row) come from a two-sided t-test (F-test) of the equality of prediction accuracy rates between countries for a given treatment (across treatments within a given country).

TABLE 2 | Predicted vs. actual behavior: prediction accuracy across countries and treatments.

	$Pr(1[PredictionRoll]) = 1$			
If 1[*ActualRoll*] =	**0**	**1**	**0**	**1**
	(p_{DR})	**(p_R)**	**(p_{DR})**	**(p_R)**
Condition	**France**		**Japan**	
PHOTO	44.2%	46.8%	38.2%	41.6%
VIDNE	45.3%	49.8%	42.5%	46.5%
VIDLO	50.0%	49.9%	36.2%	42.4%

1[*PredictionRoll*] (1[*ActualRoll*]) *is set to 1 if a subject predicts that the target player rolled a die (if the target player actually rolled a die) in the previous experiment, and to 0 otherwise.*

details of the previous hidden action game experiment, as well as their own experimental task.

Those instructions are read aloud by the experimenter, any remaining questions are immediately answered, and the experiment moves to its main stage, as described above. In addition to earnings in the experimental tasks, there is a show-up fee of 5 euros for the French participants, and 600 yen for the Japanese participants. The duration of a session was approximately 1h30 and the average total payoff was 23 euros in France and 3,175 yen in Japan[9].

3. AGGREGATE RESULTS

Table 1 provides an overview of the average prediction accuracy rates (i.e., the likelihood that a randomly chosen subject makes a correct prediction in a randomly chosen round of the experiment) across treatments and cultures. This aggregate evidence points to *(i)* no effects of varying the sources of observable information on prediction accuracy within a given culture, and *(ii)* no intercultural variation of prediction accuracy in any of the three information conditions.

As a next step of our analyses, we disaggregate those data by looking at prediction accuracy rates conditional on the target agent's actual decision—either *Roll* or *Don't roll*. We employ the statistical framework from Zylbersztejn et al. (2020) to draw a link between the predicted behavior and the actual behavior. Suppose that p_R (p_{DR}) is the probability of making a prediction *Roll* conditional on the target person actually choosing to *Roll* (*Don't roll*). $p_R = p_{DR}$ implies that subjects are unable to discriminate between trustworthy and untrustworthy target players, and make a prediction *Roll* at a constant rate (freely ranging between 0 and 1) irrespective of the trustee's underlying type. $p_R > p_{DR}$, in turn, implies that subjects are able to detect the target player's type at least partially which makes them more likely to make a prediction *Roll* for those who actually rolled a die[10]. The corresponding prediction rates are summarized in **Table 2**, and statistical support for mean comparisons is provided in **Table 3**. For each of the three information conditions (PHOTO, VIDNE, VIDLO), we regress an indicator variable 1[*PredictionRoll*] (set to 1 if one predicts that the target person rolled a die in the previous experiment, and to 0 otherwise) on another indicator variable 1[*ActualRoll*] (set to 1 if the target person actually rolled a die in the previous experiment, and to 0 otherwise), 1[*Japan*] (set to 1 for the Japanese subjects, and to 0 otherwise), as well as their interaction. The intercept (denoted α_0) captures the aggregate likelihood of predicting *Roll* for those trustees that did not actually roll a die (such that $\alpha_0 = p_{DR}$). Our key measure of interest is given by coefficients α_1 and $\alpha_1 + \alpha_3$ which provide the respective empirical estimates of the difference between p_R and p_{DR} (i.e., the extent to which subjects are able to distinguish between those who rolled and those who did not) for the French and Japanese subjects[11].

The main message that stems from this analysis is the following: only in one instance—the VIDLO condition implemented in Japan—the difference $p_R - p_{DR}$ is positive and statistically significant (testing $H_0 : \alpha_1 + \alpha_3 = 0$ yields $p = 0.013$), indicating that these subjects can tell better than chance between trustworthy and untrustworthy target agents. In the five remaining cases, we observe $p_R - p_{DR}$ to be small and not significantly different from zero[12].

3.1. The Role of Target Player's Facial Characteristics

The model reported in **Table 4** extends the analyses from **Table 3** by accounting for several individual characteristics of the target player. Beside the treatment and 1[*ActualRoll*] indicator variables, as well as their interactions (coefficients $\beta_1, \ldots, \beta_5$), the set of explanatory variables includes several facial measurements

[9]At the time when our experiments were run, the usual exchange rate oscillated around 1 euro = 130 yen.

[10]For a perfect ability to discriminate between the two types of trustees, we would have $p_R = 1$ and $p_{DR} = 0$.

[11]This specification overcomes the usual caveats of using OLS for binary choice data. First, our specification with cluster-robust variance-covariance matrix is also heteroscedasticity-robust. Second, the forecasting issue (i.e., predicted probabilities going beyond the [0; 1] range) does not arise for binary explanatory variables: here, an estimated coefficient simply boils down to the respective choice proportion in a given experimental condition.

[12]To provide further statistical support for this result, we run additional analyses based on paired *t*-test. For each subject, we calculate the rate of prediction *Roll* for untrustworthy target agents, and then compare it to analogous rate calculated for the trustworthy ones. In all conditions other than VIDLO conducted in Japan, we find Bayes factor BF_{10} between 0.15 and 0.45 for a two-sided test, clearly testifying against the alternative hypothesis of a difference between the two rates. For the remaining condition, $BF_{10} = 2.23$, thus yielding support (although not overwhelming) for the alternative hypothesis of different rates. Repeating the same exercise for standard (i.e., non-Bayesian) *t*-test yields *p*-values and conclusions in line with those reported in **Table 3**.

TABLE 3 | Predicted vs. actual behavior: regression analysis.

	PHOTO		VIDNE		VIDLO	
	Coef. (SE)	*p*	Coef. (SE)	*p*	Coef. (SE)	*p*
Intercept (α_0)	0.442	<0.000	0.453	<0.000	0.500	<0.000
	(0.042)		(0.031)		(0.025)	
1[*ActualRoll*] (α_1)	0.027	0.212	0.045	0.162	−0.001	0.955
	(0.021)		(0.032)		(0.026)	
1[*Japan*] (α_2)	−0.060	0.267	−0.028	0.535	−0.138	0.002
	(0.054)		(0.044)		(0.044)	
1[*ActualRoll*] × 1[*Japan*] (α_3)	0.007	0.816	−0.006	0.895	0.063	0.086
	(0.032)		(0.042)		(0.036)	
$H_0 : \alpha_1 + \alpha_3 = 0$		0.159		0.134		0.016
Prob > *F*		0.172		0.171		0.005
N of obs./clusters		2418/93		2236/86		2444/94

Results of OLS regression models of the individual prediction (indicator variable 1[*PredictionRoll*] = 1 *if one predicts that the target player rolled a die in the previous experiment; 0 otherwise) on a set of indicator variables:* 1[*ActualRoll*] *(set to 1 if the target player actually rolled a die in the previous experiment, and to 0 otherwise),* 1[*Japan*] *(set to 1 for the Japanese subjects, and to 0 otherwise), as well as their interaction. Observations are clustered for each individual, standard errors (SE) are cluster-robust.*

of the target agent (masculinity, asymmetry, weight-to-height ratio; coefficients $\beta_6, \beta_7, \beta_8$, respectively) and that person's sex (1[*Female*] = 1 for females, 0 for males; coefficient β_9)[13]. Furthermore, we include an indicator variable 1[*Japan*] (set to 1 for the Japanese subjects and to 0 otherwise; coefficient γ_0) and its interactions with all the previous variables (coefficients $\gamma_1, \ldots, \gamma_9$). The model is estimated with pooled data[14].

This new specification *(i)* provides robustness analysis of the effects reported in **Table 3** after controlling for a rich set of target player's observable characteristics, and *(ii)* allows for testing (through coefficients γ_i) for cultural differences with respect to any of the dimensions captured by the model.

In relation to *(i)*, the model confirms that only in one instance—the VIDLO condition implemented in Japan—relevant information can be extracted from the recordings in a way that improves prediction accuracy above chance[15].

TABLE 4 | Facial characteristics and predictions across cultures: regression analysis.

Coef. number (*i*): Variable	β_i (SE)	*p*	γ_i (SE)	*p*
0: Intercept	0.312	0.005	0.096	0.513
	(0.110)		(0.147)	
1: 1[*ActualRoll*]	0.019	0.363	0.014	0.671
	(0.021)		(0.032)	
2: 1[*VIDNE*]	0.011	0.836	0.033	0.639
	(0.052)		(0.069)	
3: 1[*VIDLO*]	0.058	0.237	−0.077	0.263
	(0.049)		(0.069)	
4: 1[*ActualRoll*] × 1[*VIDNE*]	0.019	0.625	−0.013	0.804
	(0.038)		(0.052)	
5: 1[*ActualRoll*] × 1[*VIDLO*]	−0.028	0.403	0.056	0.250
	(0.034)		(0.048)	
Target agent's characteristics				
6: Facial masculinity	0.018	<0.000	0.007	0.219
	(0.004)		(0.006)	
7: Facial asymmetry	0.003	0.292	−0.004	0.212
	(0.003)		(0.003)	
8: Facial width-to-height ratio	0.002	0.970	−0.076	0.183
	(0.042)		(0.057)	
9: 1[*Female*]	0.087	<0.000	0.007	0.822
	(0.022)		(0.030)	

Results of OLS regression models of the individual prediction (indicator variable 1[*PredictionRoll*] = 1 *if a subject predicts that the target agent rolled a die in the previous experiment; 0 otherwise) on a set of explanatory variables:* 1[*ActualRoll*] *(set to 1 if the target agent actually rolled a die in the previous experiment, and to 0 otherwise) and treatment indicator variables* 1[*VIDNE*] *and* 1[*VIDLO*] *set to 1 for a given treatment and 0 otherwise (*1[*PHOTO*] *is the omitted reference condition), as well as their interactions; target player's individual characteristics: facial masculinity, facial asymmetry, facial weight-to-height ratio, as well as sex (*1[*Female*] *is set to 1 for females, and to 0 for males). This subset of explanatory variables is associated with coefficients* β_i *(first column). The model also includes an indicator variable* 1[*Japan*] *(set to 1 for the Japanese subjects, and to 0 otherwise) as well as its interactions with all the previous variables; these explanatory variables are associated with coefficients* γ_i *(last column). Observations are clustered for each individual (7,098 observations in 273 clusters), standard errors (SE) are cluster-robust.*

Regarding *(ii)*, the model indicates that, irrespective of the culture of origin, subjects systematically condition their predictions on certain observable characteristics of the target players. It is important to note at this point that, based on our empirical data, this information should be considered as irrelevant for predictions, since neither of the four individual characteristic included in the model happens to be associated with the observed behavior in the hidden action game[16]. Nonetheless, two of these observable characteristics—facial masculinity and sex—are statistically significant predictors of

[13]The three facial measurements have been obtained from the mugshot pictures used in the PHOTO treatment. For computation, we followed standard procedures adopted from Van Leeuwen et al. (2018) and summarized in **Appendix B**. See Stirrat and Perrett (2010) and Rodríguez-Ruiz et al. (2019) for a further discussion on the potential role of these facial characteristics in cooperation detection.

[14]Estimated coefficients from a logistic regression give comparable results. The main advantage of using OLS instead of a non-linear model is that in the latter, the only meaningful way to quantitatively interpret the estimated coefficients is by computing marginal effects. However, the use of marginal effects becomes problematic in the presence of interactions terms. The literature does not provide a clear-cut solution to this issue (see Ai and Norton, 2003; Greene, 2010). Since the statistical testing of interactions is central to the exercise reported in **Table 4**, we favor OLS (which allows us to easily operationalize interaction terms in the model) over a non-linear specification.

[15]For the French sample, we test the significance of coefficients β_1 (PHOTO), $\beta_1 + \beta_4$ (VIDNE), $\beta_1 + \beta_5$ (VIDLO), neither of which is found to be significant ($p = 0.363, p = 0.231, p = 0.740$, respectively). For the Japanese data, the corresponding tests involve coefficients $\beta_1 + \gamma_1$ ($p = 0.171$), $\beta_1 + \beta_4 + \gamma_1 + \gamma_4$ ($p = 0.145$), $\beta_1 + \beta_5 + \gamma_1 + \gamma_5$ ($p = 0.018$).

[16]Two-sided ranksum test does not detect significant differences in facial masculinity ($p = 0.959$), asymmetry ($p = 0.520$), or width-to-height ratio ($p = 0.382$) between those that *Roll* ($N = 14$) and those that do not ($N = 12$). Moreover, both females and males choose to *Roll* with the same frequency (in 7 out of 13 cases); χ^2 test yields $p = 1.000$.

assessed trustworthiness. Importantly, such biased judgment of trustworthiness persists across cultures[17].

3.2. The Role of Verbal Content

So far, our experimental evidence points to a general incapacity of the French subjects to accurately predict strangers' trustworthiness from different stimuli containing nonverbal content, despite cultural proximity between the two parties. Strikingly, this failure occurs even for the strategically loaded video recordings provided in the VIDLO condition—stimuli that helps the more culturally distant Japanese subjects distinguish between the target players' types. In this section, we are asking whether and to what extent this insufficiency can be fixed by further providing the verbal content of VIDLO recordings. For this sake, we revisit the dataset from our previous experiment reported in Zylbersztejn et al. (2020). That experiment involves the same subject pool (GATE-Lab, Lyon, France) and the same video recordings, but this time with sound turned on (henceforth referred to as the VIDLO_SOUND condition)[18].

Evidence reported in the first part of **Table 5** suggests that, unlike the sound-off VIDLO condition, the VIDLO_SOUND condition with verbal content of strategic statements allows the French subjects to distinguish between the target agents' types. Even though the ability to identify untrustworthy target players does not vary between the two conditions, we observe that VIDLO_SOUND improves detection of trustworthiness. Furthermore, in line with a large body of experimental literature (see Woike and Kanngiesser, 2019, for a recent review), these data indicate that a particular facet of verbal content—a promise to *Roll*—constitutes an informative signal of cooperative intentions: target agents who made such a promise are more than twice as likely to *Roll* than the target players not making such a promise[19].

TABLE 5 | Verbal and nonverbal content in VIDLO: evidence from the French data.

Average rate of prediction *Roll* per stimulus			
If 1[*ActualRoll*] =	**0 (*N* = 12)**	**1 (*N* = 14)**	***p* (ranksum test)**
VIDLO_SOUND	47.9%	66.2%	0.024
VIDLO	50.0%	49.9%	0.918
p (signrank test)	0.814	0.035	
If 1[*PromiseRoll*] =	**0 (*N* = 10)**	**1 (*N* = 16)**	***p* (ranksum test)**
VIDLO_SOUND	47.6%	64.1%	0.045
VIDLO	54.8%	46.9%	0.119
p (signrank test)	0.445	0.015	

The unit of observation is the rate of prediction Roll observed for a given recording ($N = 26$) in a given condition. 1[ActualRoll] (1[PromiseRoll]) is set to 1 if the target player actually rolled a die (made a promise to roll a die) in the previous experiment, and to 0 otherwise.

As shown in the bottom part of **Table 5**, French subjects in the VIDLO_SOUND condition effectively pick up on this signal and attribute higher trustworthiness to promise-makers, in stark contrast to the sound-off VIDLO condition. We also note that the same holds for the Japanese sample: the respective rates are 48.2% without a promise, and 37.5% with a promise ($p = 0.118$, two-sided ranksum test). This, in turn, suggests that the nonverbal information the Japanese subjects pick up on when forming judgments is unrelated to the verbal content conveyed in the strategic statements[20].

4. CONCLUSION

Our study contributes to several strands of ongoing debate on how observing others may be helpful for predicting their behavior in social interactions. We take a cross-cultural perspective and

[17] As shown in **Table 4**, coefficients β_6 and β_9 are positive and significant. This suggests that, *ceteris paribus*, higher facial masculinity, as well as being a female, increases the likelihood of being perceived as trustworthy person by a French subject. Insignificance of coefficients β_7 and β_8, in turn, suggests that there is no statistical association between being perceived as a trustworthy person and one's facial asymmetry or width-to-height ratio. The same results hold for the Japanese sample: coefficients $\beta_i + \gamma_i$ are found to be positive and significant for $i = 6$ and $i = 9$ (both $p < 0.001$), but not for $i = 7$ ($p = 0.489$) and $i = 8$ ($p = 0.057$). Finally, a joint test of $H_0 : \gamma_6 = \gamma_7 = \gamma_8 = \gamma_9 = 0$ does not reject the joint nullity of the differences between the respective coefficients across the two samples ($p = 0.434$).

[18] In Experiment 1 reported in Zylbersztejn et al. (2020), there are three conditions: neutral mugshot pictures (analogous to the PHOTO treatment used herein), neutral videos and loaded videos (analogous to VIDNE and VIDLO used herein, with one key difference: the sound is on, so that the subjects not only watch, but also listen to the target player's statement). Compared to the present experiment, the stimuli in that experiment are provided in a slightly different manner: the total set of stimuli consists of 41 items (including the 26 stimuli employed herein), and each subject inspects a randomly drawn sequence of 20 items. Focusing on the subset of the 26 target players that are common for both experiments, in Zylbersztejn et al. (2020) each item is shown to 21 subjects on average (range: 15–30 for pictures, 16–28 for both types of videos), while in the present experiment each subject inspects all 26 items. We believe that these differences do not distort subjects' predictions, so that the observations coming from the two sources remain comparable. Exploiting the data from the PHOTO condition (in which the stimuli contain the same information in both experiments), we compare the rates of prediction *Roll* for each of the 26 items registered in the present experiment to those from Zylbersztejn et al. (2020); signrank test yields $p = 0.354$. The same exercise for the VIDNE condition—in which neutral video recordings are muted in the present experiment, and contain the target player's voice in Zylbersztejn et al. (2020)—yields $p = 0.525$. This, in turn, corroborates the previous finding from Vogt et al. (2013) that hearing a stranger's voice in a neutral context does not *per se* affect the perception of that person's cooperativeness.

[19] The respective likelihoods are 69% ($N = 16$) and 30% ($N = 10$). χ^2 test yields $p = 0.054$. Like Charness and Dufwenberg (2006), we define a promise as a statement of intent to *Roll*. Note that, as raised by Houser and Xiao (2011), the *ex post* interpretation of free-form messages is a major methodological challenge for the experimenter. The literature still lacks a common consensus on whether this should involve content analysis carried out by the experimenter (Charness and Dufwenberg, 2006), by independent coders (He et al., 2017), through an incentivized coordination game (Houser and Xiao, 2011), or by asking the subjects for their own interpretation (Servátka et al., 2011). Our classification method echoes the recent study by Schwartz et al. (2019). All statements were classified as promises or no-promises by two independent coders. The first coder classified the content of messages while preparing the transcripts of the trustees' statements. Then, another coder received a complete list of transcripts and independently classified each of them. Ties were broken by one of the authors.

[20] We note that implementing VIDLO_SOUND in the Japanese sample does not seem as a meaningful exercise due to a high degree of uncertainty as of the extent to which these subjects comprehend the verbal content of an improvised statement in French. Although their skills in foreign languages may be insufficient for understanding everything, they may nonetheless comprehend (or believe to be understanding) a part of this content (e.g, single words or sentences). This leaves an important degree of uncontrolled variation related to what a Japanese subject could potentially understand, how much, and how well, thus rendering the overall results hard to interpret.

focus on the ability to detect a stranger's proneness to conditional cooperation, or trustworthiness, based on "thin slices" of observable information. As noted by Olivola et al. (2014), many important social decisions (e.g., political elections and court sentences) are made on the basis of people's facial appearance, and individuals tend to agree when it comes to judging which faces look trustworthy[21]. Furthermore, evidence from laboratory experiments employing economic games suggests that people exhibit less trust toward partners with untrustworthy looking faces, even when given relevant information about their past behavior (Chang et al., 2010; Rezlescu et al., 2012).

Is this information actually useful for making accurate judgments? Olivola et al. (2014) and Todorov et al. (2015a) qualify "face-ism" as a judgment bias, since social inferences based on facial appearance tend to be inaccurate and unreliable. On the other hand, Bonnefon et al. (2013, 2017) argue that physical cues provided via "thin slices" of information may nonetheless contain "kernels of truth," and observing one's face, body language, way of expression may help detect cooperation in various economic interactions.

We believe that our novel experimental evidence goes some way in reconciling both of these claims. Echoing a closely related study by Tognetti et al. (2013), our experimental data point to a judgment bias that meshes well with the notion of "face-ism": subjects account for morphological traits of the target agents, even though the latter are not associated with the actual behavior. Extending these previous findings, we further document that this bias persists across cultures and attains the same magnitude in both the French and the Japanese sample.

At the same time, we believe that "kernels of truth" may well exist alongside the aforementioned biased judgments. However, our data reveals that predicting behavior in social interactions requires that "thin slices" contain direct social cues (like in our VIDLO condition), rather than being restricted to the purely physical ones (i.e., with no relation to the social context of the interaction—like in our PHOTO and VIDNE conditions). The dominant role of social context relative to physical attributes is consistent with a recent study by Jaeger et al. (2020) who show that people are generally unable to detect the trustworthiness of strangers based solely on their facial appearance. Importantly, we find that this effect varies considerably across cultures. Despite cultural distance, Japanese subjects are sufficiently attuned to the nonverbal content of strategic statements to be able to distinguish between trustworthy and untrustworthy target agents in the VIDLO condition. Within cultural proximity, French subjects tend to ignore these cues. Nonetheless, when additionally provided with verbal content (like in our auxiliary VIDLO_SOUND condition), they become capable of correctly reading a credible signal of trustworthiness—namely, a voluntary promise to cooperate. Hence, we conclude that cultural distance is not *per se* helpful or detrimental for predicting trustworthiness. Rather, it affects ways in which people exploit observable information in social interactions.

In the closing lines, we would like to mention an important limitation of our study. Both the target agents used in the experimental stimuli, as well as the sample of participants to our experiment, are drawn from rather homogeneous student populations in France and Japan. While we see our study as an important step in documenting cross-cultural differences in trustworthiness detection, we also believe that there is a need for further evidence drawn from different sets of stimuli (e.g., including ethnicities other than the Caucasian ethnicity we focus on here) and more diversified samples of participants (e.g., coming from the general population).

[21] See Todorov et al. (2015b) for a systematic review of the empirical evidence on social attribution from faces.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the GATE-Lab Research Ethics Committee based at the Groupe d'Analyse et de Théorie Economique (UMR 5824). The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

FUNDING

We acknowledge the support from the following programs operated by the French National Research Agency (Agence Nationale de Recherche): DigiCom as a part of *UCA*^{JEDI} (ANR-15-IDEX-01) and LABEX CORTEX (ANR-11-LABX-0042) as a part of Université de Lyon (ANR-11-IDEX-007), as well as the Joint Usage/Research Center at ISER, Osaka University, and Grant-in-aid for Scientific Research, Japan Society for the Promotion of Science (15H05728, 18K19954, 20H05631).

ACKNOWLEDGMENTS

We are grateful to two referees, Jean-François Bonnefon, Astrid Hopfensitz, Sonja Vogt, as well as the participants of 2019 and 2020 INDEPTH workshops at GATE in Lyon for helping us improve the paper. Ynès Bouamoud, Yuki Hamada, Yasser Nabangui, Charlotte Saucet, and Hiroko Shibata provided quality research assistance. Quentin Thévenet provided valuable assistance with software programming. AZ acknowledges VISTULA Fellowship.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2021.727550/full#supplementary-material

REFERENCES

Ai, C., and Norton, E. C. (2003). Interaction terms in logit and probit models. *Econ. Lett*. 80, 123–129. doi: 10.1016/S0165-1765(03)00032-6

Babutsidze, Z., Hanaki, N., and Zylbersztejn, A. (2021). Nonverbal content and trust: an experiment on digital communication. *Econ. Inq*. doi: 10.1111/ecin.12998. [Epub ahead of print].

Baron-Cohen, S., Wheelwright, S., Hill, J., Raste, Y., and Plumb, I. (2001). The "reading the mind in the eyes" test revised version: a study with normal adults, and adults with asperger syndrome or high-functioning autism. *J. Child Psychol. Psychiatry All. Discip*. 42, 241–251. doi: 10.1111/1469-7610.00715

Belot, M., Bhaskar, V., and van de Ven, J. (2010). Promises and cooperation: evidence from a TV game show. *J. Econ. Behav. Organ*. 73, 396–405. doi: 10.1016/j.jebo.2010.01.001

Belot, M., Bhaskar, V., and Van De Ven, J. (2012). Can observers predict trustworthiness? *Rev. Econ. Stat*. 94, 246–259. doi: 10.1162/REST_a_00146

Berg, J., Dickhaut, J., and McCabe, K. (1995). Trust, reciprocity, and social history. *Games Econ. Behav*. 10, 122–142. doi: 10.1006/game.1995.1027

Bonnefon, J.-F., Hopfensitz, A., and De Neys, W. (2013). The modular nature of trustworthiness detection. *J. Exp. Psychol*. 142:143. doi: 10.1037/a0028930

Bonnefon, J.-F., Hopfensitz, A., and De Neys, W. (2017). Can we detect cooperators by looking at their face? *Curr. Direct. Psychol. Sci*. 26, 276–281. doi: 10.1177/0963721417693352

Boone, R. T., and Buck, R. (2003). Emotional expressivity and trustworthiness: the role of nonverbal behavior in the evolution of cooperation. *J. Nonverb. Behav*. 27, 163–182. doi: 10.1023/A:1025341931128

Brown, W. M., Palameta, B., and Moore, C. (2003). Are there nonverbal cues to commitment? An exploratory study using the zero-acquaintance video presentation paradigm. *Evol. Psychol*. 1:147470490300100104. doi: 10.1177/147470490300100104

Centorrino, S., Djemai, E., Hopfensitz, A., Milinski, M., and Seabright, P. (2015). Honest signaling in trust interactions: smiles rated as gesnuine induce trust and signal higher earning opportunities. *Evol. Hum. Behav*. 36, 8–16. doi: 10.1016/j.evolhumbehav.2014.08.001

Chang, L. J., Doll, B. B., van't Wout, M., Frank, M. J., and Sanfey, A. G. (2010). Seeing is believing: trustworthiness as a dynamic belief. *Cogn. Psychol*. 61, 87–105. doi: 10.1016/j.cogpsych.2010.03.001

Charness, G., and Dufwenberg, M. (2006). Promises and partnership. *Econometrica* 74, 1579–1601. doi: 10.1111/j.1468-0262.2006.00719.x

Chovil, N., and Fridlund, A. J. (1991). Why emotionality cannot equal sociality: reply to buck. *J. Nonverb. Behav*. 15, 163–167. doi: 10.1007/BF01672218

Ekman, P. (2006). Darwin and facial expression: a century of research in review. *Malor Books*. An imprint of The Institute for the Study of Human Knowledge.

Farrell, J., and Rabin, M. (1996). Cheap talk. *J. Econ. Perspect*. 10, 103–118. doi: 10.1257/jep.10.3.103

Fetchenhauer, D., Groothuis, T., and Pradel, J. (2010). Not only states but traits-humans can identify permanent altruistic dispositions in 20 s. *Evol. Hum. Behav*. 31, 80–86. doi: 10.1016/j.evolhumbehav.2009.06.009

Fischbacher, U. (2007). z-tree: Zurich toolbox for ready-made economic experiments. *Exp. Econ*. 10, 171–178. doi: 10.1007/s10683-006-9159-4

Fischbacher, U., Gächter, S., and Fehr, E. (2001). Are people conditionally cooperative? evidence from a public goods experiment. *Econ. Lett*. 71, 397–404. doi: 10.1016/S0165-1765(01)00394-9

Frederick, S. (2005). Cognitive reflection and decision making. *J. Econ. Perspect*. 19, 25–42. doi: 10.1257/089533005775196732

Fridlund, A. J. (1994). *Human Facial Expression*. Boston, MA: Academic Press. doi: 10.1016/B978-0-12-267630-7.50012-5

Gneezy, U., and Potters, J. (1997). An experiment on risk taking and evaluation periods. *Q. J. Econ*. 112, 631–645. doi: 10.1162/003355397555217

Greene, W. (2010). Testing hypotheses about interaction terms in nonlinear models. *Econ. Lett*. 107, 291–296. doi: 10.1016/j.econlet.2010.02.014

Greiner, B. (2015). Subject pool recruitment procedures: organizing experiments with Orsee. *J. Econ. Sci. Assoc*. 1, 114–125. doi: 10.1007/s40881-015-0004-4

He, S., Offerman, T., and van de Ven, J. (2017). The sources of the communication gap. *Manage. Sci*. 63, 2832–2846. doi: 10.1287/mnsc.2016.2518

Houser, D., and Xiao, E. (2011). Classification of natural language messages using a coordination game. *Exp. Econ*. 14, 1–14. doi: 10.1007/s10683-010-9254-4

Jaeger, B., Oud, B., Williams, T., Krumhuber, E., Fehr, E., and Engelmann, J. (2020). Can people detect the trustworthiness of strangers based on their facial appearance. doi: 10.31234/osf.io/ayqeh

Kocher, M. G., Cherry, T., Kroll, S., Netzer, R. J., and Sutter, M. (2008). Conditional cooperation on three continents. *Econ. Lett*. 101, 175–178. doi: 10.1016/j.econlet.2008.07.015

Murphy, R. O., Ackermann, K. A., and Handgraaf, M. J. (2011). Measuring social value orientation. *Judgm. Decis. Mak*. 6, 771–781. doi: 10.2139/ssrn.1804189

Oda, R., Naganawa, T., Yamauchi, S., Yamagata, N., and Matsumoto-Oda, A. (2009a). Altruists are trusted based on non-verbal cues. *Biol. Lett*. 5, 752–754. doi: 10.1098/rsbl.2009.0332

Oda, R., Tainaka, T., Morishima, K., Kanematsu, N., Yamagata-Nakashima, N., and Hiraishi, K. (2021). How to detect altruists: Experiments using a zero-acquaintance video presentation paradigm. *J. Nonverb. Behav*. 31, 137–152. doi: 10.1007/s10919-020-00352-0

Oda, R., Yamagata, N., Yabiku, Y., and Matsumoto-Oda, A. (2009b). Altruism can be assessed correctly based on impression. *Hum. Nat*. 20, 331–341. doi: 10.1007/s12110-009-9070-8

Olivola, C. Y., Funk, F., and Todorov, A. (2014). Social attributions from faces bias human choices. *Trends Cogn. Sci*. 18, 566–570. doi: 10.1016/j.tics.2014.09.007

Rezlescu, C., Duchaine, B., Olivola, C. Y., and Chater, N. (2012). Unfakeable facial configurations affect strategic choices in trust games with or without information about past behavior. *PLoS ONE* 7:e34293. doi: 10.1371/journal.pone.0034293

Rodríguez-Ruiz, C., Sanchez-Pages, S., and Turiegano, E. (2019). The face of another: anonymity and facial symmetry influence cooperation in social dilemmas. *Evol. Hum. Behav*. 40, 126–132. doi: 10.1016/j.evolhumbehav.2018.09.002

Sally, D. (2000). A general theory of sympathy, mind-reading, and social interaction, with an application to the prisoners' dilemma. *Soc. Sci. Inform*. 39, 567–634. doi: 10.1177/053901800039004003

Schwartz, S., Spires, E., and Young, R. (2019). Why do people keep their promises? A further investigation. *Exp. Econ*. 22, 530–551. doi: 10.1007/s10683-018-9567-2

Servátka, M., Tucker, S., and Vadovič, R. (2011). Words speak louder than money. *J. Econ. Psychol*. 32, 700–709. doi: 10.1016/j.joep.2011.04.003

Stirrat, M., and Perrett, D. I. (2010). Valid facial cues to cooperation and trust: male facial width and trustworthiness. *Psychol. Sci*. 21, 349–354. doi: 10.1177/0956797610362647

Sylwester, K., Lyons, M., Buchanan, C., Nettle, D., and Roberts, G. (2012). The role of theory of mind in assessing cooperative intentions. *Pers. Indiv. Differ*. 52, 113–117. doi: 10.1016/j.paid.2011.09.005

Todorov, A., Funk, F., and Olivola, C. (2015a). Response to Bonnefon et al.: Limited kernels of truth in facial inferences. *Trends Cogn. Sci*. 19:422. doi: 10.1016/j.tics.2015.05.013

Todorov, A., Olivola, C. Y., Dotsch, R., and Mende-Siedlecki, P. (2015b). Social attributions from faces: determinants, consequences, accuracy, and functional significance. *Annu. Rev. Psychol*. 66, 519–545. doi: 10.1146/annurev-psych-113011-143831

Tognetti, A., Berticat, C., Raymond, M., and Faurie, C. (2013). Is cooperativeness readable in static facial features? An inter-cultural approach. *Evol. Hum. Behav*. 34, 427–432. doi: 10.1016/j.evolhumbehav.2013.08.002

Tognetti, A., Yamagata-Nakashima, N., Faurie, C., and Oda, R. (2018). Are non-verbal facial cues of altruism cross-culturally readable? *Pers. Indiv. Differ*. 127, 139–143. doi: 10.1016/j.paid.2018.02.007

Turmunkh, U., van den Assem, M. J., and Van Dolder, D. (2019). Malleable lies: communication and cooperation in a high stakes TV game show. *Manage. Sci*. 65, 4795–4812. doi: 10.1287/mnsc.2018.3159

Van den Assem, M. J., Van Dolder, D., and Thaler, R. H. (2012). Split or steal? cooperative behavior when the stakes are large. *Manage. Sci*. 58, 2–20. doi: 10.1287/mnsc.1110.1413

Van Leeuwen, B., Noussair, C. N., Offerman, T., Suetens, S., Van Veelen, M., and Van De Ven, J. (2018). Predictably angry-facial cues provide a credible signal of destructive behavior. *Manage. Sci*. 64, 3352–3364. doi: 10.1287/mnsc.2017.2727

Vogt, S., Efferson, C., and Fehr, E. (2013). Can we see inside? Predicting strategic behavior given limited information. *Evol. Hum. Behav*. 34, 258–264. doi: 10.1016/j.evolhumbehav.2013.03.003

Woike, J. K., and Kanngiesser, P. (2019). Most people keep their word rather than their money. *Open Mind* 3, 68–88. doi: 10.1162/opmi_a_00027

Zylbersztejn, A., Babutsidze, Z., and Hanaki, N. (2020). Preferences for observable information in a strategic setting: an experiment. *J. Econ. Behav. Organ.* 170, 268–285. doi: 10.1016/j.jebo.2019.12.009

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Psychology

ORIGINAL RESEARCH
published: 05 October 2021
doi: 10.3389/fpsyg.2021.743054

Long Term Effects of the COVID-19 Pandemic on Social Concerns

Esther Blanco[1,2]*, *Alexandra Baier*[1], *Felix Holzmeister*[3], *Tarek Jaber-Lopez*[4] and *Natalie Struwe*[1]

[1] Department of Public Finance, University of Innsbruck, Innsbruck, Austria, [2] The Ostrom Workshop, Indiana University, Bloomington, IL, United States, [3] Department of Economics, University of Innsbruck, Innsbruck, Austria, [4] Université Paris Nanterre, Nanterre, France

While some local, temporary past crises have boosted overall charitable donations, there have been concerns about potential substitution effects that the Covid-19 pandemic might have on other social objectives, such as tackling climate change and reducing inequality. We present results from a donation experiment ($n = 1,762$), with data collected between April 2020 and January 2021. We combine data from (i) an online donation experiment, (ii) an extended questionnaire including perceptions, actions, and motives on the Covid-19 pandemic, the climate crisis, and poverty, as well as charitable behavior and (iii) epidemiological data. The experimental results show that donations to diverse social concerns are partially substituted by donations to the Covid-19 fund; yet, this substitution does not fully replace all other social concerns. Over time we observe no systematic trend in charitable donations. In regards to the determinants of individual donations, we observe that women donate more, people taking actions against Covid-19 and against poverty donate more, while those fearing risks from poverty donate less. In addition, we observe that the population under consideration is sensitive to the needs of others, enhancing total donations for higher Covid-19 incidence. For donations to each charity, we find that trusting a given charitable organization is the strongest explanatory factor of donations.

JEL: L3, D64, Q54, I3, D9

Keywords: charitable donation, COVID-19 pandemic, climate crisis, poverty, substitution of social concerns

OPEN ACCESS

Edited by:
Lena Rademacher, University of Lübeck, Germany

Reviewed by:
Noemí Herranz-Zarzoso, University of Jaume I, Spain
Gerardo Sabater-Grande, University of Jaume I, Spain

***Correspondence:**
Esther Blanco
esther.blanco@uibk.ac.at

Specialty section:
This article was submitted to Personality and Social Psychology, a section of the journal Frontiers in Psychology

Received: *17 July 2021*
Accepted: *31 August 2021*
Published: *05 October 2021*

Citation:
Blanco E, Baier A, Holzmeister F, Jaber-Lopez T and Struwe N (2021) Long Term Effects of the COVID-19 Pandemic on Social Concerns. Front. Psychol. 12:743054. doi: 10.3389/fpsyg.2021.743054

1. INTRODUCTION

Understanding the drivers of human behavior is essential when facing global shocks such as the Covid-19 pandemic. Together with governmental actions and recommendations, the behavioral responses of citizens have shown to be a key variable in shaping the evolution of the collective action problem that the pandemic represents. A large body of literature has been dealing with the striking psychological consequences of the lockdown due to Covid-19 (see Salari et al., 2020, for an overview). Similarly, behavioral scientists have been tracking the evolution of social preferences and their correlation with health behaviors during the pandemic (see section 2 for a review). This study contributes to the literature addressing the long term (10 months) impact of the pandemic on social preferences by investigating the substitution effects in social concerns with respect to Covid-19, the climate crisis, and poverty alleviation. To the best of our knowledge, this is the first study to analyze whether the pandemic affects the social priorities during an extended time period.

Understanding the substitution in social concerns associated with the pandemic is critical in designing recovery policies. The relative weights of social concerns can affect the social acceptability of policies to "build back better." Next to the substantial impact on individuals' daily lives and health conditions brought about by the Covid-19 pandemic, there are further pressing social objectives affecting human well-being, such as alleviating global poverty, addressing the climate crisis, and promoting environmental conservation (featured in the United Nations Sustainable Development Goals; SDGS). Importantly, these objectives are interrelated with the Covid-19 pandemic. The "Covid-19 Response" to each of the SDGs (United Nations, 2020) and the report by the World Wide Fund for Nature (WWF, Jeffries B., 2020) illustrate the complex interrelations between health, poverty, and environmental conservation. But these complex interrelationships might be difficult to perceive for citizens who since the beginning of the pandemic have been facing increased stress, burdens in their daily lives, and new economic challenges. This may translate into focusing on the pandemic at the expense of other pressing issues, substituting the relevance of previous social concerns. The apprehension of such substitution effects in social concerns induced by the Covid-19 has been stressed by researchers (see, e.g., Hodges and Jackson, 2020; Naidoo and Fisher, 2020), Think Tanks (see, e.g., Zhongming et al., 2020), and political leaders (such as those of the European Union (*EU*) early on). For example, Rosenbloom and Markard (2020) have raised the concern that the Covid-19 response and recovery could affect the mitigation of the climate crisis and the continuation of the *Intergovernmental Panel on Climate Change* (*IPCC*) report (Tollefson, 2020). In addition, Mahler et al. (2020) estimate that the Covid-19 pandemic might push about 40–60 million people into extreme poverty. Furthermore, a common concern of scientists, governments, and supra-national agencies is that the pandemic might induce a financial crisis amplifying inequality and severe poverty (von Braun et al., 2020).

Within this context, we present evidence on the long-term substitution effects that the Covid-19 pandemic might have on other social priorities by means of real-life donations to charities. We collected weekly data for 8 weeks and monthly data for 8 months between April 2020 and January 2021. Our results respond to the call by the scientific community for economists to contribute to the understanding of the behavioral effects of the Covid-19 pandemic (Coyle, 2020), contributing to the efforts by the economics discipline to generate cumulative evidence aiding policy-making (see https://bit.ly/3jmBZk3). We study if and how Covid-19 concerns substituted donations to other social concerns, how substitution effects evolved over time, and the determinants of donation behavior during the pandemic.

We combine results from (i) an online donation experiment with more than 1,700 students, (ii) an extended questionnaire, and (iii) epidemiological data. In the online donation experiment, subjects are endowed with €3 that can be distributed between themselves and a list of charitable organizations which vary between treatments. In a *Baseline* setting, the list of possible recipients comprises eight charities representing diverse social concerns. To measure potential substitution effects in donations between various social concerns in light of the Covid-19 pandemic, in a *COVID-19* treatment we include the *COVID-19 Solidarity Response Fund for WHO* (*WHO Covid-19 Fund*; see https://bit.ly/3wiwJDU for details about the fund) in addition to these eight charities as a possible recipient for donations. Finally, in a *Covid-19 Only* treatment we include only the *WHO Covid-19 Fund* as a possible recipient[1]. After the donation task, participants answer an extensive questionnaire including subjects' socio-demographic characteristics; subjects' perception of how relevant a charity's work is regarding alleviating the consequences of the Covid-19 pandemic, their national or international operation, and their trustworthiness; participants' risk perceptions, actions, and motivations regarding the Covid-19 pandemic, the climate crisis, and poverty, respectively; as well as subjects' history of donation and voluntary work for charities. Our pre-registered initial theses (see pre-registration at https://aspredicted.org/3g8sd.pdf) are (i) that the Covid-19 pandemic substitutes other social concerns, (ii) that the distribution of donations changes over time with the intensity of the crisis, and (iii) that donations correlate with risk perceptions, actions, and motives at the individual level. The controlled experiment that we present allows us to test these conjectures.

As compared to previous studies focusing on aggregate levels of charitable donations to single charities (see, e.g., Andreoni, 1990; Vesterlund, 2003; Frey and Meier, 2004; Bénabou and Tirole, 2006; Ariely et al., 2009; Gneezy et al., 2014; Garcia et al., 2020) or alternative charities with the same social objectives (see, among others, Soyer and Hogarth, 2011; Schmitz, 2021), we intentionally incorporate charities that cover a wide range of social priorities (as in Eckel and Grossman, 2003; Crumpler and Grossman, 2008; Brown et al., 2017). Our study is closer to previous contributions to the literature focusing on how negative shocks on individuals' health or natural disasters affect donations to charities working on related social objectives and to charities working on other social objectives (see section 2). Blanco et al. (2020) is the only previous experimental study looking at the effect of the pandemic on relative social priorities, reporting short time effects for 2 months. As compared to this study, we incorporate two main novelties: First, we present evidence for 10 months, providing evidence on long-term substitution of social objectives for the first time. Second, we provide a broad analysis on the individual determinants of donations to charities during the pandemic. The rich database collected through the questionnaire provides insights into the factors shaping human behavior in the context of the pandemic. We also incorporate the evolution of the epidemiological situation in the analysis of the determinants of donations (similar to other studies in this field of research, e.g., Abel et al., 2020; Branas-Garza et al., 2020; Lohmann et al., 2020).

Our findings suggest a long-term substitution effect due to the Covid-19 pandemic, as has been anticipated by policy makers. This result is derived from two observations: On the

[1]Please see section 3 for a discussion of the methodological implications of changing the number of possible recipient charities and their relevance in the results.

one hand, we observe substantial donations to the *WHO Covid-19 Fund*. On the other hand, participants do not change their aggregate donations depending on whether the *WHO Covid-19 Fund* is a possible recipient. These findings suggest that people react to the context and adapt their donation behavior to the broader set of calls for donations, but the aggregate social concern (altruism) is not reduced by the pandemic. The latter represents additional evidence on the mixed results in the literature with respect to the impact of the Covid-19 on social preferences (see section 2). Notably, we do not observe systematic trends in donations over time, which is possibly driven by the pandemic having extended over a longer time period than initial forecasts suggested. With the 10 months data collection in our analysis we have not, unfortunately, reached the post-pandemic period. We observe that systematic predictors of donation for the pooled data are the 7-day incidence of Covid-19 infections, self-reported individual Covid-19 actions, and participants' gender, with women donating significantly more than men, the latter being in line with previous literature (Eckel and Grossman, 2001, 2003; Eckel et al., 2005). When analyzing each organization separately, we find that trusting the corresponding charity is the most significant predictor of donations to the respective charity. This is in line with the emphasis of Ostrom (1990) on the relevance of trust as a precondition to successfully overcome collective action challenges.

2. RELATED LITERATURE

Psychologists have devoted much effort during the Covid-19 pandemic to track the consequences of health regulations on psychological well-being (see Salari et al., 2020). At early stages of the Covid-19 pandemic, increased levels of depression, stress, and anxiety were reported for different populations (see Cao et al., 2020; Zhou et al., 2020 for college students, Wang and Zhao, 2020 for university students, and Zhang et al., 2020 for working adults in China; see Odriozola-Gonzalez et al., 2020a,b; Planchuelo-Gomez et al., 2020; Rodriguez-Rey et al., 2020 for evidence from different populations in Spain). Large scale studies have analyzed the early psychological responses to the pandemic, including concern and stress, and associated public behavior in 48 countries (Lieberoth et al., 2021). There is evidence that negative psychological effects endure for longer time periods (see, e.g., Gonzalez-Sanguino et al., 2020; Roma et al., 2020). More generally, life satisfaction (in a sample of Spanish adults) positively correlates with hope about overcoming the pandemic and negatively correlates with social phobia (Blasco-Belled et al., 2020). In addition, daily life satisfaction and the length of lockdown periods are positively correlated (Sabater-Grande et al., 2021).

Behavioral scientists have concurrently tracked the pro-social concerns during the pandemic. That is, the extent to which people care about others' well-being. Neoclassical economics commonly conceives individuals as purely self-interested decision makers, maximizing individual payoffs. Building on empirical evidence showing that individuals are similarly motivated by other-regarding preferences, such as altruism and inequality aversion supports a broader view on subjects' social preferences. In this study we focus on pro-social concerns (see Andreoni, 1989; Meier, 2007; Chaudhuri, 2011, for reviews on pro-social behavior). One way to elicit pro-social behavior is to look at donation decisions of individuals to charities. Next to looking at donation data of households from national statistics or survey measures, people's social concerns can be measured by means of experimental methods (see Levitt and List, 2007 for an overview of games used in experimental economics). A common approach to experimentally elicit prosocialty is to ask participants to decide on how to distribute a given amount of money between themselves and a charity recipient of their choice (Andreoni, 1990; Eckel and Grossman, 1996). Previous studies have identified factors systematically affecting subjects' pro-social behavior: Donation levels vary with the individual characteristics of donors (Eckel and Grossman, 2001) and the institutional context (Frey and Meier, 2004; Garcia et al., 2020), including whether there are market interactions (see, e.g., Bartling et al., 2015; Kirchler et al., 2016).

The stability of social concerns is a controversial topic. While several models characterize people as belonging to certain preference types (in the sense of latent traits), there is growing evidence that social concerns can be context-dependent, time-dependent, and vary with the experience of people in life. Individuals' pro-social preferences measured via experiments or surveys change over time due to factors like education interventions (Jakiela et al., 2015), economic shocks (Fisman et al., 2015), or natural disasters and violence (Voors et al., 2012; Cassar et al., 2017). Empirical studies using donation statistics show that fundraising interventions for natural and humanitarian disasters foster donations to charities related to the disaster, and increase donations to unrelated causes (Brown et al., 2012), but the effect on other charities fades out over time (Scharf et al., 2017). Importantly, such donations appeals have shown not to reduce donations to other (unrelated) causes (Deryugina and Marx, 2021). More specifically, Brown et al. (2012) show that unexpected donations of households after the 2004 Indian Ocean tsunami were positively correlated with planned (future) donations toward other social causes. Scharf et al. (2017) find that fundraising interventions associated with a natural or human disaster lift donations to charities related to the disaster, and donations to other (unrelated) charities for a short time but decline shortly thereafter, leading to no changes in baseline donation levels to the other charities in the longer time horizon. Similarly, Deryugina and Marx (2021) identify that an exogenous increase in demand for giving (due to tornadoes) does not reduce donations to other local charities. Thus, Deryugina and Marx (2021) conclude that "giving to one cause need not come at the expense of another." An additional line of literature addresses the question whether and how experiencing a crisis affects peoples' pro-social behavior. For example, experiencing a natural disaster has been shown to reduce donations to related causes (Eckel et al., 2007); experiencing an adverse health shock (e.g., stroke, heart attack, cancer), however, substitutes donations to other social concerns toward health-related charities (Black et al., 2020). Our take from these studies measuring the effect of experience during a crisis is that the impact could be context dependent, and

thus reinforces the need for specific research conducted for the Covid-19 pandemic.

Recent research on the effects of the Covid-19 pandemic on social preferences reports intertemporal stability of risk and time preferences (Drichoutis and Nayga, 2020) and a negative effect on generosity measured by donations in an online experiment (Branas-Garza et al., 2020). A study with students from Wuhan during the pandemic finds positive trends in altruism, trust, and risk tolerance (Shachat et al., 2020). Subjects in China that were more intensively exposed to the Covid-19 crisis reveal more anti-social behavior than those with lower exposure (Lohmann et al., 2020). Li et al. (2021) conducted an online experiment to examine the contagion of others' positive and negative donation behavior of the Covid-19 pandemic in China during and after the peak. They also investigated the impact of social anxiety on the link between the contagion of donation behaviors and the changes in the Covid-19 situation. Their results show that increased or decreased donation amounts given by other participants lead to positive or negative donation behavior, respectively. Moreover, participants' social anxiety decreased with the ease of the pandemic, and social anxiety in turn mediated the relationship between the pandemic abatement and the decrease in the contagion of positive donation behaviors.

Similarly, recent studies address how experience with the Covid-19 pandemic (Branas-Garza et al., 2020; Shachat et al., 2020) or information policies on Covid-19 affect people's pro-social behavior and pro-conservation policy support (Abel and Brown, 2020; Abel et al., 2020; Guo et al., 2020; Shreedhar and Mourato, 2020). Other studies have addressed, more broadly, the interconnections between the Covid-19 pandemic, economic well-being, and environmental conservation (see, e.g., Dobson et al., 2020; Goldthau and Hughes, 2020). Although a negative income shock due to the pandemic might decrease pro-social behavior (Almunia et al., 2020), previous evidence suggests that a collective threat can enhance cooperation, pro-social behavior, and trust (Li et al., 2020). Examining social preferences in the time of a pandemic is of special interest, as measures of social preferences have been found to correlate with health behavior. For example, people who are more pro-social are also more likely to follow hygiene recommendations to fight the pandemic (Campos-Mercade et al., 2021).

This study is a follow-up study of Blanco et al. (2020), which is—to the best of our knowledge—the first to report evidence on substitution effects between social concerns in the Covid-19 context. Blanco et al. (2020) investigate short-time changes in social concerns at the onset of the pandemic[2]. The results of Blanco et al. (2020) show a partial substitution of donations to a Covid-19-related fund at the expense of donations to other social concerns on the short run. The follow-up study presented herein is novel in two aspects: First, we examine long-time trends in social concerns, reporting data over 10 months. Second, we explore a wide set of determinants that might influence the donation behavior during the pandemic.

This study also contributes to a strand of literature investigating competition among charities, including studies using lab and field experiments. There is evidence from laboratory experiments that the total amount of charitable giving varies when changing the number of charities or campaigns (Reinstein, 2007; Soyer and Hogarth, 2011; Deck and Murphy, 2019; Schmitz, 2021), when the number of potential charities is uncertain (Eckel et al., 2020). Specifically, studies show that increasing the number of charities with similar objectives that are possible recipients increases total contributions (Soyer and Hogarth, 2011; Schmitz, 2021). Schmitz (2021) increases the list of charities from one single charity up to three and finds a weak substitution with more recipients but no changes in the overall donation amount. Soyer and Hogarth (2011) investigate competition among charities with up to 16 possible recipients. They show that the total amount of donations increases with more recipients but at a decreasing rate. There is also field evidence pointing in the same direction: A solicitation of volunteering by two charities results in increased time donations to each charity as compared to people solicited by a single charity to volunteer (Lange and Stocking, 2012). Lange and Stocking (2012) also show that subjects solicited to volunteer by two charities gave higher total monetary donations to the sum of charities than they gave when they were solicited by only one charity[3].

In sum, the evidence from the studies discussed above suggests that donations to unexpected events caused by crises do not necessarily come at the expense of donations to other charities. When people have experienced the respective events themselves, the results seem to be context dependent: There is evidence that having experienced a health shock can generate a shift in donations, leading to a substitution toward donations to health related charities at the expense of donations toward other social concerns. This calls for specific results referring to the effects of the Covid-19 pandemic. While there is a wide literature on the effect of the pandemic on psychological well-being and social preferences, there is no study investigating the long-term substitution effects on social preferences that we address in this paper. Lastly, from a methodological perspective, the previous literature suggests that increasing the number of possible recipients increases total donations.

[2]Please note that the current study includes the weekly data from April and May 2020 on which the results in Blanco et al. (2020) are based on, and uses the same treatment variations (see section 3) as the initial study. The subsequent monthly data from June 2020 to January 2021 is reported for the first time in this paper, as are the results related to several of the questionnaire items. Incorporating the relevance of the Covid-19 incidence rate on donation behavior is also a novel aspect of the current study.

[3]A related line of research looking at competition between charities examines the effect of targeting one charity out of a list of potential recipients on donations to other charities. Applied mechanisms are information priming (Harwell et al., 2015; Filiz-Ozbay and Uler, 2019) or different incentives through matching donations (i.e., the experimentalist adds a fix rate to each donation made by participants) (Gallier et al., 2019; Schmitz, 2021). Applying a matching to one charity out of a set does not generate substitution in donations between charities with similar objectives. Schmitz (2021) finds that the matching does not change total net donations to all charities. Gallier et al. (2019) observes increases in net donations to the matched charity and to other similar charities. These studies, however, differ substantially from our design, as we keep constant the incentives to donate to each of the recipient.

3. EXPERIMENTAL DESIGN

We implement three different treatment conditions, where each of the three treatments consists of a incentivized donation-to-charity task, similar to Eckel and Grossman (2003) and Eckel et al. (2005), followed by a questionnaire. In the donation task, subjects were endowed with €3 to be distributed among themselves and various charitable organizations, freely deciding how much to allocate to each charity and to themselves, if any. The list of available charities varied between treatments.

In the *Baseline* treatment, the list of charitable organizations included eight charities, namely *World Wide Fund for Nature* (*WWF*), *Doctors Without Borders* (*MSF*), *Amnesty International* (*AI*), *SOS Kinderdorf* (*SOS*), *Caritas* (*CAR*), *Licht ins Dunkel* (*LID*), *Oxfam* (*OXF*), and the *Red Cross* (*RC*). This list was chosen to reflect a broad range of social concerns. In the *COVID-19* treatment, the *WHO Covid-19 Fund* was added to the list of charitable organizations used in *Baseline*, leading to a total of nine charities. In *Covid-19 Only*, the *WHO Covid-19 Fund* was the only available recipient[4]. In all treatments the decision screen included the mission statement of each of the charities. In the *Baseline* and *COVID-19* treatments, participants could distribute their endowment across multiple charities, if any, and themselves. In all treatments, donations were matched at a rate of 25%, i.e., we donated an additional 25% to all donations made by participants. This mechanism ensures that is is socially efficient for the participants to make donations via the donation task that we offer, as opposed to keeping the full endowment themselves and making donations to their preferred charities outside of the experiment. The individual earnings of the experiment are defined by the amount (of the €3 endowment) that subjects kept for themselves. The instructions of the experiment are presented in section A of the **Supplementary Material**[5].

After completing the donation task, subjects answered a questionnaire containing subjects' socio-demographic characteristics; subjects' perception of how relevant a charity's work is regarding alleviating the consequences of the Covid-19 pandemic, the perception of the national vs. international assistance offered by the charity, and their trustworthiness; participants' risk perceptions, actions, and motives regarding the Covid-19 pandemic, the climate crisis, and poverty, respectively; as well as subjects' history of donation and work for charities (see section B of the **Supplementary Material** for the detailed survey questions). Survey items on risk perceptions, actions, and motives are *z*-standardized (across all three treatments in the main experiment). The measures used in the analyses are constructed as the sum of the standardized responses of the items belonging to the particular inventory; this measure is finally *z*-standardized again, such that all measures used in the analyses have a mean of zero and a standard deviation of one. Likewise, survey responses on participants' trust in the charities, their perceived relevance of the charities' work during the pandemic, and the perceived help of the charities during the pandemic are *z*-standardized.

Experimental Procedures

A total of 1,762 subjects (*Baseline*: $n = 581$; *COVID-19*: $n = 599$; and *Covid-19 Only*: $n = 582$) were recruited from the standard student subject pool of the University of Innsbruck, Tyrol (Austria) using *hroot* (Bock et al., 2014). As part of western Austria, Tyrol was among the regions that were worst affected by the first wave of the Covid-19 pandemic in Austria, bordering the North of Italy and the South of Germany. The region reported the first cases on February 25, 2020 and entered a lock-down of all municipalities in the region for about 7 weeks on March 16, 2020. During the period of the data collection (April 2020–January 2021), the number of cases in Tyrol varied substantially. During this period, the 7-day incidence varied between 771.3 and 0.1, with peaking values during the months of November and December 2020 and lowest values for May and June 2020 (https://bit.ly/2SOSsFM). Thus, the study covers a period of time with substantial variance in terms of the severity of the Covid-19 pandemic.

We ran the experiments online. Subjects only participated in one of the treatment conditions in a between-subjects design and could only participate once. For each date at which data was collected, invitations were made for three simultaneously running sessions, one for each treatment condition, with up to 40 participants in each treatment, leading to a total number of 16 sessions. Upon receiving the invitation, subjects were informed that this was an online experiment that would last approx. 20 min. As payment options we offered transactions via PayPal or in the form of Amazon gift cards.

We collected data in two different intervals. First, in April and May 2020, we collected weekly data on 1 day of each week, for a total of eight consecutive weeks. Thereafter, starting in June 2020, we collected monthly data on 1 day of each month for a total of eight consecutive months. Subjects were told that they could participate in the experiment as soon as they received the link which was distributed at 10 a.m., and that participation was possible until 8 p.m. on the same day; at 8 p.m., the experimental sessions would be closed and the links would be deactivated.

At the end of each experimental session, the sum of donations across all treatments was transferred to each of the organizations via bank transfers, including a matching payment of 25%. A depersonalized summary of all individual donations as well as the total amount of money paid to each organization was made available on the website of the corresponding author after each experimental session. The payment to participants was transferred within three working days by one of the co-authors.

4. RESULTS

The presentation of results is organized in two subsections. First, we focus on average treatment effects. We observe that introducing the *WHO* Covid-19 Solidarity Response fund

[4]Please see section 5 for a discussion of the implications of the methodological aspects of this design on the results. Specifically, addressing how the change in the number of recipients could affect the findings.

[5]Note that in addition to the treatments described here, we conducted a series of robustness sessions, including a 10 fold increase in endowments, with subjects making decisions over €30 in all three decision settings, and additional robustness tests. The results are reported in (Blanco et al., 2020). The observed treatment effects are robust to these changes.

significantly reduces the average sum of donations to the original eight charities. When looking at the evolution of donations over time, we do do not observe general systematic trends in donations to the different treatment conditions.

Next, we analyze the determinants of individual donations using the data from the post-experiment questionnaire as well as epidemiological data. Our main results show that systematic predictors of total donations are the epidemiological situation, gender, previous charity donations, as well as self-reported Covid-19 actions, poverty risk perceptions and actions. Further, while the 7-day incidence rates and the self-reported Covid-19 risk perceptions do correlate, the epidemiological situation does not significantly explain donations to the *WHO Covid-19 Fund*.

In the following, we will distinguish between (i) the average total donations (*avg. total*) pooled across all charities available as recipients in the respective treatment; (ii) the average sum of donations to the original eight charities in *Baseline* (and thus a subset of the charities in *COVID-19*, excluding the *WHO Covid-19 Fund*; *avg. sum-8*); and finally (iii) the donations to the *WHO Covid-19 Fund* in *COVID-19* or *Covid-19 Only* (*avg. WHO donations*).

Treatment Effects

When considering the pooled donation data from April 2020 to January 2021, we observe a substitution of social concerns under the presence of the *WHO Covid-19 Fund*. In the *COVID-19* treatment, the average donation to the eight charities is 68.8% of the endowment ($m = €2.06$, $sd = €1.02$), which is significantly lower than the average donation of 78.1% of the endowment in *Baseline* [$m = €2.34$, $sd = €1.04$; $t_{(1,178)} = 5.851$, $p < 0.001$, $n = 1,180$; see **Figure 1B**]. Moreover, a Komogorov-Smirnov for the equality of the distribution functions of *avg. sum-8* between *Baseline* and *COVID-19* indicates that the distributions of donations to the eight charities differ systematically between the two treatments ($D = 0.236$, $p < 0.001$; $n = 1,180$).

This substitution is the sum of consistent but small substitutions for each individual charity. The *avg. sum-8* is smaller in *COVID-19* than in *Baseline* for all charities (negative estimates in **Figure 1B**), despite these differences being only statistically significant for *OXF* [$t_{(1,178)} = 2.559$, $p = 0.011$; $n = 1,180$] and *WWF* [$t_{(1,178)} = 2.119$, $p = 0.034$; $n = 1,180$]. Moreover, all charities are similarly affected by the presence of the *WHO Covid-19 Fund*. In particular, we do not observe significant differences in substitution effects between the different charities, with the only exception being a marginally stronger reduction in donations to *Oxfam* as compared to the reduction in donations to *Amnesty International* [$\chi^2(1) = 3.949$, $p = 0.047$].

The differences in *avg. sum-8* between *COVID-19* and *Baseline* are not due to a change in the share of participants giving to any of the eight organizations: Pooled across the eight charities, 89.50% and 89.48% of participants choose to donate a positive amount of their endowment in *Baseline* and *COVID-19*, respectively [Pearson's χ^2-test: $\chi^2(1) < 0.001$, $p = 0.992$]. The average differences result from the fact that those who donate to any of the eight charities, indeed donate significantly lower amounts in *COVID-19* ($m = 2.31$, $sd = 0.69$) as compared to *Baseline* [$m = 2.62$, $sd = 0.69$; $t_{(1,054)} = 7.982$, $p < 0.001$]. On the charity level, the proportion of participants giving any positive amount, jointly with the amount given by those who donate, separated by treatments are shown in **Table 1**. While the share of participants donating to charity vary substantially between charities, differences between treatments are not significant, except for the *proportion* of participants donating to *OXF*, which is—as compared to *Baseline*—significantly smaller in *COVID-19* [Pearson's χ^2-test: $\chi^2(1) = 4.793$, $p = 0.029$]. Similarly, the *average amount* donated (by those participants who give a positive amount) does not significantly differ between treatments *Baseline* and *COVID-19* for any of the charities, except for the comparison regarding donations to *WWF* [$t_{(583)} = 2.246$, $p = 0.025$].

The main result on the substitution of social concerns related to the presence of the *WHO Covid-19 Fund* derives from two observations. First, the *avg. WHO donations* are substantial (*Observation 1*). Second, *avg. total* donations do not significantly differ between *Baseline* and *COVID-19*, introducing the *WHO Covid-19 Fund* (*Observation 2*).

The first observation is based on the finding that in the *COVID-19* treatment, with the list of nine charities, donations to *WHO Covid-19 Fund* amount to 8.0% of the endowment ($m = €0.24$, $sd = €0.48$). In particular, the donations to *WHO Covid-19 Fund* significantly exceed the donations to three out of the eight charities (*CAR*, *LID*, and *OXF*); for two more charities, donations do not significantly differ from donations to the *WHO Covid-19 Fund* (*SOS*, and *RC*). Moreover, when participants can only decide between donating to a Covid-19 charitable organization or keeping money for themselves (*Covid-19 Only*), donations to the *WHO Covid-19 Fund* amount to 53.3% of the endowment ($m = €1.60$, $sd = €1.12$; see **Figure 1A**).

With respect to the second observation, we do not find evidence for differences in *avg. total* donations between *Baseline* and *COVID-19*. The average total donations in *Baseline* are 78.1% of the endowment ($m = €2.34$, $sd = €1.04$; see **Figure 1A**). The aggregate donations to the full set of nine charities in the *COVID-19* treatment is slightly lower, at 76.9% of the endowment ($m = €2.31$, $sd = €1.02$), with the difference not being statistically significant [$t_{(1,178)} = 0.930$, $p = 0.353$, $n = 1,180$; see **Figure 1A**]. Thus, despite having more possible recipients, the donations do not increase in the *COVID-19* treatment.

Figure 2 displays the evolution of the *avg. sum-8* donations in *Baseline* and *COVID-19* over time. Looking at the figure we do not observe a systematic time trend in either of the two treatments. Despite the extended time under consideration (10 months) and the convulsive social situation during this period, the donation to the initial list of eight charities fluctuates up to 30 percent of the value without a clear time trend. Second, we do not observe a clear time trend in treatment effects. We observe that in five out of the first 8 weeks (April to May 2020) the *avg. sum-8* donations in *Baseline* are significantly above those in *COVID-19*. Over the summer of 2020, the difference in donations to the original eight charities in *Baseline* and *COVID-19* disappears, and returns only in October, the month that led to the beginning of the *second wave* in Austria, at a level that is comparable to that of early April 2020. Finally, in December and January the difference vanishes again. These results are consistent with the

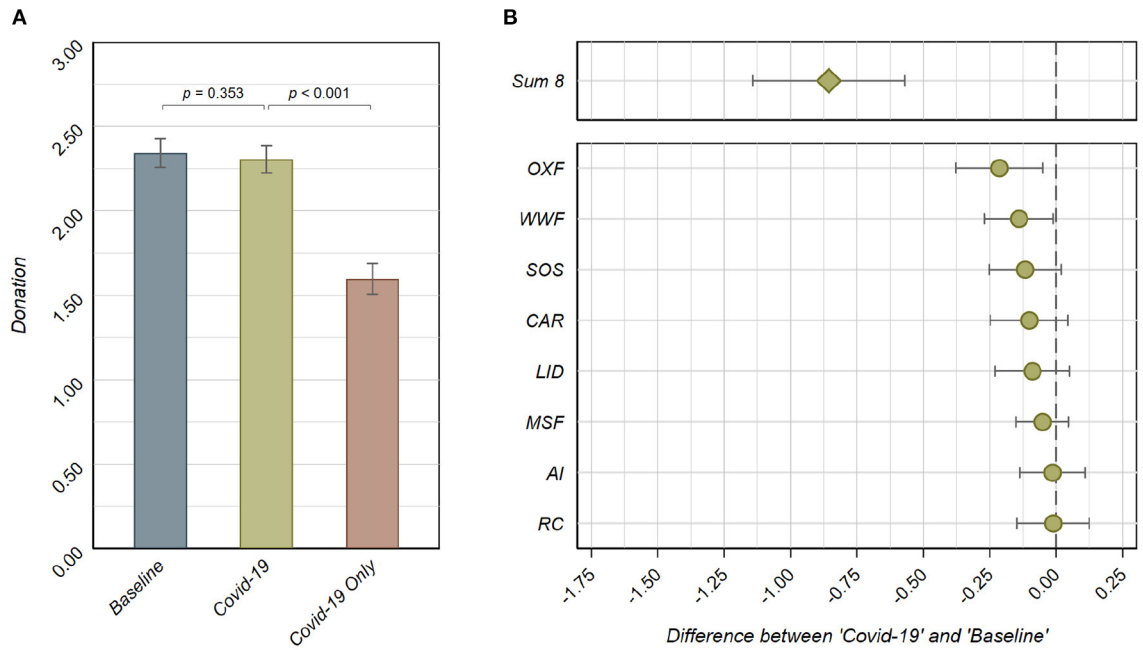

FIGURE 1 | (A) *avg. total* donations (pooled across charities) per treatment in €. *p*-values are based on Tobit regressions with €0 and €3 as the lower and upper limit, respectively (endowment €3), and robust standard errors. **(B)** Point estimates and 95% confidence intervals of the differences in *avg. sum-8* donations between the *Baseline* and the *COVID-19* treatment, based on Tobit regressions of the amount donated to the respective charitable organization on a treatment indicator for the *COVID-19* treatment (with €0 and €3 as the lower and upper limit, respectively, and robust standard errors). Negative values represent lower donations in the *COVID-19* treatment than the *Baseline* treatment. All pairwise comparisons between coefficients based on Wald tests after seemingly unrelated regressions (with robust standard errors) are insignificant, except for *OXF–AI* ($\chi^2(1) = 3.949$, $p = 0.047$). The estimate at the top indicates the difference in the sum of donations to the eight charitable organizations between the *Baseline* and the *COVID-19* treatment ($t(1178) = 5.851$, $p < 0.001$; $n = 1,180$).

TABLE 1 | Share of participants donating any positive amount, and average amounts donated by those who donate, separated by charities and treatments.

	Share of donors				Avg. Amount donated			
Charity	**Baseline (%)**	**COVID-19 (%)**	**$\chi^2(1)$**	***p*-value**	**Baseline (%)**	**COVID-19 (%)**	***t*-value**	***p*-value**
WWF	51.5	47.8	1.630	0.202	0.90	0.80	2.246	0.025
MSF	65.6	62.4	1.261	0.261	0.88	0.87	0.132	0.895
SOS	35.1	31.6	1.682	0.195	0.72	0.65	1.452	0.147
AI	44.1	44.4	0.014	0.905	0.79	0.75	0.932	0.352
CAR	24.1	20.0	2.834	0.092	0.54	0.59	0.852	0.395
LID	27.5	25.2	0.825	0.364	0.63	0.56	1.291	0.198
OXF	21.3	16.4	4.793	0.029	0.60	0.49	1.898	0.059
RC	36.1	36.7	0.043	0.835	0.76	0.71	1.050	0.295

χ^2 statistics and the corresponding p-values are reported for treatment comparisons of the share of participants; for comparisons of the average amount donated by those who donate between treatments, the t-statistics and p-values are obtained from Tobit regressions of the amount donated to the respective charitable organization on a treatment indicator for the COVID-19 treatment (with €0 and €3 as the lower and upper limit, respectively, and robust standard errors).

substitution being stronger when the epidemiological situation worsen. But these estimates need to be taken very carefully, as they are based on reduced sub-samples for each time period (roughly 40 observations per treatment each).

Supplementary Figures S2, S3 show the time evolution with respect to *Observation 1* on *avg. WHO donations* and *Observation 2* on *avg. total*, respectively. Generally, for both observations we do not find evidence for systematic variation over time. **Figure 2** shows that over time the *avg. WHO donations* in *COVID-19* and *Covid-19 Only* remain above zero throughout all 10 months in both treatments. When the *WHO Covid-19 Fund* is the only possible recipient (*Covid-19 Only*), we observe high variability during the first weeks of Spring 2020 followed by a mild increasing trend after August 2020. When the *WHO Covid-19 Fund* is one of the possible alternative recipients, we do not observe such an evolution. Indeed, the Spearman correlation between donations to the *WHO Covid-19 Fund* in treatments *COVID-19* and *Covid-19 Only* over time turns out to be close to

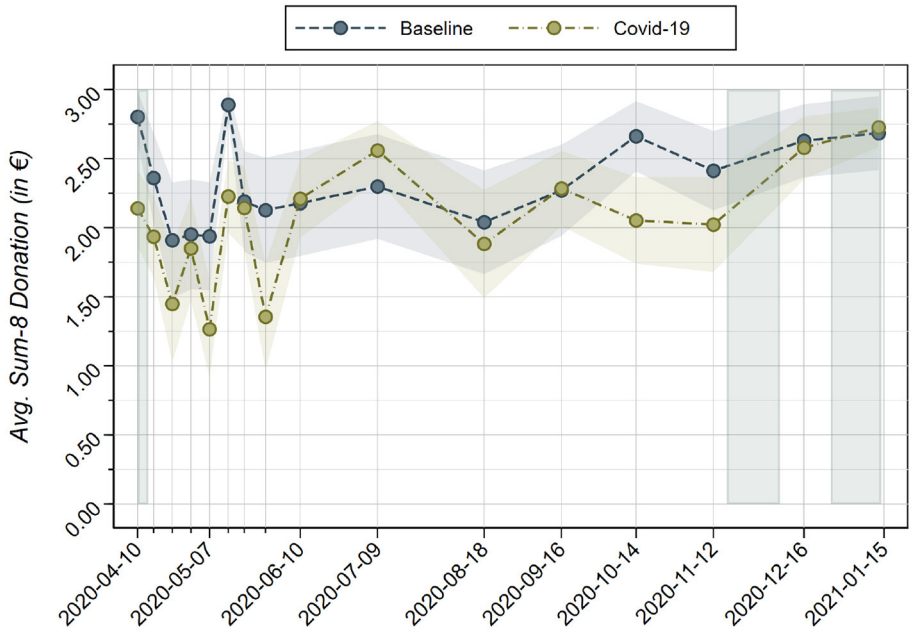

FIGURE 2 | Evolution of *avg. sum-8* donations (in €) in *Baseline* and *COVID-19* per treatment over the eight consecutive weeks plus eight consecutive months of data collection. Shaded areas indicate 95% confidence intervals. Vertically shaded areas indicate lockdown periods. The differences (based on Tobit regressions of *avg. sum-8* on a treatment indicator, with €0 and €3 as the lower and upper limit, respectively, and robust standard errors) between treatments *Baseline* and *COVID-19* are insignificant for each date, except for 2020-04-10 [$t_{(75)} = 4.094$, $p < 0.001$], 2020-04-16 [$t_{(75)} = 2.202$, $p = 0.031$], 2020-05-07 [$t_{(69)} = 2.501$, $p = 0.015$], 2020-05-14 [$t_{(83)} = 3.805$, $p < 0.001$], 2020-05-28 [$t_{(68)} = 2.644$, $p = 0.010$], and 2020-10-14 [$t_{(73)} = 3.343$, $p = 0.001$].

zero ($\rho_s = 0.021$, $p = 0.940$; $n = 16$). Finally, we observe that in October 2020 and November 2020, right before the second lock down in Austria, donations to *WHO Covid-19 Fund* in *COVID-19* are lowest; participants seem to prioritize other social concerns at that time.

Further, **Figure 3** shows the variation in *avg. total* donations over time in all three treatment conditions, as related to *Observation 2*. Remarkably, *avg. total* donations in *Baseline* do not significantly differ from *COVID-19*, except for May 14, 2020 [$t_{(83)} = 2.192$, $p = 0.031$] and October 2020 [$t_{(73)} = 2.909$, $p = 0.005$]. Aggregate pro-social concerns seem to be consistently unaffected by the presence or absence of the *WHO Covid-19 Fund* in the list of recipients throughout the 10 months of the study.

Determinants of Individual Donations

In this section, we use participants' responses to the questionnaire and epidemiological data to explore their relevance for donation behavior. On average, participants in our sample are 23.4 years old, and 58.3% of our participants are female; 44.2, 33.8, and 18.4% are of the Austrian, German, and Italian nationality, respectively. 36.6% of our sample has indicated to have donated to a charitable donation in the past 12 months (in reference to the day of participation) and 23.3% have indicated to have volunteered for a charitable organization in the past 12 months.

For each of the charities available as a potential recipient in the donation experiment, **Figure 3** presents the mean (unstandardized) survey responses on (a) trust in the charity's work, (b) its perceived relevance in fighting the consequences of the pandemic, as well as (c) the perceived level of international assistance. Generally, we observe relatively high and similar average trust levels for each of the nine charities. There are substantial differences in the perceived relevance between the charities in fighting the consequences of the Covid pandemic, with *MSF*, *RC*, and the *WHO Covid-19 Fund* showing equally high average levels, and *WWF* having the lowest score. Finally, we observe that the help of *AI*, *CAR*, *LID*, and *RC* is perceived to be more nationally oriented than that of the other charities under consideration.

Table 2 presents regression analyses for *avg. total* donations for the pooled data. Model 1 examines the impact of the 7-day incidence rate (in logs) of Covid-19 infections in Tyrol; model 2 additionally includes subjects' socio-demographic characteristic, whether they are a member of a charity, as well as their history of charitable work and donations to charities. Model 3 also incorporates participants' self-reported risk perceptions, actions, and motives related to the Covid-19 pandemic, and finally model 4 incorporates also the risk perceptions, actions, and motives related to climate change and to poverty. Looking at the epidemiological situation and Covid-19 perceptions, we observe a significant correlation between the 7-day incidence and the standardized responses to Covid-19 risks (Pearson correlation: $\rho = -0.097$, $p < 0.001$), Covid-19 actions (Pearson correlation: $\rho = -0.188$, $p < 0.001$), and Covid-19 motives (Pearson correlation: $\rho = 0.050$, $p = 0.035$). While the 7-day incidence rate is an objective measure of the epidemiological situation, the Covid-19 risks, actions, and

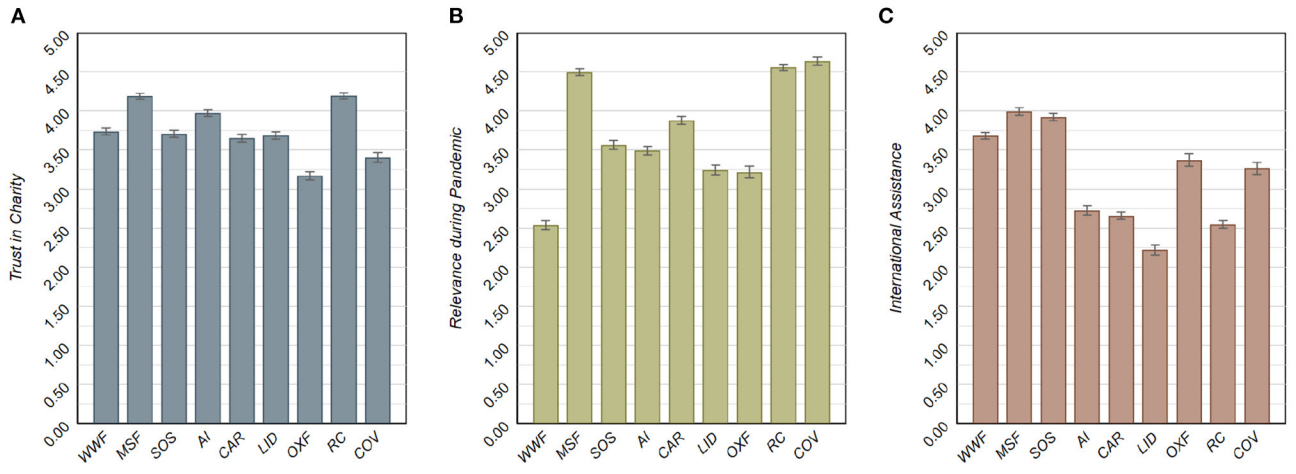

FIGURE 3 | Mean (unstandardized) survey responses on **(A)** trust on the charity, **(B)** relevance of the charity work in fighting the Covid-19 pandemic, and **(C)** the extent to which the work of the charity is perceived to be internationally oriented, separated by the eight charities and the *WHO Covid-19 Fund*. Error bars indicate 95% confidence intervals.

motives give a subjective measure of the understanding of the pandemic and the epidemiological situation.

First, we observe that the 7-day incidence as a measure for the epidemiological situation is a significant determinant of individuals' donation behavior in all model specifications. Furthermore, we report that females donate significantly more and subjects having donated to charities in the past are also associated with significantly higher total donations in the donation task. Finally, we do not find evidence that self-reported Covid-19 risk perceptions are a significant predictor of donations, but we observe that Covid-19 actions and motives show a significant and positive relationship with donations. The Covid-19 motives are however not significant after controlling for the additional variables in Model 5. Perceptions of risks associated with poverty are negatively correlated with donations, while poverty actions have a positive and significant impact. We do not observe significant effects of perceived risks, motives, or actions related to climate change on total donations.

Table 3 presents the model in Model 5 of **Table 2**, including in addition the charity-specific self-reported degree of trust on the charity, the perceived relevance of the charity in fighting the consequences of the Covid-19 pandemic, and whether the charity is perceived to provide assistance internationally or nationally. The results show that the trust in the charity have a significant positive effect on donations for each individual charity. Interestingly, the epidemiological data does not significantly correlate with donations to any of the different charities after controlling for subjects' perception about each charity. The rest of the variables significantly affect donations for some of the charities, but not generally for all of them.

5. DISCUSSION AND CONCLUSION

This paper presents evidence on long-term (10 months) substitution effects that the Covid-19 pandemic has on other social concerns. We report results from a large online experiment with 1,762 students making real-life donations to charities between April 2020 and January 2021. As apprehended by policy makers, our findings suggest a substitution effect due to the Covid-19 pandemic. The data shows that introducing the *Covid-19 Solidarity Response Fund for WHO* as a potential recipient significantly reduces the donations to the rest of eight organizations, as compared to another treatment where only eight charities comprising a wide range of social concerns are available. This result is driven by two main observations: (i) Participants donate substantial amounts to the *WHO Covid-19 Fund*; and (ii) the total donations are not significantly different when the *WHO Covid-19 Fund* is present. That is, aggregate pro-social concerns do not differ depending on whether the *WHO Covid-19 Fund* is available in the list of charitable organizations participants could donate to. This is in line with previous results for treatment effects reported in Blanco et al. (2020) for the onset of the pandemic.

It is worth emphasizing that these results differ from the results that could be expected to derive from the methodological variation of the number of recipients. Our experimental design implies that there are eight possible recipients in *Baseline*, whereas the number of possible recipients is increased to nine in *COVID-19*. In principle, this variation in the number of possible recipients could already affect the results, rather than (or in addition to) the fact that the *WHO Covid-19 Fund* is the introduced charity. The previous evidence on charity competition reviewed in section 2 suggests that the experimental design would induce higher total donations in *COVID-19* (with nine possible recipients) than in *Baseline* (with eight possible recipients). Previous studies experimentally varying the number of charities to which people can donate have consistently observed increased aggregate levels of total donations (Soyer and Hogarth, 2011; Schmitz, 2021). This is not what we observe in our data. Participants' average total donations actually turn out to be

TABLE 2 | Regression analyses of total donations (pooled across all charities and all treatments) on 7-day incidence rates and individual-level characteristics.

	(1)	(2)	(3)	(4)	(5)
7-day incidence (log)	0.126***	0.123***	0.125***	0.138***	0.132***
	(0.029)	(0.028)	(0.028)	(0.029)	(0.028)
Age (in Years)		–0.024	–0.028	–0.017	-0.017
		(0.017)	(0.017)	(0.017)	(0.018)
Female		1.020***	1.007***	0.855***	0.705***
		(0.137)	(0.136)	(0.135)	(0.134)
Germany		–0.126	-0.102	–0.066	–0.131
		(0.149)	(0.148)	(0.147)	(0.147)
Italy		–0.317	–0.278	–0.239	–0.223
		(0.179)	(0.180)	(0.180)	(0.177)
Other Country		0.204	0.180	0.199	0.179
		(0.368)	(0.366)	(0.354)	(0.355)
Charity member			–0.037	–0.059	–0.099
			(0.176)	(0.175)	(0.174)
Charity Work			0.029	0.017	–0.089
			(0.164)	(0.162)	(0.163)
Charity Donations			0.451**	0.404**	0.232
			(0.140)	(0.138)	(0.141)
Covid-19: Risks				–0.091	–0.128
				(0.078)	(0.078)
Covid-19: Actions				0.278***	0.258**
				(0.078)	(0.079)
Covid-19: Motives				0.217**	0.124
				(0.084)	(0.092)
Climate: Risks					0.135
					(0.089)
Climate: Actions					0.123
					(0.088)
Climate: Motives					–0.021
					(0.112)
Poverty: Risks					–0.158*
					(0.076)
Poverty: Actions					0.272**
					(0.089)
Poverty: Motives					0.096
					(0.098)
Constant	2.564***	2.627***	2.538***	2.324***	2.544***
	(0.102)	(0.445)	(0.443)	(0.443)	(0.457)
Observations	1,762	1,751	1,751	1,751	1,751
Pseudo R^2	0.004	0.018	0.020	0.028	0.035

*The table reports the results of Tobit regressions with €0 and €3 as the lower and upper limit, respectively. Robust standard errors are provided in parentheses. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.*

smaller in the *COVID-19* treatment with nine possible recipients as compared to the *Baseline* treatment with eight possible recipients. While the difference is not statistically significant, our results do not support an increase in total donations due to an increase in the number of possible recipients. Thus, we believe that the results reported here are a lower bound estimate of the substitution effect due to the presence of the *WHO Covid-19 Fund*.

Providing a full characterization of the impact of including the *WHO Covid-19 Fund* in the treatment comparison would require considering ten different treatment conditions: including all nine charities and a sequential exclusion of one single charity in nine additional treatments. One could then compare the strength of the different treatment effects for the exclusion of the *WHO Covid-19 Fund* as compared to the treatment effect from the exclusion of each other charity. Given the limitations with respect

TABLE 3 | Regression analyses of donations to charity (pooled across all treatments) on 7-day incidence rates and individual-level characteristics, separated by the eight charities and the *WHO Covid-19 Fund*.

	WWF	***MSF***	***SOS***	***AI***	***CAR***	***LID***	***OXF***	***RK***	***COV***
7–day incidence (log)	0.009	0.019	0.013	0.023	–0.009	0.002	–0.013	–0.040*	0.011
	(0.014)	(0.012)	(0.016)	(0.016)	(0.019)	(0.016)	(0.024)	(0.017)	(0.044)
Age (in Years)	0.018	0.007	–0.025*	0.007	–0.028*	–0.006	–0.010	–0.016	0.048
	(0.012)	(0.010)	(0.012)	(0.012)	(0.013)	(0.010)	(0.016)	(0.012)	(0.038)
Female	0.152	0.180*	0.169	0.103	0.089	0.325***	0.383**	0.252**	0.231
	(0.083)	(0.070)	(0.095)	(0.084)	(0.106)	(0.090)	(0.128)	(0.094)	(0.256)
Germany	0.076	0.102	0.105	0.088	–0.058	0.287**	–0.187	–0.247*	–0.125
	(0.083)	(0.073)	(0.098)	(0.093)	(0.113)	(0.107)	(0.131)	(0.101)	(0.265)
Italy	0.109	0.214*	–0.218	–0.000	–0.241	0.080	–0.273	0.057	–0.451
	(0.110)	(0.083)	(0.129)	(0.117)	(0.139)	(0.114)	(0.197)	(0.119)	(0.340)
Other Country	0.303	0.254	0.465*	0.154	0.495**	0.282	0.267	0.183	–1.304*
	(0.178)	(0.136)	(0.192)	(0.180)	(0.190)	(0.196)	(0.270)	(0.175)	(0.582)
Charity Member	–0.005	0.038	–0.127	0.082	0.077	0.196	–0.133	0.269*	0.002
	(0.104)	(0.084)	(0.116)	(0.100)	(0.125)	(0.110)	(0.155)	(0.126)	(0.296)
Charity Work	0.017	0.042	0.168	–0.041	–0.065	–0.121	0.146	0.100	0.128
	(0.092)	(0.078)	(0.107)	(0.103)	(0.108)	(0.099)	(0.143)	(0.110)	(0.285)
Charity Donations	0.110	–0.097	–0.030	–0.160	0.079	–0.020	0.127	–0.027	–0.216
	(0.079)	(0.067)	(0.088)	(0.087)	(0.107)	(0.092)	(0.123)	(0.097)	(0.245)
Covid–19: Risks	–0.020	–0.022	0.070	0.047	–0.054	0.035	–0.018	0.096*	0.084
	(0.044)	(0.038)	(0.050)	(0.046)	(0.061)	(0.045)	(0.067)	(0.046)	(0.146)
Covid–19: Actions	–0.067	0.044	–0.062	0.007	–0.151**	0.001	–0.153	0.001	0.024
	(0.050)	(0.039)	(0.054)	(0.052)	(0.059)	(0.051)	(0.080)	(0.055)	(0.143)
Covid–19: Motives	–0.036	–0.001	–0.069	–0.085	0.107	0.016	–0.027	–0.050	0.040
	(0.050)	(0.045)	(0.065)	(0.056)	(0.067)	(0.057)	(0.091)	(0.060)	(0.151)
Climate: Risks	0.141**	–0.054	–0.033	–0.088	–0.024	–0.009	–0.081	–0.040	–0.034
	(0.052)	(0.041)	(0.058)	(0.058)	(0.065)	(0.053)	(0.080)	(0.055)	(0.162)
Climate: Actions	0.076	–0.010	–0.013	–0.043	–0.004	–0.003	–0.037	–0.021	0.025
	(0.052)	(0.038)	(0.055)	(0.047)	(0.065)	(0.053)	(0.066)	(0.056)	(0.163)
Climate: Motives	0.161**	–0.015	–0.076	–0.004	–0.181*	–0.192**	0.147	–0.118	–0.385
	(0.058)	(0.050)	(0.074)	(0.066)	(0.078)	(0.059)	(0.090)	(0.067)	(0.220)
Poverty: Risks	0.004	–0.104**	–0.061	–0.076	–0.044	–0.016	0.058	–0.051	0.039
	(0.045)	(0.038)	(0.054)	(0.047)	(0.054)	(0.046)	(0.068)	(0.049)	(0.140)
Poverty: Actions	–0.093	0.085*	0.238***	–0.049	0.075	0.057	0.137	0.038	0.142
	(0.050)	(0.040)	(0.062)	(0.057)	(0.061)	(0.054)	(0.086)	(0.060)	(0.171)
Poverty: Motives	–0.109	0.013	–0.042	0.080	0.058	0.057	–0.207*	0.051	0.201
	(0.056)	(0.055)	(0.070)	(0.059)	(0.068)	(0.060)	(0.097)	(0.074)	(0.188)
Trust in Charity	0.352***	0.311***	0.374***	0.372***	0.376***	0.294***	0.362***	0.414***	0.545***
	(0.043)	(0.034)	(0.056)	(0.044)	(0.054)	(0.042)	(0.060)	(0.051)	(0.133)
Relevance during Pandemic	0.083*	0.036	0.103*	0.070	0.067	0.060	0.144	0.034	0.118
	(0.039)	(0.030)	(0.048)	(0.042)	(0.046)	(0.041)	(0.074)	(0.051)	(0.125)
International Assistance	–0.005	0.108***	–0.028	0.011	0.111*	–0.160***	0.164*	0.003	–0.067
	(0.036)	(0.032)	(0.040)	(0.042)	(0.047)	(0.041)	(0.063)	(0.041)	(0.119)
Constant	–0.535	–0.013	0.142	–0.349	–0.116	–0.498	–0.437	–0.011	–0.712
	(0.295)	(0.252)	(0.298)	(0.307)	(0.329)	(0.274)	(0.412)	(0.294)	(0.952)
Observations	656	685	578	651	656	489	238	696	469
Pseudo R^2	0.110	0.092	0.111	0.073	0.111	0.123	0.227	0.098	0.025

*The table reports the results of Tobit regressions with €0 and €3 as the lower and upper limit, respectively. Robust standard errors are provided in parentheses. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.*

to the number of students in the subject pool we could not run these ten treatments for each data point for 10 months. In this study we have prioritized the use of the subject pool to assess the time evolution of donation behavior. Other studies could focus on the methodological question of assessing to which extent there could be substitution effects from restricting the decision setting to other social causes, and if present, the relative size of the substitution for different social concerns.

Looking into the evolution over time, for the 10 months time period covered by our data, we do not find any indication for systematic trends in donations. As compared to the results in Blanco et al. (2020), additional analyses show no change in the total donations from the first 8 weeks to the subsequent 8 months of data in the *Baseline* nor *Covid-19 Only* treatments. We do observe however a significant increase in donations for the *COVID-19* treatment when comparing the first 8 weeks and the subsequent 8 months of data collection. We observe that for the *COVID-19* treatment there is a significant increase in donations to *AI*, *CAR*,*WWF*, *LID*, and *OXF*; there is no significant change for *MSF* and *RC*; and there is a significant decrease for the *WHO Covid-19 fund*. For the *Baseline* and the *Covid-19 Only* treatments there are no significant differences in donations to each charity generally, with the only exception of a significant increase of donations to *WWF* for the *Baseline* treatment.

When looking into the determinants of aggregate donations by participants in our study, we see that there is evidence that the worsening of the epidemiological situation, measured by the 7 day incidence, significantly increases total donations. Moreover, as expected, we find a significant gender effect, with women donating significantly higher total amounts than men. Moreover, people taking actions against Covid-19 and against poverty donate significantly more, whereas people fearing risks from poverty donate significantly less. When looking into separate donations to each charity, we find that trusting a given charitable organization is the strongest explanatory factor of why participants donate to the respective charity.

We believe that our results can be informative to policy makers, helping them better understand human behavior during global shocks such as the Covid-19 pandemic. This global health crisis has been attracting the international community's attention to the interrelation between the environment, health, and inequality in human well-being. At the same time, there is a fear that the pandemic dominates both policy and social agendas, at the expense of other social concerns. We present evidence that such substitution of social concerns is only partially present among the participants in our study. While we observe a reduction of concerns for other (non-Covid-19) social objectives, donations to charities in other domains remain at relatively high levels. This behavior seems to be stable during the pandemic; we do not find clear trends over the 10 months of our study. The aforementioned results suggests an optimistic prospect since it represents a backup for the ongoing considerations with respect to other social concerns that public administrations and charities worldwide have been pursuing before and during global crisis such as the Covid-19 pandemic. It is also worth highlighting how the participants in our study are sensitive to the needs of others, increasing total donations in times of higher incidence rates of Covid-19 infections.

The experimental methodology used in this study inevitably is subject to certain limitations. The experimental design of the donation task allows to draw causal inference with respect to the treatment effects we report, but the donation task under analysis is only a proxy of pro-social behavior in the field. Similarly, as common in economic experiments in the laboratory, our participants form a very homogeneous sample of students with similar age, education level and socio-demographic background. An additional limitation of our study is that the nature of some of our treatments, i.e. *Covid-19* and *Covid-19 Only*, might be subject to some experimenter demand effects (Zizzo, 2010) since the presence of a Covid-19 fund might be very appealing for the participants. However, recent evidence suggests that pro-social behavior in the lab—elicited using a similar donation task—significantly correlates with health behavior during the pandemic in the field (Campos-Mercade et al., 2021). Further, by nature, we cannot analyze the extent to which treatment results would have differed, had we started to collect data prior to the pandemic. The *WHO Covid-19 fund* was established only after the pandemic stroke. Finally, the data used in this project was collected for a pre-determined (and pre-registered) period of 10 months during the pandemic. Certainly, we see value in future research replicating this data collection after this pandemic is over in order to expand our understanding of the very long term effects and the behavior in the aftermath of such unprecedented global shocks.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are publicly available. This data can be found here: https://osf.io/uy7ps.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Board for Ethical Questions in Science of the University of Innsbruck. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

EB: conceptualization, experimental design, writing-original draft, supervision, project administration, and funding acquisition. AB: performing experiments and writing-original draft. FH: statistical analysis, data curation, writing-original draft, and visualization of the manuscript. TJ-L: experimental design, statistical analysis, writing-review, and editing. NS: development of the questionnaire, programming, performing experiments, writing-original draft, and project administration. All authors contributed to the article and approved the submitted version.

FUNDING

Funding was provided by FWF (No. P 32859).

ACKNOWLEDGMENTS

We thank all the participants in the sessions for their time and generous donations to the charities. We also thank Tobias Trojok for assistance in developing the questionnaire. We thank participants of the 6th Workshop on Experiments for the Environment and the 2021 Annual Meeting of the Austrian Science Association for helpful comments, as well as Loukas Balafoutas, Ivo Steimanis, Björn Vollan, and James M. Walker for their discussions and comments on the research.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2021.743054/full#supplementary-material

REFERENCES

Abel, M., and Brown, W. (2020). Prosocial behavior in the time of COVID-19: The effect of private and public role models. *IZA Discussion Paper No. 13207*.

Abel, M., Byker, T., and Carpenter, J. (2020). Socially optimal mistakes? Debiasing COVID-19 mortality risk perceptions and prosocial behavior. *IZA Discussion Paper No. 13560*.

Almunia, M., Guceri, I., Lockwood, B., and Scharf, K. (2020). More giving or more givers? the effects of tax incentives on charitable donations in the uk. *J. Public Econ*. 183:104114. doi: 10.1016/j.jpubeco.2019.104114

Andreoni, J. (1989). Giving with impure altruism: applications to charity and ricardian equivalence. *J. Polit. Econ*. 97, 1447–1458. doi: 10.1086/261662

Andreoni, J. (1990). Impure altruism and donations to public goods: a theory of warm-glow giving. *Econ. J*. 100, 464–477. doi: 10.2307/2234133

Ariely, D., Bracha, A., and Meier, S. (2009). Doing good or doing well? Image motivation and monetary incentives in behaving prosocially. *Am. Econ. Rev*. 99, 544–555. doi: 10.1257/aer.99.1.544

Bartling, B., Weber, R. A., and Yao, L. (2015). Do markets erode social responsibility? *Q. J. Econ*. 130, 219–266. doi: 10.1093/qje/qju031

Bénabou, R., and Tirole, J. (2006). Incentives and prosocial behavior. *American Economic Review* 96, 1652–1678. doi: 10.1257/aer.96.5.1652

Black, N., De Gruyter, E., Petrie, D., and Smith, S. (2020). Altruism born of suffering? the impact of an adverse health shock on pro-social behaviour. *CEPR Discussion Paper No. DP15535*.

Blanco, E., Baier, A., Holzmeister, F., Jaber-Lopez, T., and Struwe, N. (2020). "Substitution of social concerns under the covid-pandemic," in *Working Papers in Economics and Statistics, 2020-30*, University of Innsbruck.

Blasco-Belled, A., Tejada-Gallardo, C., Torrelles-Nadal, C., and Alsinet, C. (2020). The costs of the covid-19 on subjective well-being: an analysis of the outbreak in spain. *Sustainability* 12:243. doi: 10.3390/su12156243

Bock, O., Baetge, I., and Nicklisch, A. (2014). hroot: Hamburg registration and organization online tool. *Eur. Econ. Rev*. 71, 117–120. doi: 10.1016/j.euroecorev.2014.07.003

Branas-Garza, P., Jorrat, D., Alfonso, A., Espín, A., Muñoz, T., and Kovarik, J. (2020). Exposure to the Covid-19 pandemic and generosity in southern Spain. *PsyArXiv [Preprint]*. doi: 10.31234/osf.io/6ktuz

Brown, A. L., Meer, J., and Williams, J. F. (2017). Social distance and quality ratings in charity choice. *J. Behav. Exp. Econ*. 66, 9–15. doi: 10.1016/j.socec.2016.04.006

Brown, S., Harris, M. N., and Taylor, K. (2012). Modelling charitable donations to an unexpected natural disaster: Evidence from the us panel study of income dynamics. *J. Econ. Behav. Organ*. 84, 97–110. doi: 10.1016/j.jebo.2012.08.005

Campos-Mercade, P., Meier, A. N., Schneider, F. H., and Wengström, E. (2021). Prosociality predicts health behaviors during the covid-19 pandemic. *J. Public Econ*. 195:104367. doi: 10.1016/j.jpubeco.2021.104367

Cao, W., Fang, Z., Hou, G., Han, M., Xu, X., Dong, J., et al. (2020). The psychological impact of the covid-19 epidemic on college students in china. *Psychiatry Res*. 287:112934. doi: 10.1016/j.psychres.2020.112934

Cassar, A., Healy, A., and Von Kessler, C. (2017). Trust, risk, and time preferences after a natural disaster: experimental evidence from thailand. *World Dev*. 94, 90–105. doi: 10.1016/j.worlddev.2016.12.042

Chaudhuri, A. (2011). Sustaining cooperation in laboratory public goods experiments: a selective survey of the literature. *Exp. Econ*. 14, 47–83. doi: 10.1007/s10683-010-9257-1

Coyle, D. (2020). Economists must collaborate courageously. *Nature* 582:9. doi: 10.1038/d41586-020-01505-3

Crumpler, H., and Grossman, P. J. (2008). An experimental test of warm glow giving. *J. Public Econ*. 92, 1011–1021. doi: 10.1016/j.jpubeco.2007.12.014

Deck, C., and Murphy, J. J. (2019). Donors change both their level and pattern of giving in response to contests among charities. *Eur. Econ. Rev*. 112, 91–106. doi: 10.1016/j.euroecorev.2018.12.004

Deryugina, T., and Marx, B. M. (2021). Is the supply of charitable donations fixed? evidence from deadly tornadoes. *Am. Econ. Rev*. Insights, 3:383–98.

Dobson, A. P., Pimm, S. L., Hannah, L., Kaufman, L., Ahumada, J. A., Ando, A. W., et al. (2020). Ecology and economics for pandemic prevention. *Science* 369, 379–381. doi: 10.1126/science.abc3189

Drichoutis, A. C., and Nayga, R. (2020). On the stability of risk and time preferences amid the covid-19 pandemic. *Exp. Econ*. 13, 1–36. doi: 10.1007/s10683-021-09727-6

Eckel, C. C., and Grossman, P. J. (1996). Altruism in anonymous dictator games. *Games Econ. Behav*. 16, 181–191. doi: 10.1006/game.1996.0081

Eckel, C. C., and Grossman, P. J. (2001). Chivalry and solidarity in ultimatum games. *Econ. Inq*. 39, 171–188. doi: 10.1111/j.1465-7295.2001.tb00059.x

Eckel, C. C., and Grossman, P. J. (2003). Rebate versus matching: does how we subsidize charitable contributions matter? *J. Public Econ*. 87, 681–701. doi: 10.1016/S0047-2727(01)00094-9

Eckel, C. C., Grossman, P. J., and Johnston, R. M. (2005). An experimental test of the crowding out hypothesis. *J. Public Econ*. 89, 1543–1560. doi: 10.1016/j.jpubeco.2004.05.012

Eckel, C. C., Grossman, P. J., and Milano, A. (2007). Is more information always better? an experimental study of charitable giving and hurricane katrina. *Southern Econ. J*. 388–411. doi: 10.1002/j.2325-8012.2007.tb00845.x

Eckel, C. C., Guney, B., and Uler, N. (2020). Independent vs. coordinated fundraising: Understanding the role of information. *Eur. Econ. Rev*. 127:103476. doi: 10.1016/j.euroecorev.2020.103476

Filiz-Ozbay, E., and Uler, N. (2019). Demand for giving to multiple charities: an experimental study. *J. Eur. Econ Assoc*. 17, 725–753. doi: 10.1093/jeea/jvy011

Fisman, R., Jakiela, P., and Kariv, S. (2015). How did distributional preferences change during the great recession? *J. Public Econ*. 128, 84–95. doi: 10.1016/j.jpubeco.2015.06.001

Frey, B. S., and Meier, S. (2004). Social comparisons and pro-social behavior: Testing "conditional cooperation" in a field experiment. *Am. Econ. Rev*. 94, 1717–1722. doi: 10.1257/0002828043052187

Gallier, C., Goeschl, T., Kesternich, M., Lohse, J., Reif, C., and Römer, D. (2019). "Inter-charity competition under spatial differentiation: sorting, crowding, and spillovers," in *ZEW-Centre for European Economic Research Discussion Paper*, 19–039.

Garcia, T., Massoni, S., and Villeval, M. C. (2020). Ambiguity and excuse-driven behavior in charitable giving. *Eur. Econ. Rev*. 124:103412. doi: 10.1016/j.euroecorev.2020.103412

Gneezy, U., Keenan, E. A., and Gneezy, A. (2014). Avoiding overhead aversion in charity. *Science* 346, 632–635. doi: 10.1126/science.1253932

Goldthau, A., and Hughes, L. (2020). Protect global supply chains for low-carbon technologies. *Nature* 585, 28–30. doi: 10.1038/d41586-020-02499-8

Gonzalez-Sanguino, C., Ausin, B., Castellanos, M. A., Saiz, J., Lopez-Gomez, A., Ugidos, C., et al. (2020). Mental health consequences during the initial stage of the 2020 coronavirus pandemic (covid-19) in spain. *Brain Behav. Immun*. 87, 172–176. doi: 10.1016/j.bbi.2020.05.040

Guo, Y., Shachat, J., Walker, M. J., and Wei, L. (2020). Viral social media videos can raise pro-social behaviours when an epidemic arises. *ESI Working Paper* 1–19. doi: 10.1007/s40881-021-00104-w

Harwell, H., Meneses, D., Moceri, C., Rauckhorst, M., Zindler, A., and Eckel, C. (2015). *Did the Ice Bucket Challenge Drain the Philanthropic Reservoir*? Economic Research Laboratory, Texas A&M University, Working Paper.

Hodges, K., and Jackson, J. (2020). Pandemics and the global environment. *Sci. Adv*. 6. doi: 10.1126/sciadv.abd1325

Jakiela, P., Miguel, E., and Te Velde, V. L. (2015). You've earned it: estimating the impact of human capital on social preferences. *Exp. Econ*. 18, 385–407. doi: 10.1007/s10683-014-9409-9

Kirchler, M., Huber, J., Stefan, M., and Sutter, M. (2016). Market design and moral behavior. *Manag. Sci*. 62, 2615–2625. doi: 10.1287/mnsc.2015.2246

Lange, A., and Stocking, A. (2012). The complementarities of competition in charitable fundraising. *Congressional Budget Office Washington, DC Working Paper* 32.

Levitt, S. D., and List, J. A. (2007). What do laboratory experiments measuring social preferences reveal about the real world? *J. Econ. Perspect*. 21, 153–174. doi: 10.1257/jep.21.2.153

Li, K. K., Huang, B., Tam, T., and Hong, Y.-Y. (2020). Does the covid-19 pandemic affect people's social and economic preferences? evidence from china. *SSRN 3690072*.

Li, S., Liu, X., and Li, J. (2021). The contagion of donation behaviors changes along with the abatement of the covid-19 pandemic: an intertemporal survey experiment. *Front. Psychol*. 12:1485. doi: 10.3389/fpsyg.2021.585128

Lieberoth, A., Lin, S.-Y., Stöckli, S., Han, H., Kowal, M., Gelpi, R., et al. (2021). Stress and worry in the 2020 coronavirus pandemic: relationships to trust and compliance with preventive measures across 48 countries in the covidistress global survey. *R. Soc. Open Sci*. 8:200589. doi: 10.1098/rsos.200589

Lohmann, P., Gsottbauer, E., You, J., and Kontoleon, A. (2020). Social preferences and economic decision-making in the wake of covid-19: experimental evidence from china. *SSRN 3705264*.

Mahler, D., Lakner, C., Castaneda-Aguilar, R. A., and Wu, H. (2020). *The impact of Covid-19 (Coronavirus) on global* poverty: why Sub-Saharan Africa might be the region hardest hit. *World Bank Blogs*. Blogs, 20. Available online at: https://bit.ly/30wvyEl.

Meier, S. (2007). Do subsidies increase charitable giving in the long run? matching donations in a field experiment. *J. Eur. Econ. Assoc*. 5, 1203–1222. doi: 10.1162/JEEA.2007.5.6.1203

Naidoo, R., and Fisher, B. (2020). Reset sustainable development goals for a pandemic world. *Nature* 583, 198–201. doi: 10.1038/d41586-020-01999-x

Odriozola-Gonzalez, P., Planchuelo-Gomez, A., Irurtia, M. J., and de Luis-Garcia, R. (2020a). Psychological effects of the covid-19 outbreak and lockdown among students and workers of a spanish university. *Psychiatry Res*. 290:113108. doi: 10.1016/j.psychres.2020.113108

Odriozola-Gonzalez, P., Planchuelo-Gomez, A., Irurtia-Muniz, M. J., and de Luis-Garcia, R. (2020b). Psychological symptoms of the outbreak of the covid-19 crisis and confinement in the population of spain. *PsyArXiv*. doi: 10.31234/osf.io/mq4fg

Ostrom, E. (1990). *Governing the Commons: The Evolution of Institutions for Collective Action. Political Economy of Institutions and Decisions*. Cambridge: Cambridge University Press.

Planchuelo-Gomez, A., Odriozola-Gonzalez, P., Irurtia, M. J., and de Luis-Garcia, R. (2020). Longitudinal evaluation of the psychological impact of the covid-19 crisis in spain. *J. Affect. Disord*. 277, 842–849. doi: 10.1016/j.jad.2020.09.018

Reinstein, D. (2007). *Substitution Between (and Motivations for) Charitable Contributions: An Experimental Study*. Economics Discussion Papers 2935, University of Essex, Department of Economics.

Rodriguez-Rey, R., Garrido-Hernansaiz, H., and Collado, S. (2020). Psychological impact and associated factors during the initial stage of the coronavirus (covid-19) pandemic among the general population in spain. *Front. Psychol*. 11:1540. doi: 10.3389/fpsyg.2020.01540

Roma, P., Monaro, M., Colasanti, M., Ricci, E., Biondi, S., Di Domenico, A., et al. (2020). A 2-month follow-up study of psychological distress among italian people during the covid-19 lockdown. *Int. J. Environ. Res. Public Health* 17:8180. doi: 10.3390/ijerph17218180

Rosenbloom, D., and Markard, J. (2020). A COVID-19 recovery for climate. *Science* 368, 447. doi: 10.1126/science.abc4887

Sabater-Grande, G., Garcia-Gallego, A., Georgantzis, N., and Herranz-Zarzoso, N. (2021). When will the lockdown end? confinement duration forecasts and self-reported life satisfaction in spain: a longitudinal study. *Front. Psychol*. 12:874. doi: 10.3389/fpsyg.2021.635145

Salari, N., Hosseinian-Far, A., Jalali, R., Vaisi-Raygani, A., Rasoulpoor, S., Mohammadi, M., et al. (2020). Prevalence of stress, anxiety, depression among the general population during the covid-19 pandemic: a systematic review and meta-analysis. *Glob. Health* 16, 1–11. doi: 10.1186/s12992-020-00589-w

Scharf, K. A., Smith, S., and Wilhelm, M. (2017). Lift and shift: the effect of fundraising interventions in charity space and time. CESifo Working Paper Series No. 6694, Available online at SSRN: https://ssrn.com/abstract=3074331 (accessed October 18, 2017).

Schmitz, J. (2021). Is charitable giving a zero-sum game? the effect of competition between charities on giving behavior. *Manag. Sci*. doi: 10.1287/mnsc.2020.3809

Shachat, J., Walker, M. J., and Wei, L. (2020). The impact of the Covid-19 pandemic on economic behaviours and preferences: experimental evidence from Wuhan. *Working Paper*, 20–33.

Shreedhar, G., and Mourato, S. (2020). Linking human destruction of nature to COVID-19 increases support for wildlife conservation policies. *Environ. Resour. Econ*. 76, 963–999. doi: 10.1007/s10640-020-00444-x

Soyer, E., and Hogarth, R. M. (2011). The size and distribution of donations: effects of number of recipients. *Judg. Decis. Making* 6, 616–628.

Tollefson, J. (2020). Can the world's most influential climate report carry on? *Nature*. doi: 10.1038/d41586-020-01047-8

United Nations (2020). Sustainable development goals report 2020. Availble online at: https://bit.ly/3foFdks

Vesterlund, L. (2003). The informational value of sequential fundraising. *J. Public Econ*. 87, 627–657. doi: 10.1016/S0047-2727(01)00187-6

von Braun, J., Zamagni, S., and Sorondo, M. S. (2020). The moment to see the poor. *Science* 368, 214. doi: 10.1126/science.abc2255

Voors, M. J., Nillesen, E. E. M., Verwimp, P., Bulte, E. H., Lensink, R., and Van Soest, D. P. (2012). Violent conflict and behavior: a field experiment in burundi. *Am. Econ. Rev*. 102, 941–964. doi: 10.1257/aer.102.2.941

Wang, C., and Zhao, H. (2020). The impact of covid-19 on anxiety in chinese university students. *Front. Pychol*. 11:1168. doi: 10.3389/fpsyg.2020.01168

WWF, Jeffries B. (2020). The Loss of Nature and Rise of Pandemics-Protection Human and Planatory Health. Italy: WWF. Available online at: https://d2ouvy59p0dg6k.cloudfront.net/downloads/the_loss_of_nature_and_rise_of_pandemics___protecting_human_and_planetary_health.pdf

Zhang, S. X., Wang, Y., Rauch, A., and Wei, F. (2020). Unprecedented disruption of lives and work: health, distress and life satisfaction of working adults in china one month into the covid-19 outbreak. *Psychiatry Res*. 288:112958. doi: 10.1016/j.psychres.2020.112958

Zhongming, Z., Linong, L., Wangqiang, Z., and Wei, L. (2020). Open letter to global leaders: A healthy planet for healthy people. *The Club of Rome*. Available online at: https://bit.ly/2DvreMp

Zhou, N., Cao, H., Liu, F., Wu, L., Liang, Y., Xu, J., et al. (2020). A four-wave, cross-lagged model of problematic internet use and mental health among chinese college students: disaggregation of within-person and between-person effects. *Dev. Psychol*. 56, 1009. doi: 10.1037/dev0000907

Zizzo, D. J. (2010). Experimenter demand effects in economic experiments. *Exp. Econ*. 13, 75–98. doi: 10.1007/s10683-009-9230-z

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychology

ORIGINAL RESEARCH
published: 06 October 2021
doi: 10.3389/fpsyg.2021.732248

Does Whistleblowing on Tax Evaders Reduce Ingroup Cooperation?

Philipp Chapkovski[1], Luca Corazzini[2] and Valeria Maggian[2]*

[1] National Research University Higher School of Economics, Russian Federation, Moscow, Russia, [2] Department of Economics and VERA, University of Venice "Ca' Foscari", Venice, Italy

Whistleblowing is a powerful and rather inexpensive instrument to deter tax evasion. Despite the deterrent effects on tax evasion, whistleblowing can reduce trust and undermine agents' attitude to cooperate with group members. Yet, no study has investigated the potential spillover effects of whistleblowing on ingroup cooperation. This paper reports results of a laboratory experiment in which subjects participate in two consecutive phases in unchanging groups: a tax evasion game, followed by a generalized gift exchange game. Two dimensions are manipulated in our experiment: the inclusion of a whistleblowing stage in which, after observing others' declared incomes, subjects can signal other group members to the tax authority, and the provision of information about the content of the second phase before the tax evasion game is played. Our results show that whistleblowing is effective in both curbing tax evasion and improving the precision of tax auditing. Moreover, we detect no statistically significant spillover effects of whistleblowing on ingroup cooperation in the subsequent generalized gift exchange game, with this result being unaffected by the provision of information about the experimental task in the second phase. Finally, the provision of information does not significantly alter subjects' (tax and whistleblowing) choices in the tax evasion game: thus, knowledge about perspective ingroup cooperation did not alter attitude toward whistleblowing.

Keywords: tax evasion, whistleblowing, ingroup cooperation, spillover effects, laboratory experiment JEL classification: H26, C90, D02 PsychoINFO classification: 2900, 4200

OPEN ACCESS

Edited by:
Tarek Jaber-Lopez, Université Paris Nanterre, France

Reviewed by:
Matthias Kasper, University of Vienna, Austria
Charles Noussair, University of Arizona, United States

****Correspondence:***
Philipp Chapkovski
fchapkovskiy@hse.ru

Specialty section:
This article was submitted to Personality and Social Psychology, a section of the journal Frontiers in Psychology

Received: *28 June 2021*
Accepted: *13 September 2021*
Published: *06 October 2021*

Citation:
Chapkovski P, Corazzini L and Maggian V (2021) Does Whistleblowing on Tax Evaders Reduce Ingroup Cooperation? Front. Psychol. 12:732248. doi: 10.3389/fpsyg.2021.732248

INTRODUCTION

Tax evasion and tax fraud represent a major concern all over the world[1], subtracting fiscal resources that are needed to finance public goods and questioning the effectiveness and fairness of tax systems.

Whistleblowing by citizens has recently gained increased attention as an effective and viable strategy to contrast tax evasion. For instance, according to the IRS Whistleblower Office, between 2007 and 2016, information submitted by whistleblowers has helped the United States government to recover $3.4 billion of tax revenue[2].

[1] According to the most recent United States Internal Revenue Service tax gap report (Internal Revenue Service, 2019), the average annual gross tax gap was of $441 billion in tax years 2011–13 (slightly over 16 percent of total tax liability). In 2016, the VAT gap in Europe was estimated to be equal to EUR 147.1 billion, 12.3% of the total expected VAT revenue (Internal Revenue Service, 2019).

[2] 2016 Annual Report to the Congress of the Internal Revenue Service (https://www.irs.gov/pub/irs-prior/p5241--2017.pdf), retrieved on November 2, 2018.

Despite the potential fiscal benefits of whistleblowing, the number of studies analyzing its determinants and socio-economic consequences is still limited. In this respect, while there is evidence showing that trust in the government represents an important determinant of the decision to blow the whistle on tax evaders (Antinyan et al., 2020) a research question that remains unexplored is whether whistleblowing can undermine the quality of social interactions within communities. As numerous studies have been shown, those who dare to report the norm violation or crime committed by their own group members are indeed under risk of being stigmatized by their communities (Woldoff and Weiss, 2010). Ostracism of snitchers goes far beyond socially vulnerable groups (such as ethnic minorities, prisons, or districts with high crime rate), including school classes (Morris, 2010) and police departments. Apart from the potential retaliation of the norm violator, whistleblowers also risk to be victim of actions of other members of their reference group, who usually prefer not to work with them (Reuben and Stephenson, 2013). In particular, even when anonymity is fully assured, the whistleblower's actions might be perceived as undermining ingroup trust (Wallmeier, 2019), so that whistleblowing could negatively affect future group cooperation.

In this paper, we report results of a laboratory experiment aimed at: (i) investigating the effects of whistleblowing on tax evasion; and (ii) assessing its potential consequences on ingroup trust and cooperation.

Our experiment includes two consecutive phases. In the first phase, we implement a simple tax evasion game in which participants, randomly assigned to group of five members according to a fixed matching protocol, have to decide the amount of their income they want to report to the central authority in order to pay taxes. In case of auditing, if the declared income is lower than the actual one, the individual has to pay the back taxes on the undeclared income plus a fine.

In the second phase, participants play a generalized gift exchange game. In particular, subjects simultaneously decide how much of their endowment to send to other group members, knowing that the amount sent will be doubled by the experimenter.

We manipulate two main dimensions of our experimental design: the presence of a whistleblowing mechanism and the provision of information at the beginning of the first phase about the content of the experimental task in the second phase. Concerning the first dimension, we distinguish between *Whistleblowing* and *NoWhistleblowing* treatments. In the *Whistleblowing* treatments, after all income declaration choices have been made, each subject is given the possibility to blow the whistle on others so to increase their probability of being audited by the tax authority. Moving to the second manipulated dimension, in the *Information* treatment, information about the content of the experimental task in the second part is provided at the beginning of the experiment, while in the *NoInformation* treatment subjects learn about the second phase only at the end of the tax evasion game. Thus, the information manipulation allows us to investigate whether being aware about the forthcoming cooperative task in the second phase strategically affects the efficacy of whistleblowing and tax evasion in the first phase, making group subjects more reluctant to blow the whistle on other group members.

Our results are summarized as follows. First, whistleblowing is effective in reducing tax evasion as well as in improving the precision of tax auditing. Indeed, participants blow the whistle on ingroup members who misreport their income and the risk of being signaled to the tax authority increases the overall level of tax compliance. Second, we detect no statistically significant spillover effects of whistleblowing on ingroup cooperation in the subsequent generalized gift exchange game, with this result being unaffected by subjects' information about the experimental task in the second part.

The rest of the paper is organized as follows. Section "LITERATURE REVIEW" summarizes the related literature while in Section "EXPERIMENTAL DESIGN" we introduce our experimental design and the experimental procedures implemented. In Section "RESULTS" we present our results and discuss possible explanations. Section "DISCUSSION" concludes and suggests directions for future research.

LITERATURE REVIEW

In this study we investigate the existence and sign of cross-contexts spillover effects of whistleblowing on ingroup trust. Near and Miceli, 1985 (page 4) define whistle-blowing as "the disclosure by organizational members (former or current) of illegal, immoral, or illegitimate practices under the control of their employers, to persons or organizations that may be able to effect action". This widely used definition refers to the hierarchical type of relations where the reported hold structurally more powerful positions than those who report (Loyens, 2013). The main focus of this paper is instead peer reporting whistleblowing, defined as "a lateral control attempts that occur when an in-group member discloses a peer's wrongdoing to higher authorities outside the group" (Trevino and Victor, 1992). In the rest of the paper we will use the terms 'whistleblowing' and 'peer reporting' interchangeably.

Our paper relates to the recent and flourishing literature that investigates the within- or across-context spillovers of policy interventions, which focuses mostly on how they might affect prosocial norms and social preferences beyond those behaviors directly targeted by the institutions (Peysakhovich and Rand, 2016; dAdda et al., 2017; Galbiati et al., 2018; Ghesla et al., 2019). In the laboratory experiment by Engl et al. (2020), participants sequentially play two identical public good games, such that cooperation is institutionally enforced only in the first one. They find evidence of significant positive spillover effects of the institution, meaning that it increases cooperation also in the unregulated game, affecting preferences and beliefs about others' attitude to cooperate. Furthermore, Galeotti et al. (2021) show how policy interventions can exert unintended behavioral effects that go beyond their original scope. More specifically, in their quasi-experiment, both fraudsters and non-fraudsters in public transport when exposed to ticket inspections were more likely to misappropriate money in a different unrelated context, providing evidence

of negative spillover effects of deterrence institutions on intrinsic honesty.

Whether, and under which conditions, whistleblowing represents an effective instrument to curb tax evasion is an intriguing research question that is gaining increasing attention in recent years. Breuer (2013) experimentally investigates whether incentivization of whistleblowing is effective for fostering tax compliance and shows that whistleblowing is successful in limiting tax evasion, even without monetary incentives. Bazart et al. (2020) experimentally study the impact of a whistleblowing-based audit scheme upon taxpayers' reporting decisions. They design an experiment aiming at comparing the relative efficiency of whistleblowing opportunities compared to a standard random-based audit scheme, keeping operating costs constant for the tax administration (neither rewards nor denunciation costs are considered). Their findings confirm that whistleblowing-based audit scheme decreases the monetary amount of evasion, improves the targeting of evaders and raises the tax levy. In their experimental study, Masclet et al. (2019) investigate the effect of whistleblowing programs on tax evasion providing information to participants on the use of the tax revenues in three dynamic treatments: (i) a baseline treatment where tax evaders are obliged to pay taxes on the undeclared income and a penalty if audited, (ii) an information treatment in which participants are also informed about the income declaration rates of all other group members and (iii) a denunciation treatment in which each participant has the possibility to blow the whistle on others. They find that monitoring alone does not increase the declared income while allowing for blowing the whistle decreases tax evasion; moreover, informing participants that the tax revenue was used to finance an environmental public good has no significant impact on either tax compliance or peer reporting. However, the role of information about other tax payers seems to affect the tax compliance rate according to a non-trivial relationship (see the corresponding section of the metastudy examining main factors affecting tax evasion Alm, 2019). On the one hand, if an individual knows that his neighbors are cheating with taxes, he will be more likely to evade taxes as well (Alm et al., 2017). On the other hand, the threat of public disclosure of tax evaders' identity may serve as an effective deterrrent: the cross-cultural study run by Alm et al. (2017) reveals indeed that when the photos of tax evaders were shown to the rest of the group, full compliance raised from 38% to 57%.

Nyreröd and Spagnolo (2021) investigate the effects of introducing economic incentives to stimulate whistleblowing and show that rewarding whistleblowers is associated with a reduction in misbehaviors. Amir et al. (2018) extends the analysis to the indirect effects of the introduction of a whistleblowing program in 2013 in Israel to combat tax evasion. Their findings support the hypothesis that, despite the limited direct effect on tax collection, whistleblowing indirectly increases tax revenues through deterrence.

The effect of whistleblowing programs is not limited only to the tax evasion schemes. They are also proved to have a strong deterrent effect as an antitrust measure (Apesteguia et al., 2007; Hinloopen and Soetevent, 2008). The way a whistleblowing scheme is designed to fight against cartels is usually different from what is observed in tax compliance because, in contrast to the individual crime of tax evasion, the creation of a cartel implies a collusion between group members. Thus, a law maker has to show leniency toward whistleblowers, whose degree affects the effectiveness of the program (Chen and Rey, 2013), something which also depends on the intrinsic motives of the whistleblower (Heyes and Kapur, 2009). Buckenmaier et al. (2020) show that introducing the possibility to blow the whistle on others both reduces the probability that subjects collude and accept bribes and increases tax compliance. More importantly, they also document strong spillover effects of leniency programs, with a strong time persistence of the effects of the whistleblowing program after its removal. Our experimental study is aimed at shedding light on another potential spillover effect of whistleblowing. Indeed, as long as whistleblowing is interpreted as a non-cooperative institution that is mainly intended to punish other group members, institutionalizing the possibility of individuals to denounce each other's wrongdoing might finally result in an erosion of ingroup trust, making coordination and cooperation for mutual benefit more difficult to achieve. Ingroup trust is indeed a necessary component of group cohesion (Fonseca et al., 2019), which in turn affects a group's ability to successfully participate in cooperation and coordination games (Gächter et al., 2017). When an individual makes a decision about peer reporting, he might undermine this loyalty, lowering other members' willingness to cooperate. However, the relations between group loyalty and norm violation are complex. On the one hand, loyalty can decrease norm violations within groups (Hildreth et al., 2016) while, on the other hand, people tend to perceive loyal but dishonest actions as more ethical than disloyal but honest ones (Hildreth and Anderson, 2018).

Whistleblowing has been also investigated in different contexts, including corruption and the work environment. In particular, depending on the level of interdependency of work tasks, the work environment represents a further important context in which ingroup trust and whistleblowing institutions are strongly related to each other (Lau and Liden, 2008). Concerning how whistleblowing affects, and is affected by, awareness about future interactions in the workplace, there are important papers that are close to ours. In a hierarchical framework, Wallmeier (2019) investigates the emergence of fraudulent whistleblowing. More specifically, in his laboratory experiment, a manager and an employee play a modified version of a trust game. Before interacting with the employee, the manager can engage in embezzlement, which in turn exerts a negative externality on a third party. The employee observes possible misbehavior and may report it to an external authority. He finds that both introducing an incentivized and an anonymous reporting mechanism increases fraudulent whistleblowing and discourages subsequent group cooperation. Finally, Reuben and Stephenson (2013) investigate a situation in which individuals have the opportunity to blow the whistle on those who lie for personal advantage and found that whistleblowers are indeed ostracized. However, differently from these papers, anonymity of the whistleblower is fully assured in our study, which in turn removes the possibility of ostracism

and direct retaliation. In this respect, beside its deterrence effects, our experimental design is aimed at assessing the indirect effects exerted by whistleblowing in the tax evasion game of the first phase on the level of ingroup trust and cooperation in the different, generalized gift exchange context subjects participate in the second phase.

EXPERIMENTAL DESIGN

The experiment consists of two consecutive phases. In the first phase of the experiment, individuals participate in 10 rounds of a tax evasion game, while in the second phase they play a generalized gift exchange game for five rounds. In both phases, subjects always interact with the same group members. Indeed, at the beginning of the experiment, groups of five subjects are randomly formed and their composition is kept constant throughout the two phases.

In each round of the first phase of the experiment, each individual is assigned with a gross income expressed in ECUs (Experimental Currency Units). In particular, the gross income of each subject is an integer number that is randomly drawn from a uniform distribution between 100 and 240. Given her gross income, each subject chooses how much to declare to the central tax authority for tax payments, knowing that, on the declared amount, she will pay a flat tax rate of 30%. In each period, the declared income of one of the five group members is randomly selected (thus corresponding to a probability of 20%) and audited by the tax authority to verify its conformity with the gross income. If the subject under-declares her gross income, then, in addition to the due taxes on the gross income, she will pay a fine that is set equal to the evaded taxes (namely, the 30% of the difference between the gross and the declared income). If the subject fully declares her gross income, then the audit mechanism does not produce any further effect on her payoffs. Once the declaration choice is submitted, information about others' gross and declared incomes is provided. Finally, at the end of every period, each subject is informed about her payoffs and whether her choice has been selected for auditing.

With respect to the *NoWhistleblowing* treatment, in the *Whistleblowing* treatment the only difference is that once all declaration choices are submitted and information about others' gross and declared incomes is provided, each subject can blow the whistle on other group members. In particular, each subject is given the possibility to signal one of the four remaining group members to the tax authority. Then, the computer randomly selects one whistleblower. If the whistleblower effectively blew the whistle on one group member, then her choice is implemented, and the declared income of the signaled subject is audited. On the other hand, if the whistleblower decided not to blow the whistle on anybody, then, as in the *NoWhistleblowing* treatment, one of the group members is randomly selected and her declared income audited. Finally, no information is given to the audited subject on whether audit was due to random selection or to whistleblowing by other group members.

While most real-life leniency programs provide whistle-blowers with some indulgence for their own violations, our experimental design does not entail any bonuses in monetary or non-monetary form for those denouncing other tax evaders. This non-incentivized whistleblowing design is standard in tax evasion experiments [see, for instance, Bazart et al. (2020)], representing a conservative test to measure individuals' propensity for blowing the whistle: if we observe peer reporting without extra motives, we expect such a behavior to occur even with a higher frequency when individuals are positively incentivized to do so. In a similar vein, in our experiment the tax revenues plus the fines are not returned back to the common pool. Masclet et al. (2019) experimentally compared peer-reporting (whistleblowing) treatments with and without positive externalities and found no difference in whistleblowing frequency when participants were informed that collected taxes were used to purchase carbon credits.

In the second phase of the experiment, participants play the generalized gift exchange game. In each of the five periods of the second phase, each subject receives an endowment of 100 ECUs and chooses how much to send to the remaining group members. Whatever she sends is doubled by the experimenter and distributed equally among the remaining four group members. Therefore, social welfare is maximized if everyone sends the maximum amount to peers. This game is a variation of the standard public good game where an individual share of investment to a public good is not returned to the initial investor. Unlike a strain of the experimental literature that uses the sequential gift exchange game (Charness, 1996; Charness and Haruvy, 2002), in our experiment participants have to make their choices simultaneously. Additionally, instead of providing a gift to one single member of their group (Kanitsar, 2019), in our design each individual provides a gift to all other group members. Besides allowing for very simple and short instructions, our choice to implement a generalized gift exchange game characterized by simultaneous decisions was driven by our research objective, namely to investigate whether having experienced a tax evasion game with or without the possibility to blow the whistle on other group members affect the individual's beliefs about the overall level of cooperation of other players, and the individual decision to give as a consequence.

Apart from the inclusion of a whistleblowing stage, our experimental design also manipulates the provision of information about the content of the second phase before the tax evasion game is played. While in the *NoInformation* treatments, participants are informed about the second phase of the experiment only after completing the tax evasion game, in the *Information* treatments all participants learn, since the beginning of the experimental session, the content and instructions of the generalized gift exchange game of the second phase. The purpose of the information manipulation is to investigate whether tax evasion and attitude to blow the whistle are affected by subjects' awareness about the fact that, in the subsequent phase, they will participate with their group members in game in which results strongly depend on the level of ingroup trust. Even if anonymity is fully assured, whistleblowing might indeed undermine ingroup trust, making cooperation in the generalized gift exchange game more difficult to achieve. By anticipating these considerations, individuals might therefore

be more reluctant to blow the whistle on others, nullifying the effectiveness of whistleblowing in curbing tax evasion. The combination of the two manipulated dimensions generates results in a 2 × 2 design, and henceforth we will refer to the four treatments with the following labels: *NoWhistle_NoInfo*, *Whistle_NoInfo*, *NoWhistle_Info* and *Whistle_Info*.

Experimental Procedures

The experiment was run between September and December 2019 at the CERME (Center for Experimental Research in Management and Economics) laboratory, in Ca' Foscari University of Venice (Italy). 240 subjects (59% female), recruited through ORSEE (Greiner, 2015), participated in the experiment. Totally, we run 12 experimental sessions, with 60 subjects per treatment. Most of participants were undergraduate students (75.4%), enrolled in Economics (72.5%). Sessions were randomly assigned to treatments so that all participants within the same session were assigned to the same treatment and none participated in more than one treatment[3].

The experiment was computerized by using o-Tree (Chen et al., 2016). Each session lasted around 75 min (including time for reading the instructions aloud, answering private questions, and paying) and the average payment was 13.5 euro, including a show-up fee of 3 euro. Although subjects participated in 15 rounds, to avoid wealth effects, only one of the 15 rounds was effectively used to determine final payments. Specifically, at the end of the experiment, the experimenter first selected one of two phases by tossing a coin. Then, given the phase, the experimenter randomly picked one of the corresponding rounds.

RESULTS

In this section, we present our results. Given the partner-matching protocol of our experiment, we perform both: (i) two-sample Mann–Whitney tests (MW) and (ii) Somers' D median difference tests (Newson, 2002) at the group level, and we report results of (i) only unless the two tests give different results[4].

Tax Evasion Game

First, we describe the effect of whistleblowing on tax evasion.

In **Figure 1**, we show the proportions of gross incomes declared by subjects in the four treatments, both over the 10 periods of the first phase (left-handed Panel) and by period (right-handed Panel). Our data confirm that blowing the whistle is indeed effective in increasing the average proportion of reported income, being equal to 0.65 in the treatments in which subjects cannot signal others' choices to the tax authority (*NoWhistle_NoInfo* and *NoWhistle_Info*) and equal to 0.80 in the treatments including the whistleblowing stage (*Whistle_NoInfo* and *Whistle_Info*), with this difference being highly significant ($p = 0.001$, MW). The same result is observed when making a pairwise comparison between *Whistle_Info* and *NoWhistle_Info* ($p = 0.038$, MW; $p = 0.158$, Somers' D), as well as between *Whistle_NoInfo* and *NoWhistle_NoInfo* ($p = 0.021$, MW). Additionally, the decrease in the proportion of the reported income across periods is starker in absence of the deterrence mechanism than in treatments including the whistleblowing stage.

Finally, we see no effect of the information manipulation on the effectiveness of whistleblowing (*Whistle_Info* vs. *Whistle_NoInfo*, p = 0.862, MW).

Figure 2 provides a more detailed picture of the frequencies of the relative reported share of income in each treatment. We observe that individuals are more likely to report an income equals to zero when whistleblowing is not allowed than in the *Whistle_NoInfo* and *Whistle_Info* treatments.

As it can be seen in **Table 1**, where we report the proportion of full compliers, intermediary compliers and full non-compliers, the most striking difference across treatments is indeed the substantial fall of full non-compliers as soon as the possibility to blow the whistle on others is introduced (from 11% and 18% respectively in the *NoWhistle_NoInfo and NoWhistle_Info* treatments to 2.8% and 3.5% in the *Whistle_NoInfo* and *Whistle_Info* treatments).

In **Table 2**, we report parametric results from a series of Multilevel models, with standard errors that are clustered at both the group and subject level, using the proportion of gross incomes declared by subjects in each of the 10 rounds of the first phase as dependent variable[5].

In Model 1, *Endowment* takes a value from 100 to 240 (in integer numbers). *Info* is equal to one in the treatments in which information about the second phase of the experiment was provided prior to the beginning of the first phase and 0 otherwise. Similarly, *Whistleblowing* takes a value of 1 in the treatments in which participants were allowed to blow the whistle on other ingroup members in the tax evasion game of the first part of the experiment, and 0 otherwise. *Period* is a time counter, and it is introduced in the regressions to account for the effect of experience in the tax evasion game. Model 2 is augmented by adding the interaction term *InfoXWhistleblowing*.

Model 3 includes participants' gender and information about the previous period. In particular, *Proportion_report_prev_period* stands for the individual proportion of income reported in the previous period, while *Audited_prev_period* consists in a binary variable indicating whether, in the previous period, the participant was audited or not.

Finally, in Model 4, we add *Economics*, which takes a value of 1 if the participants' field of study is Economics and 0 otherwise, as well as a series of categorical variables extracted from the post experimental questionnaire[6]. Previous studies (Jackson and Milliron, 1986; Richardson, 2006) have indeed shown how both

[3]In **Supplementary Appendix Table A3** in the **Supplementary Appendix** we report the per-treatment main socio-demographic characteristics of our sample.

[4]When performing the Mann–Whitney U-test, we average data at the group level and treat each group as an independent observation. The rank-order statistics Somers' D looks at the individuals' choices accounting for the presence of clusters at the group level (each experimental session included groups) in the data.

[5]See **Supplementary Appendix Table A.1** in the **Supplementary Appendix** for the results of a series of Tobit models (with left and right censoring at 0 and 1, respectively) with errors clustered at the group level. Results remain virtually unchanged across specifications.

[6]The questionnaire (originally written in Italian) is reported in the **Supplementary Appendix**.

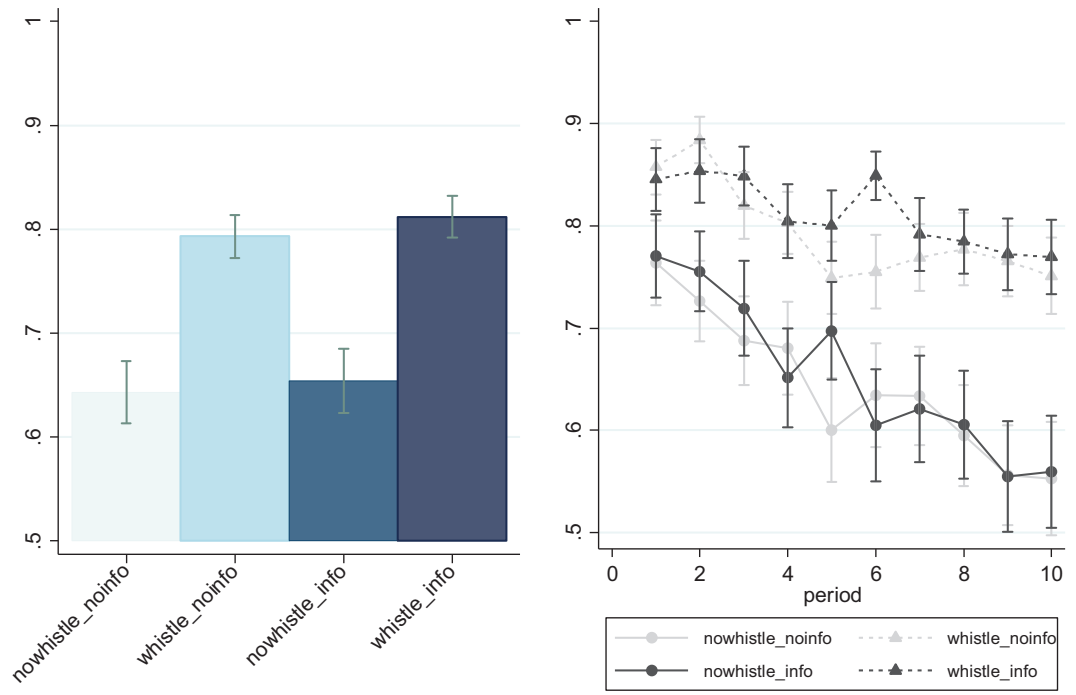

FIGURE 1 | Proportions of gross incomes declared by subjects in the tax evasion game, by treatment (left-handed Panel) e by treatment and period (right-handed Panel), *N* = 240. Error bars, mean ± SEM.

"demographic (i.e., gender), "economic" (such as income level and marginal tax rates) and "behavioral" (such as fairness and tax morale) characteristics can motive tax evasion so we controlled these factors through a series of independent variables. More specifically, to take into consideration that members of high income families might be more likely to evade taxes as well as the effects of increasing marginal tax rates on income declarations, we include *Income_family*, *Relative_wealth* and *Perceived_tax* in our regression. Both *Income_family* and *Relative_wealth* take a value from 1 (very low) to 10 (very high) and define the participant's perception of the income of her own family as well as her perception of the relative position of the family's income with respect to the average Italian family, respectively, while *Perceived_tax* takes a value from 1 to 12 and expresses the perceived tax rate paid by the participant, in 5% income brackets (with 1 being "less than 10%" and 12 being "above 60%"). On the same vein, *High_tax* measure the strength of the subject's belief on whether the tax rate affects individual willingness to pay taxes.

Given the negative relationship with fairness and tax evasion (Richardson, 2006), we also add *Fair_tax*, which indicates which tax rate would be considered as fair. Attitude toward risk might affect tax evasion when in presence of audit schemes and penalties, the variable *Risk_level* thus measures individual risk aversion and takes a value from 0 to 10, with higher numbers expressing lower levels of risk aversion. In order to control for the subject's attitude toward tax evasion, we include *Risk_audit, Reciprocal_evasion* and *Ineff_gov* as covariates in the regression. The three variables indicate how strongly the subject agrees on a 10-point scale (with 1 indicating complete disagreement and 10 complete agreement) with the statement that citizens do not pay taxes if they perceive that the audit risk is low, other citizens do not pay taxes, and collected taxes are inefficiently implemented, respectively. Expecting tax morale to possibly negatively affect tax evasion (Torgler, 2003) we include as regressor*Tax_morality, which* measures the strength of the subject's belief on whether morality affects individual willingness to pay taxes, while we also control for the level of perceived trust (*Trust*) and concern about helping others as a moral duty (*Help_others*).

From Model 1, whistleblowing significantly increases the proportion of reported income and, therefore, represents a valid instrument to limit tax evasion[7]. Differently, the effect of providing information about the second phase of the experiment before letting subjects to declare their income in the tax evasion game does not affect the amount of evaded taxes. Looking at Models 2 to 4, the interaction term between Whistleblowing and Info never reaches significance, meaning that the proportion of income reported by participants when they are allowed to blow the whistle is not affected by being aware about the gift exchange game in the second phase of the experiment. Although the coefficient of the endowment is significant at the 5% level in

[7]In **Supplementary Appendix Table A4** in the **Supplementary Appendix** we provide a more detailed analysis of the whistleblowing behavior, defined as the per period number of whistleblower's signals (from 0 to 4) on a group member as a function of her relative proportion of reported income within the group.

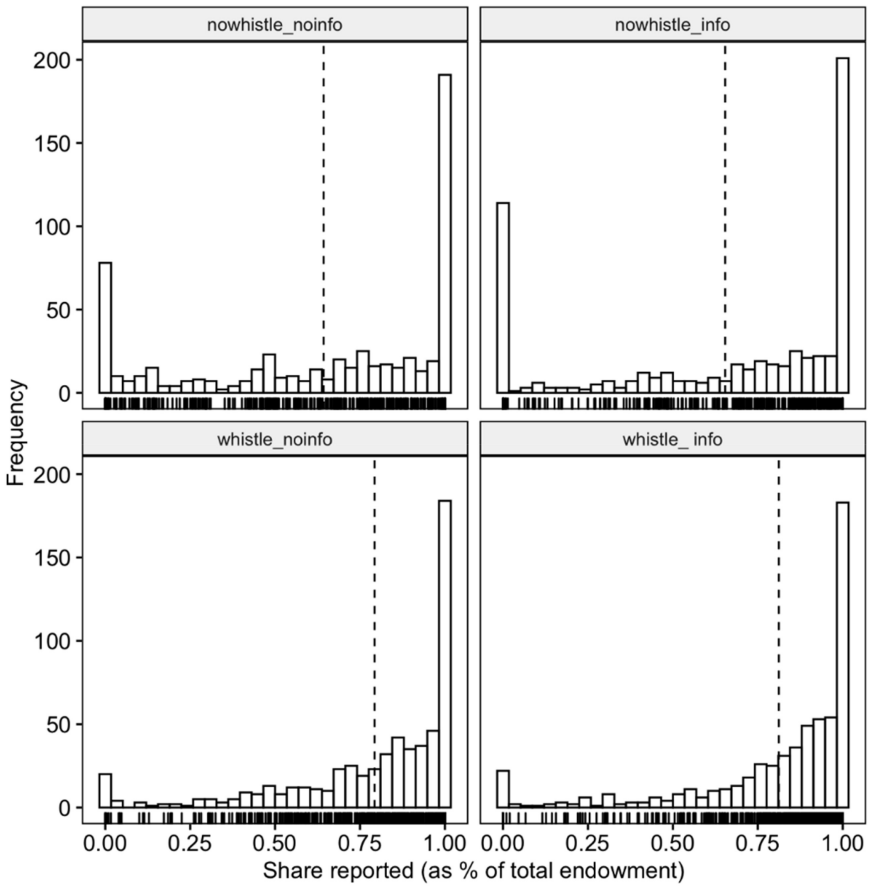

FIGURE 2 | Frequency of proportion of reported income per treatment.

TABLE 1 | Proportion of full compliers, intermediary and full non-compliers per treatment.

Treatment	Full compliers	Intermediary compliers	Full non-compliers
nowhistle_noinfo	29.3%	59.3%	11.3%
nowhistle_info	32.7%	49.3%	18.0%
whistle_noinfo	27.8%	69.3%	2.8%
whistle_ info	27.0%	69.5%	3.5%

Model 1, it presents a small magnitude, suggesting that it exerts only limited effects on participants' decision to evade taxes.

As participants gain experience in the tax evasion game, they are less likely to fully report their income, as shown by the significant and negative coefficient of the time trend in all models.

Model 3 further analyses the dynamic pattern followed by choices in the tax evasion game. The proportion of reported income is positively correlated across periods and being audited in the previous period decreases the amount evaded in the current one. As expected, the level of risk aversion is significant and negatively correlated with tax evasion: an increase of one unit in risk propensity decreases the proportion of reported income by about 0.02.

In order to better investigate the effects of being audited on the subsequent choices in the tax evasion game, the last two columns of **Table 2** focus on the sessions with and without whistleblowing, separately. We find evidence of the bomb-crater effect of tax audits (Mittone et al., 2017) only in the *NoWhistleblowing* treatments while, as expected, in the Whistleblowing sessions being audited in the previous period significantly increases the proportion of income reported in the current period, as it suggests participants that other in-group members might have blown the whistle on them. Interestingly, as shown by the coefficient of *Help_others* in the model focusing on the sessions with Whistleblowing, the more individuals think that helping others represents a moral duty, the higher the proportion of income reported, underlying the importance of moral values in determining tax evasion.

Generalized Gift Exchange Game

Our aim is to identify whether allowing individuals to blow the whistle on others in the tax evasion game and the information about the subsequent phase of the experiment exerted any effect on their contribution decisions in the generalized gift exchange game in the second phase. On average, participants contributed 24.75 tokens in the *Whistleblowing* treatments and 33.13 tokens

TABLE 2 | The determinants of the proportion of income reported in the tax evasion game: Multilevel models, with standard errors clustered at both at the group and at the subject level.

Independent variables	Model 1	Model 2	Model 3	Model 4	Whistle	NoWhistle
Info	0.015	0.010	0.011	0.025	0.021	0.042
	(0.040)	(0.057)	(0.052)	(0.050)	(0.049)	(0.048)
Whistleblowing	0.155***	0.150***	0.138***	0.137***		
	(0.040)	(0.057)	(0.052)	(0.050)		
Endowment	−0.0003062**	−0.0003063	−0.0003618	−0.0003641	−0.0001289	−0.0006132
	(0.0001207)	(0.0001207)	(0.0001287)	(0.0001286)	(0.0001391)	(0.0002104)
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Period	−0.017***	−0.017***	−0.015***	−0.015***	−0.009***	−0.020***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)
InfoXWhistleblowing		0.010	0.012	0.001		
		(0.080)	(0.073)	(0.070)		
Female			0.111***	0.061**	0.003	0.072
			(0.026)	(0.027)	(0.024)	(0.044)
Proportion_report_prev_period			0.105***	0.104***	0.190***	0.099***
			(0.022)	(0.022)	(0.032)	(0.031)
Prev_audited			−0.054***	−0.053***	0.047***	−0.149***
			(0.013)	(0.013)	(0.015)	(0.022)
Economics				−0.058**	−0.010	−0.091*
				(0.028)	(0.022)	(0.047)
Income_family				0.007	0.001	0.016
				(0.011)	(0.009)	(0.020)
Relative_wealth				0.003	0.009	−0.002
				(0.012)	(0.010)	(0.019)
Perceived_tax				−0.012*	0.011**	−0.038***
				(0.007)	(0.006)	(0.012)
Fair_tax				0.020**	0.0000744	0.046***
				(0.009)	(0.0082447)	(0.015)
Risk_audit				0.001	0.005	−0.002
				(0.006)	(0.005)	(0.010)
Risk_level				−0.019***	−0.019***	−0.027***
				(0.005)	(0.004)	(0.009)
Reciprocal_evasion				−0.008	−0.001	−0.024**
				(0.007)	(0.006)	(0.011)
Tax_Morality				−0.004	−0.005	−0.007
				(0.005)	(0.004)	(0.009)
Ineff_gov				0.006	−0.005	0.006
				(0.006)	(0.005)	(0.011)
High_tax				−0.005	−0.011**	−0.000
				(0.006)	(0.005)	(0.011)
Trust					0.003	−0.026**
					(0.006)	(0.011)
Help_others					0.029***	0.011
					(0.007)	(0.011)
Constant	0.787***	0.790***	0.656***	0.835***	0.558***	1.197***
	(0.041)	(0.046)	(0.049)	(0.090)	(0.083)	(0.175)
Observations	2400	2400	2160	2160	1080	1080
Log likelihood	−141.753	−141.746	−131.309	−117.171	257.351	−218.791
Wald chi2	133.505	133.523	169.087	211.584	161.056	170.058
p	0.000	0.000	0.000	0.000	0.000	0.000

Table 2 *reports estimates of a series of Multilevel regression models. The dependent variable is the reported proportion of income in each period of the tax evasion game. Clustered standard errors at the group level and at the individual level appear in parentheses. ***, ** and * indicate significance at the 1% level, 5% level and 10% level, respectively.*

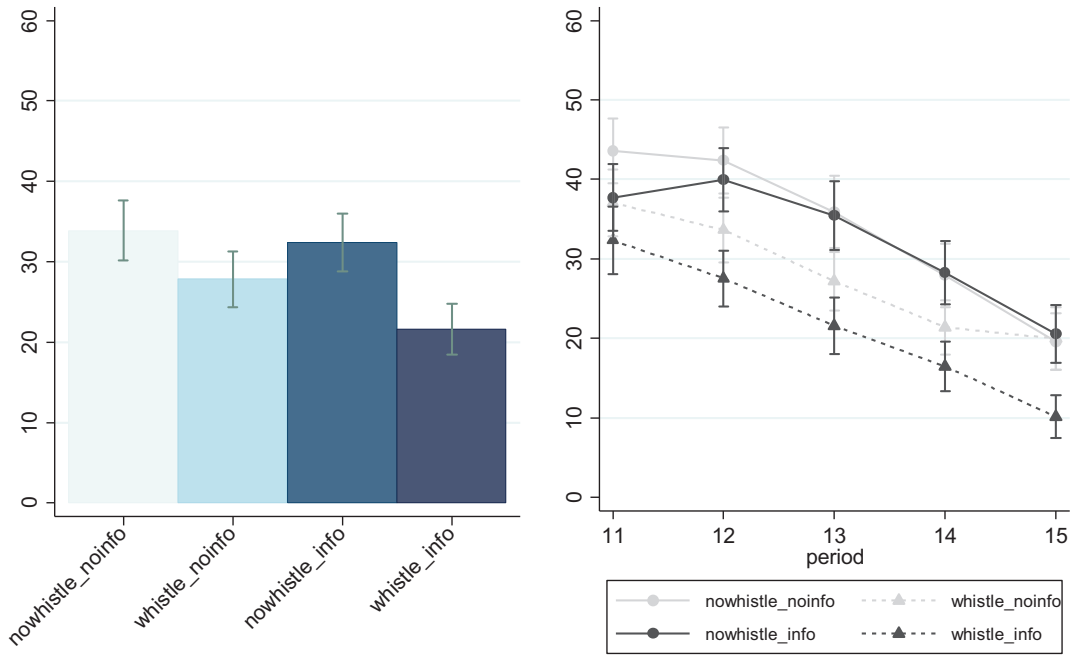

FIGURE 3 | Average contributions in the Generalized Gift Exchange Game by treatment (left-handed Panel) e by treatment and period (right-handed Panel). Error bars, mean ± SEM.

in the *NoWhistleblowing* treatments. Thus, whistleblowing tends to reduce cooperation in the subsequent game, though this effect is not significant (p = 0.143, MW; p = 0.058, Somers' D-test, 48 clusters).

In **Figure 3**, we report the average contribution in the *Whistleblowing* and *NoWhistleblowing* treatments, respectively. Allowing individuals to blow the whistle on others results in a slight reduction of contributions in the second phase of the experiment, in particular in the setting in which subjects receive information about the generalized gift exchange game before making their tax evasion choices (p = 0.133, MW; p = 0.078, Somers' D). Instead, we document no significant effects in the setting in which the information about the task in the second phase is provided only at the end of the tax evasion game (p = 0.453, MW).

In **Table 3**, we report a series of multilevel models with standard errors that are clustered at both the group and subject level and where the dependent variable is the number of tokens contributed to the Generalized Gift Exchange Game[8].

In order to investigate whether allowing individuals to blow the whistle on others in the tax evasion game affects their contributions in the second phase, in Model 1 we include *Whistleblowing*, *Info* and *Period* as regressors. We observe that whistleblowing is indeed marginally significant in decreasing ingroup contributions in the gift exchange game. However, the effect disappears when information about the second phase of the experiment is not provided at the beginning of the experimental session, as shown by the coefficient of the variable *Whistleblowing* in Model 2.

In Model 3, we also add *Contribution_prev_period,* which stands for the individual contribution in the previous period, and *Group_contribution_prev_period*, that consists in a continuous variable expressing the average contributions of the remaining 4 group members in the previous period. We find a strong evidence in favor of in group reciprocity, whereby the average contribution made by a subject increases in the average number of tokens contributed by group members in the previous period. *Proportion_report_1st_part*, *Group_proportion_report_1st_part* and *N_audited* are built upon subjects' behavior in the tax evasion game, and respectively indicate subject's average reported income, the average income reported by the remaining 4 group members, and the number of times the

[8]See **Supplementary Appendix Table A.2** in **Supplementary Appendix A** for the results of a series of Tobit models, left censored at zero, with clustered standard errors at the group level. Results are almost unchanged. The only remarkable difference relies on the effect of N_audited. In the *Whistleblowing sessions*, the higher the number of times an individual was audited in the tax evasion game (and the higher the number of whistleblowers' signals on the subject), the lower her contributions in the gift exchange game is. The opposite effect is instead observed in the *NoWhistleblowing* sessions, suggesting that being audited might have an educative effect on future cooperation.

TABLE 3 | Multilevel regressions. Amount contributed in the Generalized Gift Exchange game.

Independent variables	Model 1	Model 2	Model 3	Model 4	Whistle	NoWhistle
Whistleblowing	−8.408**	−6.030	1.026	1.241		
	(4.075)	(5.743)	(2.301)	(2.305)		
Info	−3.865	−1.487	1.302	1.816	−3.572*	1.532
	(4.075)	(5.743)	(2.143)	(2.155)	(2.167)	(2.173)
Period	−5.261***	−5.261***	−2.416***	−2.430***	−1.970*	−3.479***
	(0.439)	(0.439)	(0.712)	(0.710)	(1.029)	(0.982)
InfoXWhistleblowing		−4.757	−3.649	−4.272		
		(8.122)	(3.030)	(3.039)		
Contribution_prev_period			0.510***	0.504***	0.448***	0.523***
			(0.026)	(0.026)	(0.038)	(0.036)
Group_contribution_prev_period			0.260***	0.263***	0.157**	0.309***
			(0.044)	(0.044)	(0.072)	(0.056)
Proportion_report_1st_part			0.480	−0.882	−4.200	−0.634
			(3.628)	(3.694)	(7.766)	(4.344)
Group_proportion_report_1st_part			−7.800	−7.419	6.839	−14.775**
			(5.347)	(5.382)	(9.614)	(6.931)
Female			1.217	0.713	0.885	−0.647
			(1.606)	(1.625)	(2.277)	(2.328)
N_audited			0.093	0.064	−1.266	1.312
			(0.632)	(0.639)	(0.944)	(0.958)
Economics				−2.393	−6.189**	1.150
				(1.757)	(2.439)	(2.538)
Trust				0.080	0.524	0.026
				(0.436)	(0.604)	(0.642)
Help_others				0.700	0.593	0.663
				(0.461)	(0.669)	(0.637)
Tax_morality				−0.206	−0.079	−0.413
				(0.294)	(0.391)	(0.448)
Constant	103.457***	102.267***	39.025***	38.174***	32.499*	52.396***
	(6.710)	(7.004)	(11.123)	(11.748)	(17.167)	(16.167)
Observations	1200	1200	960	960	480	480
Log likelihood	−5581.4845	−5581.3136	−4390.5422	−4388.0923	−2184.389	−2192.5022
Wald Chi2	148.792	149.172	646.418	654.638	229.456	423.267
p	0.000	0.000	0.000	0.000	0.000	0.000

Table 3 *presents the coefficients from a series of Tobit regressions left-censored at zero. The dependent variable is the amount contributed in each period of the generalized gift exchange game. Clustered standard errors at the session level appear in parentheses. ***, ** and * indicate significance at the 1% level, 5% level and 10% level, respectively.*

participant was audited. Estimates indicate that results in the first phase of the experiment do not exert significant effects on the decisions in the gift exchange game. Similarly, Model 4 suggests that both the individual level of trust and willingness to help others do not significantly affect participants' contributions.

Finally, in the last two columns of **Table 3**, we restrict our analysis on the *Whistleblowing* and *NoWhistleblowing* treatments. It is worth noticing that, when whistleblowing is introduced, providing information about the gift exchange game before playing the tax evasion game decreases contributions in the second phase, as shown by the negative and marginally significant coefficient of *Info*. Surprisingly, in the *NoWhistleblowing* sessions, the average income reported by the other 4 group members in the tax evasion game has a negative effect on individual contribution in the gift exchange game.

DISCUSSION

In this paper, we investigated the interaction between ingroup cooperation and whistleblowing. Stemming from the previous literature, we conjectured that whistleblowing may have exerted some unintended adverse effects, undermining the group morale, and compromising its ability for collective actions. If that would be the case, then even the positive effect the whistleblowing might have on tax payments could be outweighed by negative externalities of such institution.

Our results reject the existence of adverse spillover effects from the tax evasion game to the generalized gift exchange game: although the whistleblowing somewhat discouraged contributions in the generalized gift exchange game, when controlling for other factors this difference is not significantly different from zero.

Moreover, the main driving force behind our experiment was to observe whether the shadow of the future cooperation deter participants from blowing the whistle on tax evaders. Indeed, if whistleblowing is perceived as that, it would be the case that this can be one of the mechanisms that explain the reluctance of agents to blow the whistle. Being aware that whistleblowing would suppress the ingroup cooperation, the rational profit-maximisers would avoid to report tax evaders within their group. The results of our experiments do not confirm this intuition.

These results are good news for policy makers who try to promote whistleblowing as a means of horizontal control to fight the tax evasion or other norm-violating behavior. However, the lack of the effect may mean that we need to consider some other uncounted factors. For instance as Kennedy and Schweitzer (2018) have shown, whistleblowers are generally perceived as more trustworthy than individuals who stayed idle. Since these two effects push the cooperation rate to the opposite direction the net effect is hard to predict.

Additionally, as in most experimental studies, our study abstracts away from many elements of real life in order to cleanly identify the specific links between tax evasion, whistleblowing and cooperation. While in our experiment tax evaders are asked to pay a fine if they got caught, it would be interesting to allow participants to track the identities of ingroup members from round to round so to investigate the role of reputational considerations when evading taxes. Similarly, in our experiment, retaliation against whistleblowers is not possible, a phenomenon that might indeed refrain individuals from denouncing others' wrongdoing. Finally, in our experiment the money collected through taxes are not meant to finance the provision of a public good. In such a situation, the benefits from higher levels of tax compliance due to whistleblowing might outweigh the possible decline in future cooperation. Future studies might evaluate the effects of these additional factors, in a framework where adopting a broader view in evaluating the efficacy of an institution allows to inform policies on the complex dynamics between tax evasion, whistle blowing and ingroup cooperation.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

ACKNOWLEDGMENTS

Financial support from "Fondi Primo Insediamento", University of Venice "Ca' Foscari" is gratefully acknowledged. The contribution of Philipp Chapkovski was developed within the framework of the Basic Research Program at the National Research University Higher School of Economics (HSE).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2021.732248/full#supplementary-material

REFERENCES

Alm, J. (2019). What motivates tax compliance? *J. Econ. Surv.* 33, 353–388. doi: 10.1111/joes.12272

Alm, J., Bloomquist, K. M., and McKee, M. (2017). When you know your neighbour pays taxes: information, peer effects and tax compliance. *Fisc. Stud.* 38, 587–613. doi: 10.1111/1475-5890.12111

Amir, E., Lazar, A., and Levi, S. (2018). The deterrent effect of whistleblowing on tax collections. *Eur. Account. Rev.* 27, 939–954. doi: 10.1080/09638180.2018.1517606

Antinyan, A., Corazzini, L., and Pavesi, F. (2020). Does trust in the government matter for whistleblowing on tax evaders? Survey and experimental evidence. *J. Econ. Behav. Organ.* 171, 77–95. doi: 10.1016/j.jebo.2020.01.014

Apesteguia, J., Dufwenberg, M., and Selten, R. (2007). Blowing the whistle. *Econ. Theory* 31, 143–166. doi: 10.1007/s00199-006-0092-8

Bazart, C., Beaud, M., and Dubois, D. (2020). Whistleblowing vs. random audit: an experimental test of relative efficiency. *Kyklos* 73, 47–67. doi: 10.1111/kykl.12215

Breuer, L. (2013). Tax compliance and whistleblowing–The role of incentives. *Bonn. J. Econ.* 2, 7–44.

Buckenmaier, J., Dimant, E., and Mittone, L. (2020). Effects of institutional history and leniency on collusive corruption and tax evasion. *J. Econ. Behav. Organ.* 175, 296–313. doi: 10.1016/j.jebo.2018.04.004

Charness, G. (1996). *Attribution and Reciprocity in a Simulated Labor Market: An Experimental Investigation. Economics Working Papers 283, Department of Economics and Business*. Fabra: Universitat Pompeu.

Charness, G., and Haruvy, E. (2002). Altruism, equity, and reciprocity in a gift-exchange experiment: an encompassing approach. *Games Econ. Behav.* 40, 203–231. doi: 10.1016/S0899-8256(02)00006-4

Chen, D. L., Schonger, M., and Wickens, C. (2016). oTree—An open-source platform for laboratory, online, and field experiments. *J. Behav. Exp. Finance* 9, 88–97. doi: 10.1016/j.jbef.2015.12.001

Chen, Z., and Rey, P. (2013). On the design of leniency programs. *J. Law Econ.* 56, 917–957. doi: 10.1086/674011

dAdda, G., Capraro, V., and Tavoni, M. (2017). Push, don't nudge: behavioral spillovers and policy instruments. *Econ. Lett.* 154, 92–95. doi: 10.1016/j.econlet.2017.02.029

Engl, F., Riedl, A., and Weber, R. A. (2020). Spillover effects of institutions on cooperative behavior, preferences, and beliefs. *Prefer. Beliefs August* 3:2020. doi: 10.2139/ssrn.3666456

Fonseca, X., Lukosch, S., and Brazier, F. (2019). Social cohesion revisited: a new definition and how to characterize it. *Innov. Eur. J. Soc. Sci. Res.* 32, 231–253. doi: 10.1080/13511610.2018.1497480

Gächter, S., Starmer, C., and Tufano, F. (2017). *Revealing the Economic Consequences of Group Cohesion*. Nottingham, UK: University of Nottingham.

Galbiati, R., Henry, E., and Jacquemet, N. (2018). Dynamic effects of enforcement on cooperation. *Proc. Natl. Acad. Sci. U.S.A.* 115, 12425–12428. doi: 10.1073/pnas.1813502115

Galeotti, F., Maggian, V., and Villeval, M. C. (2021). Fraud deterrence institutions reduce intrinsic honesty. *Econ. J.* 131, 2508–2528. doi: 10.1093/ej/ueab018

Ghesla, C., Grieder, M., and Schmitz, J. (2019). Nudge for good? Choice defaults and spillover effects. *Front. Psychol.* 10:178. doi: 10.3389/fpsyg.2019.00178

Greiner, B. (2015). Subject pool recruitment procedures: organizing experiments with ORSEE. *J. Econ. Sci. Assoc.* 1, 114–125. doi: 10.1007/s40881-015-0004-4

Heyes, A., and Kapur, S. (2009). An economic model of whistle-blower policy. *J. Law Econ. Organ.* 25, 157–182. doi: 10.1093/jleo/ewm049

Hildreth, J. A. D., and Anderson, C. (2018). Does loyalty trump honesty? Moral judgments of loyalty-driven deceit. *J. Exp. Soc. Psychol.* 79, 87–94. doi: 10.1016/j.jesp.2018.06.001

Hildreth, J. A. D., Gino, F., and Bazerman, M. (2016). Blind loyalty? When group loyalty makes us see evil or engage in it. *Organ. Behav. Hum. Decis. Process.* 132, 16–36. doi: 10.1016/j.obhdp.2015.10.001

Hinloopen, J., and Soetevent, A. R. (2008). Laboratory evidence on the effectiveness of corporate leniency programs. *RAND J. Econ.* 39, 607–616. doi: 10.1111/j.0741-6261.2008.00030.x

Internal Revenue Service. (2019). *Federal Tax Compliance Research: Tax Gap Estimates for Tax Years 2011-2013*. Ogden, UT: Internal Revenue Service.

Jackson, B. R., and Milliron, V. C. (1986). Tax compliance research: findings, problems, and prospects. *J. Account. Lit.* 5, 125–165.

Kanitsar, G. (2019). Solidarity through punishment: an experiment on the merits of centralized enforcement in generalized exchange. *Soc. Sci. Res.* 78, 156–169. doi: 10.1016/j.ssresearch.2018.12.012

Kennedy, J. A., and Schweitzer, M. E. (2018). Building trust by tearing others down: when accusing others of unethical behavior engenders trust. *Organ. Behav. Hum. Decis. Process.* 149, 111–128. doi: 10.1016/j.obhdp.2018.10.001

Lau, D. C., and Liden, R. C. (2008). Antecedents of coworker trust: leaders' blessings. *J. Appl. Psychol.* 93, 1130. doi: 10.1037/0021-9010.93.5.1130

Loyens, K. (2013). Towards a custom-made whistleblowing policy. Using grid-group cultural theory to match policy measures to different styles of peer reporting. *J. Bus. Ethics* 114, 239–249. doi: 10.1007/s10551-012-1344-0

Masclet, D., Montmarquette, C., and Viennot-Briot, N. (2019). Can whistleblower programs reduce tax evasion? Experimental evidence. *J. Behav. Exp. Econ.* 83:101459. doi: 10.1016/j.socec.2019.101459

Mittone, L., Panebianco, F., and Santoro, A. (2017). The bomb-crater effect of tax audits: beyond the misperception of chance. *J. Econ. Psychol.* 61, 225–243. doi: 10.1016/j.joep.2017.04.007

Morris, E. W. (2010). "Snitches end up in ditches" and other cautionary tales. *J. Contemp. Crim. Justice* 26, 254–272. doi: 10.1177/1043986210368640

Near, J. P., and Miceli, M. P. (1985). Organizational dissidence: the case of whistle-blowing. *J. Bus. Ethics* 4, 1–16. doi: 10.1007/BF00382668

Newson, R. (2002). Parameters behind "nonparametric" statistics: Kendall's tau, Somers' D and median differences. *Stata J.* 2, 45–64. doi: 10.1177/1536867X0200200103

Nyreröd, T., and Spagnolo, G. (2021). Myths and numbers on whistleblower rewards. *Regul. Gov.* 15, 82–97. doi: 10.1111/rego.12267

Peysakhovich, A., and Rand, D. G. (2016). Habits of virtue: creating norms of cooperation and defection in the laboratory. *Manag. Sci.* 62, 631–647. doi: 10.1287/mnsc.2015.2168

Reuben, E., and Stephenson, M. (2013). Nobody likes a rat: on the willingness to report lies and the consequences thereof. *J. Econ. Behav. Organ.* 93, 384–391. doi: 10.1016/j.jebo.2013.03.028

Richardson, G. (2006). Determinants of tax evasion: a cross-country investigation. *J. Int. Account. Audit. Tax.* 15, 150–169. doi: 10.1016/j.intaccaudtax.2006.08.005

Torgler, B. (2003). *Tax Morale: Theory and Empirical Analysis of Tax Compliance*. Doctoral dissertation, University of Basel, Basel.

Trevino, L. K., and Victor, B. (1992). Peer reporting of unethical behavior: a social context perspective. *Acad. Manage. J.* 35, 38–64. doi: 10.5465/256472

Wallmeier, N. (2019). *The Hidden Costs of Whistleblower Protection*. Mimeo, NY: University of Hamburg. Available online at: https://ssrn.com/abstract=3111844.

Woldoff, R. A., and Weiss, K. G. (2010). Stop snitchin': exploring definitions of the snitch and implications for urban Black communities. *J. Crim. Justice Pop. Cult.* 17, 184–223.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychology

ORIGINAL RESEARCH
published: 21 October 2021
doi: 10.3389/fpsyg.2021.761168

Esteemed Colleagues: A Model of the Effect of Open Data on Selective Reporting of Scientific Results

*Eli Spiegelman**

Economics and Social Sciences, CEREN, EA 7477, Burgundy School of Business – Université Bourgogne Franche-Comté, Dijon, France

Open data, the practice of making available to the research community the underlying data and analysis codes used to generate scientific results, facilitates verification of published results, and should thereby reduce the expected benefit (and hence the incidence) of p-hacking and other forms of academic dishonesty. This paper presents a simple signaling model of how this might work in the presence of two kinds of cost. First, reducing the cost of "checking the math" increases verification and reduces falsification. Cases where the author can choose a high or low verification-cost regime (that is, open or closed data) result in unraveling; not all authors choose the low-cost route, but the best do. The second kind of cost is the cost to authors of preparing open data. Introducing these costs results in that high- and low-quality results being published in both open and closed data regimes, but even when the costs are independent of research quality open data is favored by high-quality results in equilibrium. A final contribution of the model is a measure of "science welfare" that calculates the ex-post distortion of equilibrium beliefs about the quality of published results, and shows that open data will always improve the aggregate state of knowledge.

Keywords: open data, signaling game model, research ethics, esteem, replication crisis, replication crisis in psychology, academic dishonesty behaviors, academic dishonesty and misconduct

OPEN ACCESS

Edited by:
Ismael Rodriguez-Lara, University of Granada, Spain

Reviewed by:
Joaquin Gomez-Minambres, Lafayette College, United States
Gonzalo Olcina, University of Valencia, Spain

***Correspondence:**
Eli Spiegelman
eli.spiegelman@bsb-education.com

Specialty section:
This article was submitted to Personality and Social Psychology, a section of the journal Frontiers in Psychology

Received: *19 August 2021*
Accepted: *27 September 2021*
Published: *21 October 2021*

Citation:
Spiegelman E (2021) Esteemed Colleagues: A Model of the Effect of Open Data on Selective Reporting of Scientific Results. Front. Psychol. 12:761168. doi: 10.3389/fpsyg.2021.761168

INTRODUCTION

Experimental work in the social sciences is currently undergoing a replication crisis (Ioannidis, 2005; Stevens, 2017; Obels et al., 2020). The Open Science Collaboration (2015) successfully replicated 36 out of 100 experiments published in high-ranking psychology journals; Camerer et al. (2016, 2018) find reproducibility rates of around 61% in economics experiments. In a survey of 1,500 scientists, Baker (2016) found that 70% had failed to replicate another researcher's results, and 50% had failed to replicate their own. There are many potential sources of these phenomena, but one of the most direct is that researchers are being less than completely forthright about the nature of the results they publish. Their incentives to do so are clear: on the "demand side," tenure and promotions, successful grant proposals, and even informal esteem from colleagues are all examples of how researchers get some utility from the perception of having done important work, whether or not such perceptions are rigorously supported by the data. Furthermore, on the "supply side," the inherent complexity of interpreting empirical data implies that the "true" result is rarely completely unambiguous. Even setting aside cases (which nevertheless do exist) of outright fraud or fabrication of data, it may often be possible for otherwise principled and honest researchers to lean on their

results as it were, engaging in gentle falsification, or "p-hacking," for instance through selective analysis or reporting of results.

Even "partial dishonesty" can have negative social effects, as it generates an unwarranted image of the state of scientific knowledge. For instance, gender differences in risk aversion, with females less willing to take risks than males, long represented a "stylized fact" that emerged from studies designed to address other questions. Publishing confirmatory results lent credibility to such papers by showing that they fit with the existing body of knowledge, but also perpetuated a particular description of the social nature of gender. However, a meta-analysis by Filippin and Crosetto (2016) subsequently showed that the effect was, if not illusory, then much more fragile than had previously been estimated. Subsequent verification of previous work in this sense represents scientific progress and at the same time a progressive view of gender.

Perhaps the central assumption of this paper is that such "fact-checking," systematically applied to the accumulated body of published results, should act as a kind of disciplining tool on what gets published in the first place: researchers may be tempted to inflate the "importance" of their results in order to acquire a certain esteem from the research or wider community, but a downward revision of the importance induces an esteem penalty, so it is preferable to honestly present results of minor importance, rather than being caught in such inflation or falsification. A potential lever to encourage the disciplining verification is *open data*, which refers to the practice of making the underlying data and analysis codes used to generate results available to the research community, along with the paper itself. This clearly facilitates verification; so long as it also increases the probability of some third party actually engaging in such verification, it should thereby reduce the expected benefit (and hence the incidence) of p-hacking and other forms of academic dishonesty. The very top journals in many fields, for instance in economics, psychology and marketing, require open publication of data and analysis codes with the paper. However, the requirement is far from systematic. For instance, at the time of this writing 9 of the top 20 economics journals merely "encouraged" open data submissions. Furthermore, such encouragement is not generally effective (Tenopir et al., 2011). Alsheikh-Ali et al. (2011) investigated 500 published papers coming from high impact journals from various scientific fields, finding that only 9% had their raw data stored online publicly. Womack (2015) reached a similar conclusion; from a sample of 4,370 papers published in 2014 in the highest impact journals, only 13% made their data publicly available online.

The idea that researchers are motivated to publish "important" results due to a mechanism of esteem indicates a link to signaling models, which form the basis of the theoretical construction in this paper. The signal structure has several inter-related layers, which are developed sequentially. First, the presentation of the published paper itself should be considered as a signal of the underlying quality of the scientific result obtained. This is modeled as a relatively "cheap" signal: "authors" in the model are privately informed of the quality of their results, and can present them as whatever they choose. "Readers" are motivated to identify dishonest presentation, although verification is costly. Section 2 shows that in equilibrium, as might be expected, the lower this cost, the more verification—and the less falsification—occurs. The second layer of signaling is the choice of open or closed data, that is, of high or low verification costs. Intuitively, a "nothing to hide" principle choosing high verification costs should be taken as a bad signal, and indeed Section 3 shows that a case where the author can choose a high or low verification cost regime (that is, open or closed data) results in unraveling. All high-quality results, which require no falsification, will be published in open data, which allows readers to identify any result published in closed data as being of low quality, making falsification impossible.

These results are promising, but seem to conflict with the empirical patterns described above in which adoption of open data is very low. In this regard, a potentially important second kind of cost not included in the model is the cost to authors of securely and accessibly storing their data in open repositories. Surveys have shown that this process is perceived as a significant barrier to researchers in opening their data (Stodden, 2010; Marwick and Birch, 2018; Chawinga and Zinn, 2019). Section 4 of the paper extends the model to incorporate these costs as well, assuming that they distribute idiosyncratically across authors, and independently of the quality of results obtained. The main result is that high- and low-quality results will be published in both open and closed data regimes, but that open data will be favored by high-quality results. The structure of the equilibrium implies that the falsification among the low-quality results published in the open-data regime is *higher* than it would be in a single, high-cost (closed) regime. However, a final contribution of the model is a measure of "science welfare" that calculates the ex-post distortion of equilibrium beliefs about the quality of published results, and shows that open data will always improve the aggregate state of knowledge. The paper finishes with a discussion of these results in the context of the literature on open data in the social sciences.

SELECTIVE REPORTING AND VERIFICATION GIVEN VERIFICATION COSTS

Interaction Structure: The Prestige Game

The interaction is called a *prestige game*, indicating the interpretation of the utility functions that benefit is largely determined by the equilibrium beliefs about the quality of a piece of research produced. The game has two players: an author A and a representative reader B. The author does some research, reaching a result of stochastic quality q. For simplicity, suppose that there are two possible qualities H and L, represented as real numbers with $H > L$, and probability p of reaching result H. The quality is not directly observable, but is represented through the published paper; we denote by $\hat{q}$ the published description of the quality, and write it with lower case to distinguish it from the true quality of the result. That is, $\hat{q} \in \{h, l\}$, where h and l are taken to conventionally indicate H and L, respectively. In the standard manner of games of incomplete information, it will sometimes be convenient to refer to A players who observe $q = H$ as being of

type A_H, and those who observe $q = L$ as A_L. A's utility is therefore the prestige, or esteem that she experiences, or more concretely the expected value of q, upon announcing a quality of $\hat{q}$.

Denote $s_q \coloneqq \Pr\left[\hat{q} = h \mid q\right]$, the probability that a result of q is represented as if it were H. This means that s_L is probability of the kind of selective analysis or reporting alluded to above, that inflates results, or makes a low-quality result appear to be higher than it is. The model abstracts from the cost of engaging in this falsification (and in practice, inflation probably takes no more effort on A's part than would an "honest" analysis); the focus is on B's choice to look more deeply into the results. Specifically, at cost k, the reader B may verify A's work; denote the probability that B does so as v.[1]

Assume this verification process correctly identifies q to the public. If it turns out that $\hat{q} \neq q$, then B gets some esteem (benefit) and if $\hat{q} > q$, then A suffers a cost. The purpose of these assumptions is to reflect the social processes of prestige following the revision of scientific results. The basic assumptions are

- A downward revision ($\hat{q} > q$) generates a "large" cost C to A, and for simplicity awards the same benefit to B.
- An upwards revision ($\hat{q} < q$) has no inherent effect of A, other than the revision of the perceived quality itself, and awards a "small" benefit ε to B.
- If there is no revision of quality, there is no effect on A's esteem, although B must still pay the cost k.

To this point, the model has five parameters, which will be supposed to be in the order $H > L > C > \varepsilon > k$. The order of L and C is not important, but both benefits of verification must be greater than the cost k or non-verification is trivial. Although the model is simpler if $C \geq L$, which means that it is at least as good not to write a paper at all as to have a low-quality result be revealed as deceptive, the results below are based on the less restrictive, inverse case. If B does not verify the results of the research, the quality is taken to be its equilibrium average, conditional on $\hat{q}$.

This basic game can be represented as a special case of a signaling game, as in the game tree in **Figure 1**. Here Nature moves first at the central node, choosing the quality of the research H or L. Then A moves, choosing a quality to declare, h or l. Finally, B decides whether to verify the results (with probability v), or not [with probability $(1 - v)$]. The solution concept will be a sequential equilibrium of this game, allowing for mixed strategies. The appearance of expected values in the payoffs is the only departure from standard theory.

Equilibria

Signaling games are generally characterized by three sets: a set T of types representing the private information of a message sender, a set M from which signals may be drawn, and a set V of possible actions the message receiver may use in response. An equilibrium consists of strategies from T to M and from M to V, together with a set of beliefs over T given the realization of the message, such that each strategy is a best response to the other taking the beliefs into account, and the beliefs are consistent with the signaling strategies, following Bayes' Rule where possible. In this model, clearly $T = \{H, L\}$, $M = \{h, l\}$, $V = \{\text{verify, do not}\}$, and the beliefs are induced by

$$\Pr\left[q = H \mid \hat{q} = h\right] = \frac{ps_H}{ps_H + (1-p)\,s_L}$$

$$\Pr\left[q = H \mid \hat{q} = l\right] = \frac{p\,(1-s_H)}{p\,(1-s_H) + (1-p)\,(1-s_L)},$$

so long as these are defined.

It may be noted that this model does not satisfy the so-called *single-crossing property*, a simplification common in signaling games which generates a sorting of sender types, so "higher" types always send weakly "higher" messages in equilibrium. While it will always be the case that for A_H, being verified is at least as good as not being verified, while for A_L getting verified is (weakly) always worse, B prefers to verify A_H after a message of l and A_L after a message of h.[2] That is, while the single-crossing property holds with respect to B's actions, it does not hold with respect to the messages that induce those actions.

An important distinction among signaling models differentiates cheap talk, in which utility does not depend directly on the message sent, from costly signaling in which message senders can demonstrate something concerning their type directly through the signal sent. In the context modeled here, the message itself has no costs; however, the effect of B's choice on A does depend directly on the message sent. The general form of the utility function $U_A(q, \hat{q}, v)$ therefore is not generally constant in $\hat{q}$ (and in particular not when $v = 1$), so this can be considered a model of "impure" cheap talk. The single-crossing property is maintained in canonical models of cheap talk through a fixed and common-knowledge "bias" of the sender's preference with respect to the receiver's, meaning a divergence between the sender's (type-dependent) preferred action and the optimal action for the receiver to take, conditional on sender type. While the preferred action depends on the type of the sender, that is, the degree of "conflict" is constant; a key result is that the lower the degree of conflict, the more information may be transmitted in equilibrium. Another way of seeing the violation of the single-crossing property in the current model is that the bias depends on A's type. A_H has preferences not particularly at odds with those of B, while A_L has clear incentives to dissemble.

Finally, the fact that A's utility when $v = 0$ depends directly on beliefs is important mainly at the level of interpretations. In standard models, these beliefs are instrumental, and matter because they generate behavior that impacts utility. However, these reactions are an (optimal) mapping from the beliefs generated by the different strategies in equilibrium. In the current case, this mapping is direct: the utility to a researcher of having published a particular result is defined as the perception of its quality. This means that message choice by one type of A

[1]Costly verification puts this model somehow between cheap talk games in which the message is unverifiable, and signaling games as in Spence (1973).

[2]To compare to canonical education signaling, it would be as though an employer prefers to give a high wage to low-productivity employees and *vice versa*.

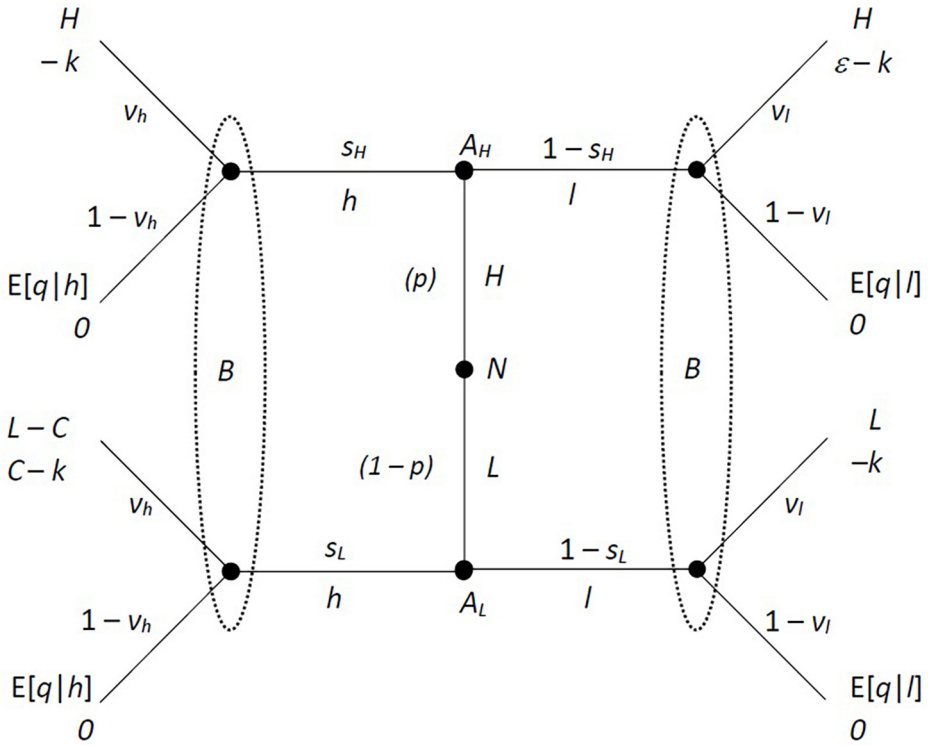

FIGURE 1 | Extensive form of the basic game structure.

imposes a kind of externality on the other if verification does not occur, but the effect can be thought of as a continuous "choice" by B of how to interpret each signal. Of course, there is no inherent incentive involved in this "choice of beliefs," other than the restriction imposed by Bayes' rule, but formally speaking, whether the choice is determined by optimal behavior or by application of Bayes' rule is irrelevant to A.

Equilibrium requires a mapping from q to the signal from A; from the signal to verification from B; and a set of beliefs over A types following each possible signal. The beliefs are defined as above. Each A-type optimal strategy consists of a simple decision rule, while B has a rule for each possible signal received. It is easy to see the following:

A_H chooses $s_H = 1$ if

$$v_h H + (1 - v_h)\, E\left[q|h\right] \geq v_l H + (1 - v_l)\, E\left[q|l\right], \qquad (1)$$

while for $s_L = 1$, A_L requires that

$$v_h\,(L - C) + (1 - v_h)\, E\left[q|h\right] \geq v_l L + (1 - v_l)\, E\left[q|l\right]. \qquad (2)$$

Concerning B, verification of a signal h requires

$$\frac{(1-p)\, s_L}{p s_H + (1-p)\, s_L} C \geq k \qquad (3)$$

while for the signal l the threshold is

$$\frac{p\,(1 - s_H)}{p\,(1-s_H) + (1-p)\,(1-s_L)} \varepsilon \geq k. \qquad (4)$$

Supplementary Appendix 1 investigates the equilibria of this game. These are described in Result 1, below

Result 1: equilibria of the prestige game

Consider the game described above, and suppose in addition that $p < \frac{C}{C+\varepsilon}$. *Then there are three equilibrium components, depending on k.*

A. *If* $k \geq (1-p)\, C$, *then there are no separating equilibria, but any mixed strategy profile* $s_H = s_L = s\varepsilon\ [0, 1]$ *can stand as an equilibrium, with no verification of any results by B.*
B. *If*$k\text{textless}\,(1-p)\, C$, *then there is a unique semi-separating equilibrium in which* $s_H = 1$; $s_L = \frac{p}{1-p}\frac{k}{C-k}$; $v_h = \frac{(H-L)\left(1-\frac{k}{C}\right)}{(H-L)\left(1-\frac{k}{C}\right)+C}$; *and* $v_l = 0$. *This equilibrium will be known as the* k-game.
C. *If* $k < p\varepsilon$, *then there is also a pooling equilibrium in which* $s_L = s_H = 0$, *and* $v_l = v_h = 1$.

Figure 2 illustrates the equilibrium falsification rate as k changes.

The additional restriction on p in Result 1 rules out a somewhat perverse set of equilibria in which the probability of a high-quality result is large enough that, in equilibrium, is it those declared as $\hat{q} = l$ that appear "suspicious," and are preferentially investigated. This does not seem to correspond to the real-world situation of scientific publishing, and to the extent that the C is "large" and ε is "small," the ratio in the condition will be close to unity, so the restriction is relatively mild.

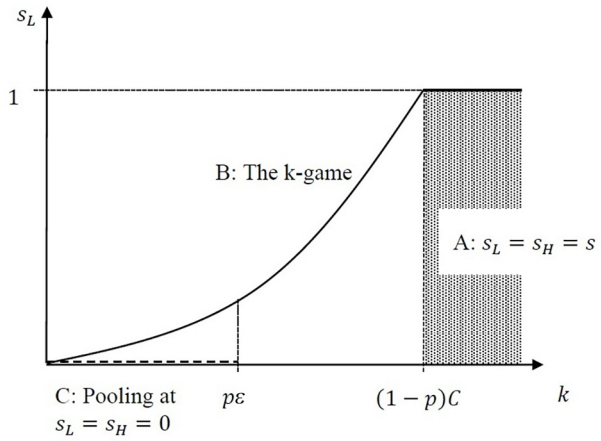

FIGURE 2 | Equilibria of the prestige game. Dashed line indicates pooling equilibria for very low values of *k*, and shaded area indicates that any pooling behavior can stand as an equilibrium if *k* is high enough.

More interesting are the effects of changes in k on the kinds of equilibria that exist. First, part A describes a situation in which the cost of verification is "too high." In particular, B does not verify any "babbling" equilibrium, where both types of A choose the same strategy, so the signal is uninformative. As a result, any such non-communicative strategy profile stands as an equilibrium. The esteem awarded to any signal is $pH + (1 - p)\ L$, so neither type of A has any reason to deviate, although if $s = 1$, then an Intuitive Criterion argument along the lines of Cho and Kreps (1987) could lead A_H to deviate, choosing $s_H = 0$ if B expected this.[3]

Part B of Result 1 describes the *k-game*, which will be the central focus of the analysis below. If $k < (1-p)\,C$, then B will verify a babbling equilibrium concentrated on the signal h, driving A_L – but not A_H – away from that strategy and generating separation. The separation can't be complete, though, or B would stop verifying, leading A_L to move back in. Therefore, the form of the equilibrium is in semi-separation, with A_H always sending the message h, while A_L mixes, sometimes sending the deceptive signal h and otherwise the honest one, l. Since in this equilibrium all results announced as l are actually of quality L, B verifies only the signal h. The average esteem to A_L is precisely L, with the costs of exposure (when verified) after sending message h exactly balancing out the benefits of deception (when not verified). While the closed-form solution is not intuitive, it is easy to see that the esteem to A_H, by contrast, which amounts to $v_h H + (1 - v_h)\,E[q \mid h]$, is strictly between H and L; as standard in games of incomplete information, deceptive behavior by the "low" types exerts an externality on the "high."

Finally, part C states if $k < p\varepsilon$, then B will verify a degenerate signal profile on either message.[4] Unlike the all-h profile, by contrast, the all-l one does stand as an equilibrium. A_H gets the payoff of H, and A_L gets the payoff of L. The former can do no better, and so long as the off-path beliefs also threaten verification should the latter deviate to signal h, A_L strictly prefers the equilibrium action. Notice that these off-path beliefs are not defined in equilibrium, but for instance, any sequence of "trembles" such that both types of A have equal probability of deviating in each element of the sequence will generate a sequential equilibrium of the same form. This equilibrium will not be the focus in what follows. It has a somewhat interesting interpretation as a "possible world" of scientific publishing; authors are "modest," never claiming high importance of their results, and readers systematically check all results, arriving at a complete state of knowledge about each. In this sense it seems like a "healthy" state of affairs, although it does impose costs on B. However, it is not particularly realistic as a description of the field, and moreover the equilibrium is not very robust to changes in the model. In particular, it relies on A_H being indifferent between claiming a high importance and having it revealed by B's verification. To the extent that there might also be some esteem from recognizing the importance of one's own work, or "embarrassment" from presenting results of $q = H$ with the label l, A_H should also suffer a cost when verified, which lead A_H to deviate from this equilibrium.

Comparative Statics: Result of Changes in *k*

The interpretation of open data in this model is to reduce the cost of verifying results; data sharing reduces k, and the model therefore gives some predictions about how open data might affect behavior. We see immediately that at least in the k-game, as k falls, v_h rises and s_L falls. Keeping in mind the possibility of a "corner solution" (part A of Result 1), this can be stated as follows

Corollary 1: so long as it reduces k to less than C (1 – p), open data will reduce the degree of inflation of weak results due to selective reporting.

Corollary 2: if open data reduces the level of inflation of weak results, it also increases the degree of verification.

An interesting implication of these corollaries concerns what might be termed *science welfare*. The goal of science is to have as accurate a picture of the functioning of the world as possible. To this extent, inflation of results, which distorts the impression that readers have of their significance, can be seen as reducing the overall quality of the scientific endeavor. While a scalar measure of "quality" does not directly map to distortion of the message contained in the results, it seems plausible that in presenting a "low quality" result of $q = L$ with the signal h, researchers will also change its overall message or real-world implications. Of course, it is common knowledge in the model that the unconditional probability of a high-quality result is p; rational expectations in equilibrium ensure that overall this remains the case. Moreover, the signal l perfectly identifies low-quality results in the k-game,

[3]The equilibrium utility is $E[q \mid h] = pH + (1-p)\,L$, and exactly one type ($A_H$) is in the situation where, if they sent the message l, and B believed that only they would send that signal, then it would be optimal for them to do so, as B would verify that signal and A_H would gain utility H. However, this deviation is not itself an equilibrium, as B would in this case also verify the signal h, leading A_L to deviate.

[4]The condition $p < \frac{C}{C+\varepsilon}$ serves formally to guarantee that $p\epsilon < (1-p)\,C$, so the latter threshold is lower than the former.

so from a scientific point of view there is no problem there. On the other hand, the public "state of knowledge" following a signal of h is not equal to H, which can be interpreted as a science welfare loss.

Following this interpretation, define a *science welfare loss function* as the *rate of unverified inflation of results*. A simple form for this is

$$W = -(1-v_h)\, s_L. \tag{5}$$

Expression (5) indicates that either full verification or fully honest reporting would be enough to reduce the science welfare loss to zero. Since v_h falls and s_L rises with k, the final corollary of this section can be stated

Corollary 3: reducing the costs of verification increases science welfare.

THE EFFECT OF DIFFERENT POSSIBLE LEVELS OF *k*

The results above show that imposing open data, if it sufficiently reduces the cost of verifying existing results, should unambiguously improve the quality of research publication. What happens when k is a choice by A? This is the scenario where the journal or discipline does not require open data, but allows authors to opt into it. This is modeled as allowing A to select from a set of possible values of k, called *regimes*. To model shared or unshared data, suppose that there are two regimes, $k_o < k_c$, indicating that *o*pen data has a lower verification cost than does *c*losed data. Suppose both regimes are parametrized such that the k-game is the equilibrium played. The choice of regime is made after Nature decides on q, simultaneously with $\hat{q}$.[5]

The structure of the interaction in any regime is identical to the prestige game described in section "Introduction," and as mentioned, the focus will largely be on k-game equilibria in each regime, or k-game components to the "double-game" equilibrium including regime choice. In this respect, the principal element that this new layer of regime choice adds to the strategic context is to endogenize the probability that a paper is of quality H in any regime; this value will now be determined by the equilibrium distribution of A-types that select into each regime. Define $\varphi(k)$ as the probability that a paper under regime k is of quality H. That is, if authors finding quality H choose k with probability $\lambda_k(H)$ and papers of quality L choose k with probability $\lambda_k(L)$, then

$$\varphi(k) = \frac{p\lambda_k(H)}{p\lambda_k(H) + (1-p)\,\lambda_k(L)}. \tag{6}$$

This value of $\varphi(k)$ will replace p in each regime k, with the rest of the prestige game analysis following as in section "Introduction." Supposing that the parameters are such that, in each regime, the k-game would be played in isolation, the following holds

Result 2: Equilibrium under regime choice

Consider a 2-regime prestige game where the k-game conditions hold in each regime, $k_o < k_c$, and define λ and φ as above. In any equilibrium

A. $\lambda_{k_o}(H) = 1; \frac{p}{(1-p)}\frac{k_o}{C-k_o} < \lambda_{k_o}(L) \leq 1$
B. *Behavior follows the k-game in the k_o regime; A_L always plays l in k_c*
C. *Off-path beliefs mimic the k-game in the k_c regime*

Result 2 establishes an "unraveling" effect under free regime choice. A_H strictly prefers the open to the closed regime, both because verification is more frequent there, and because the expected value of unverified messages is higher. Only the second of these is an advantage for A_L, of course, and so low-quality types follow high into the open regime only up to the point of indifference by B. As a result, the open regime contains a mix of types, and the k-game is played there, while only L-quality results are ever published in the closed regime. This then further implies that falsification is impossible in the closed regime, and so all those who enter it (honestly) declare l. However, the result that only the l-signal is made in the closed data equilibrium means that the equilibrium does not determine beliefs following a signal of h in the closed regime. Interestingly, several plausible off-path beliefs—for instance that only A_H (or only A_L) would choose this signal—turn out to upset the equilibrium. However, the equilibrium can be maintained with beliefs that re-create the k-game that would have been played in that regime, if A_H did not deviate toward the open data. These off-path beliefs also admit an interesting interpretation in terms of "non-stigmatization." In essence, while the equilibrium never has high-quality results in the k_c regime, it relies on the belief that this "could happen" off the equilibrium path, and so signals of type h in k_c are not systematically verified. On the other hand, B would also ascribe such a "tremble" to A_L types with some probability, so some verification occurs. In other words, the unobserved, counterfactual declarations of h in the high-cost regime are not given particularly bad (or good) interpretations in the model.

Another interesting feature of this model concerns the indeterminacy noted in A_L's strategy in part A. Overall, the rate of falsification in the k_o regime must be such that B is indifferent between verifying or not, but any combination of entry to that regime and falsification once there that generates this overall rate can stand. Consider generally the falsification rate in each regime (subscripted by regime rather than by q since only $q = L$ results in falsification in the k-game equilibrium):

$$s_k^* = \frac{\varphi(k)}{1-\varphi(k)}\frac{k}{C-k}$$

This can be combined with (6) to give

$$s_k^* = \frac{p}{(1-p)}\frac{k}{C-k}\frac{\lambda_k(H)}{\lambda_k(L)}. \tag{7}$$

Expression (7) implies that the more H-quality papers select into a regime relative to L in equilibrium, the more the L-quality papers in that regime will falsify their results. It is tempting

[5] As pointed out by a reviewer, since nothing "happens" in the model between A's choices, they are strategically simultaneous even if separated temporally in practice. If, for instance, —due perhaps to preregistration—authors had to commit to a regime before observing q, one can conjecture that the results might be more strongly affected.

to interpret this as "trying to fit in with a better pool"; the equilibrium falsification rate must leave B indifferent between verifying or not. L-quality papers have to falsify more in equilibrium as the relative frequency of H increases, in order to balance out increased risk to B of paying the cost k without getting any benefit. Because in the specific equilibrium of Result 2, $\lambda_{k_o}(H) = 1$, moreover, this means that A_L's equilibrium strategy can be determined up to

$$s^*_{k_o}\lambda_{k_o}(L) = \frac{p}{1-p}\frac{k_o}{C-k_o}, \tag{8}$$

and any combination of the terms on the left that satisfy (8) are equivalent for the equilibrium. The entry and falsification rates are jointly determined, in other words, but the overall level of falsification—and therefore verification—in the open-data regime is the same, whether it represents a large fraction of the A_L types falsifying to a moderate extent, or a smaller fraction falsifying more consistently.

This is important because it implies that the expected value of an unverified signal h does not change with λ_k. Recall from the formula in Result 1 (B) that the verification rate does not depend on p, which intuitively is because this rate serves in equilibrium to leave B indifferent between signals conditional on having observed $q = L$. The same holds in the two-regime setting. So long as neither $\lambda_k(L)$ nor $\lambda_k(H)$ are equal to zero,

$$\mathrm{E}\left[q|h,k\right] = L\cdot\Pr\left[L|h,k\right] + H\cdot\Pr\left[H|h,k\right]$$

$$= L\frac{s_k(1-p)\lambda_k(L)}{s_k(1-p)\lambda_k(L)+p\lambda_k(H)} + \mathrm{H}\frac{p\lambda_k(H)}{s_k(1-p)\lambda_k(L)+p\lambda_k(H)}$$

$$= L\frac{\frac{p}{(1-p)}\frac{k}{C-k}\frac{\lambda_k(H)}{\lambda_k(L)}(1-p)\lambda_k(L)}{\frac{p}{(1-p)}\frac{k}{C-k}\frac{\lambda_k(H)}{\lambda_k(L)}(1-p)\lambda_k(L)+p\lambda_k(H)}$$

$$+\mathrm{H}\frac{p\lambda_k(H)}{\frac{p}{(1-p)}\frac{k}{C-k}\frac{\lambda_k(H)}{\lambda_k(L)}(1-p)\lambda_k(L)+p\lambda_k(H)}$$

$$= H - \frac{k}{C}(H-L).$$

Stated plainly, adding a low-k regime to a costlier one may result in different A types choosing different regimes, and if it does, then the equilibrium effects of this will be balanced by changes in the falsification rates in each regime. But the verification rate in any regime (provided it maintains its k-game structure in equilibrium) will not change with the addition of another regime.[6]

Remark 1:

The equilibrium results of selection of A-types into different regimes include adjustment of falsification rates, with higher rates in the regime containing more A_H types; it does not affect the verification rates in either regime, compared to the single-regime k-game.

Combined with the unraveling result, this implies that, while there is a continuum of equilibria, with some fraction of A_L between zero and $\frac{p}{(1-p)}\frac{k_o}{C-k_o}$ choosing the high-cost regime (closed data), the low cost regime absorbs the falsification, and the science welfare is not affected by which equilibrium occurs. This follows directly from Remark 1. Science welfare was defined as the overall rate of unverified falsification, and neither of those quantities (verification rates or overall falsification) are affected in this model by the addition of a high-cost regime that attracts only A_L.

Result 3: Science welfare in the two-regime, free-choice model is determined by the costs of the lower-cost regime.

COSTS OF PREPARING OPEN DATA

The model from section "The Effect of Different Possible Levels of k" has some interesting characteristics, but is ultimately not quite satisfactory. It induces a correlation between regime choice and quality, suggesting that results published in open data should be, on average, of higher quality than others. But it at once predicts a multiplicity of equilibria with respect to A_L's regime choice, and also quite starkly that in any of them, all results published in the high-cost regime should be declared as low-quality, and the "unraveling" in terms of science welfare is complete. In addition, while the "no-stigmatization" result is anecdotally interesting, the off-path beliefs are at once arbitrary, imposed for no other reason than supporting the equilibrium, and rather precise, requiring a specific relationship between two different kinds of deviation. An extension that ensures that A_H may sometimes opt for the k_c regime even in the presence of multiple k-games "solves" many of these issues, yielding sharper predictions with more intuitive interpretation, at the cost of an additional assumption and parameter.

In surveys, one of the principal reasons that researchers cite for not participating in open data is the time and effort costs of doing so (Stodden, 2010; Marwick and Birch, 2018; Chawinga and Zinn, 2019). In the model so far, on the other hand, the choice of regime has been costless. Suppose, therefore, that there are still two possible levels of k, $k_o < k_c$, and that each determines a separate k-game into which authors select. In addition, there is a utility penalty K to player A for choosing k_o due to the time and effort costs of opening the data. The goal of this assumption is to make it so that some, but not all, of the A_H players choose the k_c regime, so it requires idiosyncratic costs to generate the differences. For simplicity more (perhaps) than realism, suppose that K distributes across A players randomly according to a continuous distribution $G(K)$ that is independent of q. For notational convenience, also normalize $H - L = 1$.

These assumptions induce a change in the equilibrium structure. Intuitively, A players of both types with high enough costs choose the closed regime, while those with low costs choose the open. This is driven by higher verification rates in the open regime, which make it preferable to A_H, and therefore increase

[6]Naturally, high-cost regimes will still have higher falsification and lower verification rates than low-cost ones do.

the prestige (expected value) of unverified publications there. However, if there is a cost to entering the open regime, and the benefit is conditional on either being an A_H type or not being verified, then there is no reason why A_L would ever choose that regime and then announce l. In short, for A_L, the strategy (k_c, l) dominates the strategy (k_o, l), and rather than announcing l in the k-game of the open regime, A_L goes to the closed one. This implies that all publications in the open regime are announced as h. On the other hand, while this change appears to affect behavior in important ways, the informational content of the equilibrium can be preserved, as k-game structure of the open data regime is maintained by the rate of entry to the regime, rather than the rate of falsified signaling within it. Dominance of the closed-data regime for results announced as l eliminates one of A_L's strategic margins to allow probabilistic verification by B, and hence the multiplicity of equilibria found above, but the other strategic margin remains available, preserving the basic game intuition. This is summarized in Result 4.

Result 4: Consider a two-regime environment with $k_o < k_c$ and idiosyncratic costs of entry to the k_o regime. Then

A. *There is a unique set of equilibrium entry rates to the open regime, which satisfies $\lambda_{k_o}(H) > \lambda_{k_o}(L)$*
B. *$s_{k_o} = 1$ for both A_H and A_L*
C. *The k-game is played in the closed regime among the residual, high-K A-types*

Part (B) of Result 4 follows from the dominance argument above. Part (C) follows from the presence of both types of A-player in the closed regime. To see part (A), note that if B does not verify, then A_L will enter if costs are low enough, while if B always verifies, then A_L will never enter. Therefore B must be indifferent to justify probabilistic verification. Building from expression (7), this implies that it must be that in equilibrium

$$1 = \frac{p}{(1-p)}\frac{k_o}{C-k_o}\frac{\lambda_{k_o}(H)}{\lambda_{k_o}(L)}. \tag{9}$$

Expression (9) shows that in equilibrium, more high-type authors choose the open regime than closed, justifying the inequality in part (A). Furthermore, it shows that entry in equilibrium must be in a fixed ratio. The unique level at which this ratio can stand in equilibrium is determined by threshold values (K_H^*, K_L^*) such that (9) holds when $(\lambda_{k_o}(H), \lambda_{k_o}(H)) = (G(K_H^*), G(K_L^*))$, and also

$$v_o H + (1-v_o)\, E\left[q \mid h, k_o\right] = EU\left[k_c \mid H\right] + K_H^* \tag{10}$$

$$v_o (L-C) + (1-v_o)\, E\left[q \mid h, k_o\right] = EU\left[k_c \mid L\right] + K_L^* \tag{11}$$

Expressions (10) and (11) indicate that for each type $T = H$, L of A, there is a threshold cost K_T^* such that those with cost greater than K_T^* choose the closed regime, while those with lower costs choose the open. The extra cost of data preparation must be exactly balanced by a higher expected payoff in the open regime for both types at this threshold.

It is clear from inspection of (10) and (11) that any level of v_o will determine a pair (K_H^*, K_L^*). Moreover, since, as v_o rises from zero to one, the left-hand side of (10) rises, while that of (11) falls, the difference or ratio between the implied levels of K_H^* and K_L^* is monotonic in v_o. Thus, there can be only one level of v_o that also satisfies the specific ratio determined in (9). To see that there is at least one, notice that first that (9″) implies that

$$G(K_L^*) = \frac{p}{(1-p)}\frac{k}{C-k}G(K_H^*) < G(K_H^*) \longrightarrow K_L^* < K_H^* \tag{9''}$$

Next, Remark 1 implies that in the k-game in the closed regime, $EU\left[k_c \mid L\right] = L$, while

$$EU\left[k_c \mid H\right] = \frac{(H-L)\left(1-\frac{k_c}{C}\right)}{(H-L)\left(1-\frac{k_c}{C}\right)+C}H + \frac{C}{(H-L)\left(1-\frac{k_c}{C}\right)+C}\left[H-\frac{k_c}{C}(H-L)\right]$$

$$EU\left[k_c \mid H\right] = \frac{\left(1-\frac{k_c}{C}\right)}{\left(1-\frac{k_c}{C}\right)+C}H + \frac{C}{\left(1-\frac{k_c}{C}\right)+C}\left[H-\frac{k_c}{C}\right]$$

$$EU\left[k_c \mid H\right] = \frac{\left(1-\frac{k_c}{C}\right)}{\left(1-\frac{k_c}{C}\right)+C}H + \frac{C}{\left(1-\frac{k_c}{C}\right)+C}H - \frac{C}{\left(1-\frac{k_c}{C}\right)+C}\frac{k_c}{C}$$

$$EU\left[k_c \mid H\right] = H - \frac{k_c}{\left(1-\frac{k_c}{C}\right)+C} > L$$

Inserting these values into (10) and (11) and investigating the boundary conditions, we see that when $v_o = 0$,

$$E\left[q \mid h, k_o\right] = H - \frac{k_c}{\left(1-\frac{k_c}{C}\right)+C} + K_H^* \tag{10''}$$

$$E\left[q \mid h, k_o\right] = L + K_L^*. \tag{11''}$$

Combining these implies that

$$K_H^* - K_L^* = L - \left[H - \frac{k_c}{\left(1-\frac{k_c}{C}\right)+C}\right] < 0. \tag{12}$$

The inequality in expression (12) means that there are "too many" A_L types entering the open regime when verification is "low enough." Specifically, the threshold cost for A_L is higher than that for A_H, which means that the ratio would be greater than unity, and cannot be accommodated in (9). On the other hand, the implicit threshold of K_L^* hits zero when verification is equal to its (single-regime) equilibrium level in the open regime, as then the expected value to A_L of both regimes equals L. This is clearly "too few" A_L-types entering. Because (10) and (11) are both continuous in v_o, there must therefore be a single level of entry that satisfies all conditions, establishing the result.

Regarding science welfare in this configuration, as is intuitive, the costs to using open data, or more exactly the resultant distortions they induce, increase equilibrium falsification relative to the model in section "The Effect of Different Possible Levels of k." But it is interesting to note that the distortion comes from two different sources. First, the presence of A_H types in the closed-data regime allows the A_L types who choose that regime to falsify with some probability, which was impossible above and contributes to a larger overall rate. Also, however, a corollary to the argument above concerning the entry rate into the open regime is that the verification rate there must be lower than it would be in a single, open regime. Specifically, expression (11) implies that the threshold preparation cost K_L^* drives a wedge between the expected utilities of the two regimes for A_L. Since in the single-regimes, expected utility was equal to precisely L in both regimes, and the k-game in the closed regime implies that this remains the case in the current model, it follows that expected utility must be higher for A_L in the open regime. This then requires that the verification level be lower than its single-regime level.

Furthermore, it is immediate that a reduction in the preparation cost distribution—for instance in the sense of stochastic dominance—would reduce the levels of $\left(K_L^*, K_H^*\right)$ that satisfy (9), (10), and (11), and therefore reduce this distortion, increasing science welfare. Indeed, the model in section "The Effect of Different Possible Levels of k" can be taken as a limiting case of that in section "Costs of Preparing Open Data," when costs are reduced to zero. The result is therefore as follows:

Result 5: Science welfare with preparation costs is reduced both by the entry of high-quality work into the closed data regime, and also by distortions of the verification rate in the open data regime.

Corollary 3: A leftward shift in the distribution of preparation costs will reduce these distortions and increase science welfare.

DISCUSSION

The model in this paper investigated ways in which open data can leverage social esteem to discipline the reporting of scientific results. The key assumptions were (1) authors get a direct utility benefit from the public (equilibrium) perception of the quality of work they do; (2) readers get some utility benefit from discovering that the presented quality of a given result is inaccurate; (3) discovery of inflated inaccuracy, in which low-quality results are presented as high, imposes a utility cost on authors; (4) readers must incur a cost in order to check the accuracy of the presented results. These assumptions were selected to reflect potentially important elements of the publishing process, and set up a model in which open data—one of whose primary goals is to reduce the cost to readers of replicating or recreating published results—could have an influence on the tendency to misrepresent.

The model can be seen as an application of signaling games to the case of scientific publications. While this is not a specific subject that has received much theoretical treatment, signaling games generally represent of course a vast and rich field, from which much more is taken for this paper than is contributed. The structure of simple signaling games is very standard, and has been well-understood since Spence (1973); the application here used standard refinements such as sequential equilibrium (Kreps and Wilson, 1982a,b) and, to a limited extent the Intuitive Criterion (Cho and Kreps, 1987). The idea that the signal is designed to represent some otherwise unobservable quality that matters to the signal receiver also indicates links, for instance, to literature on advertising (see Bagwell, 2007). A modest theoretical innovation, designed to reflect the esteem-based nature of the benefit to the author of discovering important results, sees utility in the model as based directly on beliefs about the signal sender's type, rather than—as is perhaps more common in economic interactions—based on the receiver's reaction to those beliefs. But as mentioned, this is essentially a difference in interpretation and has little influence on the formal structure of the game.

Another departure from the standard signaling game that might be found in any advanced microeconomics course is the fact that in this model, there are effectively two sequential signals. Section "Selective Reporting and Verification Given Verification Costs" of the paper described a semi-separating k-game equilibrium in which authors of low-quality work partially imitated high quality, and showed that the lower the cost of verification, the less falsification there will be. A measure of *science welfare loss*, defined as the equilibrium level of unverified falsification of results, was found to be decreasing in the cost of verification. Section "The Effect of Different Possible Levels of k" then extended this to a case in which there were two possible levels of this cost or regimens—reflecting open and closed data—in which case the choice of one regime or the other could be seen as a second level of signaling. It found that while some low-quality work might use the high-cost signal, there was partial "unraveling" in that some low-quality work would also be presented with a low verification cost. This is basically a second level of semi-separation in regime choice. Interestingly, while behaviorally the model in section "The Effect of Different Possible Levels of k" did not pin down what the equilibrium distribution of low-quality work signals would be, the science welfare was the same regardless of whether the high-cost regime existed or not. In terms of the equilibrium level of distortion, open data completely crowded out closed.

In the model from section "The Effect of Different Possible Levels of k," both the quality signal and the regime choice were essentially *cheap talk*, imposing no costs on the authors who chose them. In line with survey data and introspective evidence, section "Costs of Preparing Open Data" then extended the model to make using the open data regime costly relative to the closed. This resulted in some high-quality work being submitted in each regime, and increased the science welfare loss proportionally.

What does this model tell us about open data as a tool for strengthening the scientific publishing process? First, to the extent that readers get some benefit from correcting mistakes they find in the literature, facilitating this with open data should act as a disciplining tool for the presentation of results. Open data, in other words, should "work." Furthermore, while the interpretation of player B in the model is as a representative reader who may spend effort to check results of published work, it is worth mentioning that any other effect that reduced the

cost of close inspection of results should have similar effects. For instance, incentives for careful reviewing at the peer review stage, or institutional procedures on the part of employers or scientific journals could be implemented to reduce the opportunity cost of verification.[7] Second, however, the model shows that this relies on the costs of preparing open data not being too high. In particular, the more high-quality work that is submitted in closed data, the greater the science welfare loss in equilibrium. Conversely, if the preparation costs are pushed down to zero, there is no need to impose open data on the scientific community; high-quality work will select into the low-verification-cost regime, and the residual work that goes into the high-cost regime will not affect the overall level of distortion in the literature, although interestingly, the few low-quality results that are published in open data will be more likely to be falsified when they are in a "stronger pool."

The theoretical results from Sections "The Effect of Different Possible Levels of k" and "Costs of Preparing Open Data" both predict an overall correlation between the adoption of open data and research quality. This fits well with the existent empirical literature showing that papers published under open data have higher citation counts than those without (Pienta et al., 2010; Marwick and Birch, 2018). The results in these papers are correlational, and it is conceivable that the open data itself increased citation count through encouraging others to build on the published results—indeed that is the preferred interpretation in the literature. To this extent, the model is useful in supplying a justification for a separate causal interpretation of the data (see Soeharjono and Roche, 2021; for an early formal treatment see Verrecchia, 1990).

One of the more interesting implications these results may have concerns educational policies. Preparing data for open publication requires a specific set of skills, and explicitly training young academics in these skills seems bound to reduce their cost to doing so later. From the perspective of the model in section "Costs of Preparing Open Data," this would result in the kind of "leftward shift" in the function $G(K)$ that would reduce the equilibrium distortion rate. Similarly, part of the training in empirical work could be specifically in replicating existing studies using open data, or performing meta-analyses. Such measures would have the effect in the model of reducing k in any regime, which would increase verification rates and reduce falsification in all of them, again improving science welfare. Measures such as these might be better even than imposing open data on publication in the field. Even well-prepared data after all can only be verified by willing B-players. Also, the costs to verification and data preparation should be taken into account in a wider welfare criterion. Although equilibrium verification implies that agents are at least as well off incurring those costs as not, their final utility will obviously be improved if the costs are lower. From an even broader, "libertarian paternalist" perspective it may also be preferable to develop a system in which agents choose the "right" actions for themselves than one in which they are forced to do so. Such an argument has philosophical merit, and also utilitarian appeal, as those who are forced to engage in any action will be the most likely to try to find loopholes to avoid it.

[7] The author thanks a reviewer for highlighting this consideration.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

ES: construction and resolution of the theoretical model.

FUNDING

The author gratefully acknowledges financing from the ANR through the ISITE-BFC International Coach program (ANR-15-IDEX-003, PI Uri Gneezy).

ACKNOWLEDGMENTS

The author thank Antoine Malézieux for discussion and inspiration on the topic and attendant literature, and Claude Fluet, Theo Offerman, and Jeroen van de Ven for useful comments on a draft of the model.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2021.761168/full#supplementary-material

REFERENCES

Alsheikh-Ali, A. A., Qureshi, W., Al-Mallah, M. H., and Ioannidis, J. P. (2011). Public availability of published research data in high-impact journals. *PLoS One* 6:e24357. doi: 10.1371/journal.pone.0024357

Bagwell, K. (2007). The economic analysis of advertising. *Hand. Indust. Organ.* 3, 1701–1844.

Baker, M. (2016). 1,500 scientists lift the lid on reproducibility. *Nature* 533, 452–455. doi: 10.1038/533452a

Camerer, C. F., Dreber, A., Forsell, E., Ho, T. H., Huber, J., Johannesson, M., et al. (2016). Evaluating replicability of laboratory experiments in economics. *Science* 351, 1433–1436. doi: 10.1126/science.aaf0918

Camerer, C. F., Dreber, A., Holzmeister, F., Ho, T. H., Huber, J., Johannesson, M., et al. (2018). Evaluating the replicability of social science experiments in nature and science between 2010 and 2015. *Nat. Hum. Behav.* 2, 637–644. doi: 10.1038/s41562-018-0399-z

Chawinga, W. D., and Zinn, S. (2019). Global perspectives of research data sharing: a systematic literature review. *Libr. Inf. Sci. Res.* 41, 109–122. doi: 10.1016/j.lisr.2019.04.004

Cho, I., and Kreps, D. (1987). Signaling games and stable equilibria. *Q. J. Econ.* 102, 179–222. doi: 10.2307/1885060

Filippin, A., and Crosetto, P. (2016). A reconsideration of gender differences in risk attitudes. *Manage. Sci.* 62, 3138–3160. doi: 10.1287/mnsc.2015.2294

Ioannidis, J. P. (2005). Why most published research findings are false. *PLoS Med.* 2:e124. doi: 10.1371/journal.pmed.0020124

Kreps, D. M., and Wilson, R. (1982a). Reputation and imperfect information. *J. Econ. Theory* 27, 253–279. doi: 10.1016/0022-0531(82)90030-8

Kreps, D. M., and Wilson, R. (1982b). Sequential equilibria. *Econometrica* 50, 863–894.

Marwick, B., and Birch, S. E. P. (2018). A standard for the scholarly citation of archaeological data as an incentive to data sharing. *Adv. Archaeol. Pract.* 6, 125–143. doi: 10.1017/aap.2018.3

Obels, P., Lakens, D., Coles, N. A., Gottfried, J., and Green, S. A. (2020). Analysis of open data and computational reproducibility in registered reports in psychology. *Adv. Methods Pract. Psychol. Sci.* 3, 229–237. doi: 10.1177/2515245920918872

Open Science Collaboration (2015). Estimating the reproducibility of psychological science. *Science* 349:aac4716. doi: 10.1126/science.aac4716

Pienta, A. M., Alter, G. C., and Lyle, J. A. (2010). "The enduring value of social science research: the use and reuse of primary research data," in *Paper Presented at "The Organisation, Economics and Policy of Scientific Research" Workshop*, (Torino).

Soeharjono, S., and Roche, D. G. (2021). Reported individual costs and benefits of sharing open data among Canadian academic faculty in ecology and evolution. *BioScience* 71, 750–756. doi: 10.1093/biosci/biab024

Spence, A. M. (1973). Job market signaling. *Q. J. Econ.* 87, 355–374. doi: 10.2307/1882010

Stevens, J. R. (2017). Replicability and reproducibility in comparative psychology. *Front. Psychol.* 8:862. doi: 10.3389/fpsyg.2017.00862

Stodden, V. (2010). *The Scientific Method in Practice: Reproducibility in the Computational Sciences*. MIT Sloan Research Paper No. 4773-10. Available online at: http://dx.doi.org/10.2139/ssrn.1550193

Tenopir, C., Allard, S., Douglass, K., Aydinoglu, A. U., Wu, L., Read, E., et al. (2011). Data sharing by scientists: practices and perceptions. *PLoS One* 6:e21101. doi: 10.1371/journal.pone.0021101

Verrecchia, R. E. (1990). Information quality and discretionary disclosure. *J. Account. Econ.* 12, 365–380. doi: 10.1016/0165-4101(90)90021-U

Womack, R. P. (2015). Research data in core journals in biology, chemistry, mathematics, and physics. *PLoS One* 10:e0143460. doi: 10.1371/journal.pone.0143460

Conflict of Interest: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychology

ORIGINAL RESEARCH
published: 16 November 2021
doi: 10.3389/fpsyg.2021.760258

Reducing the Cost of Being the Boss: Authentic Leadership Suppresses the Effect of Role Stereotype Conflict on Antisocial Behaviors in Leaders and Entrepreneurs

Lucas Monzani[1], Guillermo Mateu[2,3], Alina S. Hernandez Bark[4] and José Martínez Villavicencio[5]*

[1] Ivey Business School, University of Western Ontario, London, ON, Canada, [2] Department of Finance, Law and Control, Burgundy School of Business, University Bourgogne Franche-Comté, CEREN, EA 7477, Dijon, France, [3] Department of Accounting, University of Valencia, Valencia, Spain, [4] Department of Social Psychology, Goethe University Frankfurt, Frankfurt, Germany, [5] Instituto Tecnológico de Costa Rica, Cartago, Costa Rica

OPEN ACCESS

Edited by:
Tarek Jaber-Lopez,
Université Paris Nanterre, France

Reviewed by:
David Pascual-Ezama,
Complutense University of Madrid, Spain
Chetan Sinha,
O. P. Jindal Global University, India
Lara Ezquerra,
University of the Balearic Islands, Spain

***Correspondence:**
Guillermo Mateu
guillermo.mateu@bsb-education.com

Specialty section:
This article was submitted to Personality and Social Psychology, a section of the journal Frontiers in Psychology

Received: 17 August 2021
Accepted: 11 October 2021
Published: 16 November 2021

Citation:
Monzani L, Mateu G, Hernandez Bark AS and Martínez Villavicencio J (2021) Reducing the Cost of Being the Boss: Authentic Leadership Suppresses the Effect of Role Stereotype Conflict on Antisocial Behaviors in Leaders and Entrepreneurs. Front. Psychol. 12:760258. doi: 10.3389/fpsyg.2021.760258

What drives entrepreneurs to engage in antisocial economic behaviors? Without dismissing entrepreneurs' agency in their decision-making processes, our study aims to answer this question by proposing that antisocial economic behaviors are a dysfunctional coping mechanism to reduce the psychological tension that entrepreneurs face in their day-to-day activities. Further, given the overlap between the male gender role stereotype and both leader and entrepreneur role stereotypes, this psychological tension should be stronger in female entrepreneurs (or any person who identifies with the female gender role). We argue that besides the well-established female gender role – leader role incongruence, female entrepreneurs also suffer a female gender role – entrepreneur role incongruence. Thus, we predicted that men (or those identifying with the male gender role) or entrepreneurs (regardless of their gender identity) that embrace these roles stereotypes to an extreme, are more likely to engage in antisocial economic behaviors. In this context, the term antisocial economic behaviors refers to cheating or trying to harm competitors' businesses. Finally, we predicted that embracing an authentic leadership style might mitigate this effect. We tested our predictions in two laboratory studies (Phase 1 and 2). For Phase 1 we recruited a sample of French Business school students (N = 82). For Phase 2 we recruited a sample of Costa Rican male and female entrepreneurs, using male and female managers as reference groups (N = 64). Our results show that authentic leadership reduced the likelihood of entrepreneurs and men of engaging in antisocial economic behaviors such as trying to harm one's competition or seeking an unfair advantage.

Keywords: entrepreneur role stereotype, female entrepreneurship, gender-entrepreneur role incongruence, leader-entrepreneur role incongruence, antisocial behaviors, economic games

INTRODUCTION

For a time, Elizabeth Holmes was a true inspiration for female entrepreneurs. Young, charismatic, and successful in Silicon Valley, the "girl boss" reigned triumphant over a sector infamous for its hyper-masculine "bro culture" (Cook, 2020). Yet, as CEO of Theranos, Mrs. Holmes faked the results of clinical trials and reported doctored information to her shareholders. The actions of

Holmes and other unethical female leaders created a headache for those scholars that related the female anatomical sex to a higher frequency of ethical behaviors at work (Borkowski and Ugras, 1998; Whitley et al., 1999; Childs, 2012).

As Mrs. Holmes and many other young entrepreneurs in the health sector found out the hard way (e.g., Mr. Martin Shkreli – "the Pharma Bro"), unethical business practices do not pay in the long run. While antisocial economic behaviors might bring results in the short term, engaging in antisocial economic behaviors leads to adverse long-term outcomes for leaders and entrepreneurs, their employees, their ventures capitalists, and other stakeholders. Without dismissing a person's agency as a driver of unethical behavior in leaders and entrepreneurs, we asked ourselves if Mrs. Holmes' unethical behavior was just a matter of individual differences (e.g., anatomical sex)? Or could these antisocial economic behaviors be a dysfunctional way of copying with the "cost of being the boss"?

There are three reasons why answering our research questions matters. First, such understanding would explain recurring issues in the entrepreneurship literature (Hughes et al., 2012; Jennings and Brush, 2013), such as why men are more likely to become entrepreneurs than women. Second, it would explain why some entrepreneurs decide to engage in unethical business practices, such as the antisocial economic behaviors that Mrs. Holmes and Mr. Shkreli displayed while leading their ventures. Third, scholars might use our findings to design interventions that deter entrepreneurs from engaging in unethical business practices and prevent future harm to shareholders and other stakeholders.

Role Congruency Theory (RCT; Eagly and Karau, 2002) is a valuable theoretical anchor for our research efforts. RCT explains well why women and other minorities suffer a double bind and prejudice when seeking or occupying leadership roles. Unfortunately, RCT does not explain the nuances of how this mechanism would work outside the traditional context of corporate firms. Whereas RCT would explain why women might suffer from reduced access to venture capital, it does not explain why female leaders might engage in the hyper-masculine antisocial economic behaviors that Mrs. Holmes displayed as the founder of her firm. Thus, by extending RCT to the female entrepreneurship arena, we provide a valuable theoretical contribution that informs the practice of leadership and entrepreneurship.

The main objective of this study is to determine if female entrepreneurs make antisocial decisions as a dysfunctional way of copying with the psychological tension created by simultaneously occupying incongruent social roles. To this end, we conducted two laboratory studies in two western countries. In a sample of business school students, Phase 1 tests our predictions about the effects of role conflict among three future roles on antisocial decisions employing two behavioral games. Phase 2 tests main and interactive effects of the same roles on antisocial decisions in a sample of Costa Rican entrepreneurs and managers.

LITERATURE REVIEW

A myriad of studies supported the propositions of RCT (Eagly and Karau, 2002). RCT proposes that women suffer a prejudice that prevents them from (a) reaching leadership roles in corporations, and by which (b) women are evaluated more harshly than men in a leadership position (Koenig et al., 2011). RCT invokes cognitive dissonance as the psychological mechanism driving said prejudice toward female leaders. RCT claims that when the characteristics of a person occupying a role misalign with the stereotypical expectations toward a given role, "this inconsistency lowers the evaluation of the group member as an actual or potential occupant of the role" (Eagly and Karau, 2002, p. 574).

Following the logic behind RCT, we argue that female entrepreneurs suffer a similar (or even stronger) prejudice than female managers, given that "entrepreneur" is also a social role ("think entrepreneur – think male;" Laguía et al., 2018). We expect entrepreneurs to suffer the effects of an additional cognitive dissonance (regardless of their anatomical sex or gender identity), which arises from the conflicting stereotypical expectations toward the leader and entrepreneur role. This logic also suggests that female entrepreneurs will suffer conflicting expectations toward three instead of two social roles.

This "triple bind and prejudice" should then result in a stronger psychological tension than the one suffered by their male counterparts. Whereas there are always functional ways of reducing psychological tension, antisocial economic behaviors seem to result from dysfunctional copying mechanisms (self-stereotyping; in-extremis identity trade-off). We unpack this last claim in the following section and summarize our predictions in **Figure 1**.

Embracing Role Stereotypes In-Extremis and Antisocial Economic Behaviors

Role stereotypes describe the "ideal" representations of social roles in a social group or culture. In turn, these ideal representations are incorporated into a person's sense of self through a psycho-social process called socialization (Hartley, 1959). Once a role stereotype is internalized into the self, it drifts from the stream of consciousness and starts eliciting automatic behavioral as responses to external stimuli. Because role stereotypes only describe "ideal" representations, said representations are susceptible to change across time (history) and space (cultures), and might always reflect the reality behind the stereotype. The present study focuses on three Western and contemporary role stereotypes and predicts what occurs if these stereotypes are embraced in-extremis.

For a virtue ethics view, the "in-extremis" adjective refers to virtues becoming vices due to an excessive display of said virtue. For example, an excessive display of the three character strengths that compose the virtue of Courage (Bravery, Persistence, and Integrity) might lead to recklessness, zealously, and self-righteousness (Crossan et al., 2013, 2017). Virtue would reside using one's practical wisdom to avoid any "in-extremis" behavior. Embracing in-extremis a social role is consistent

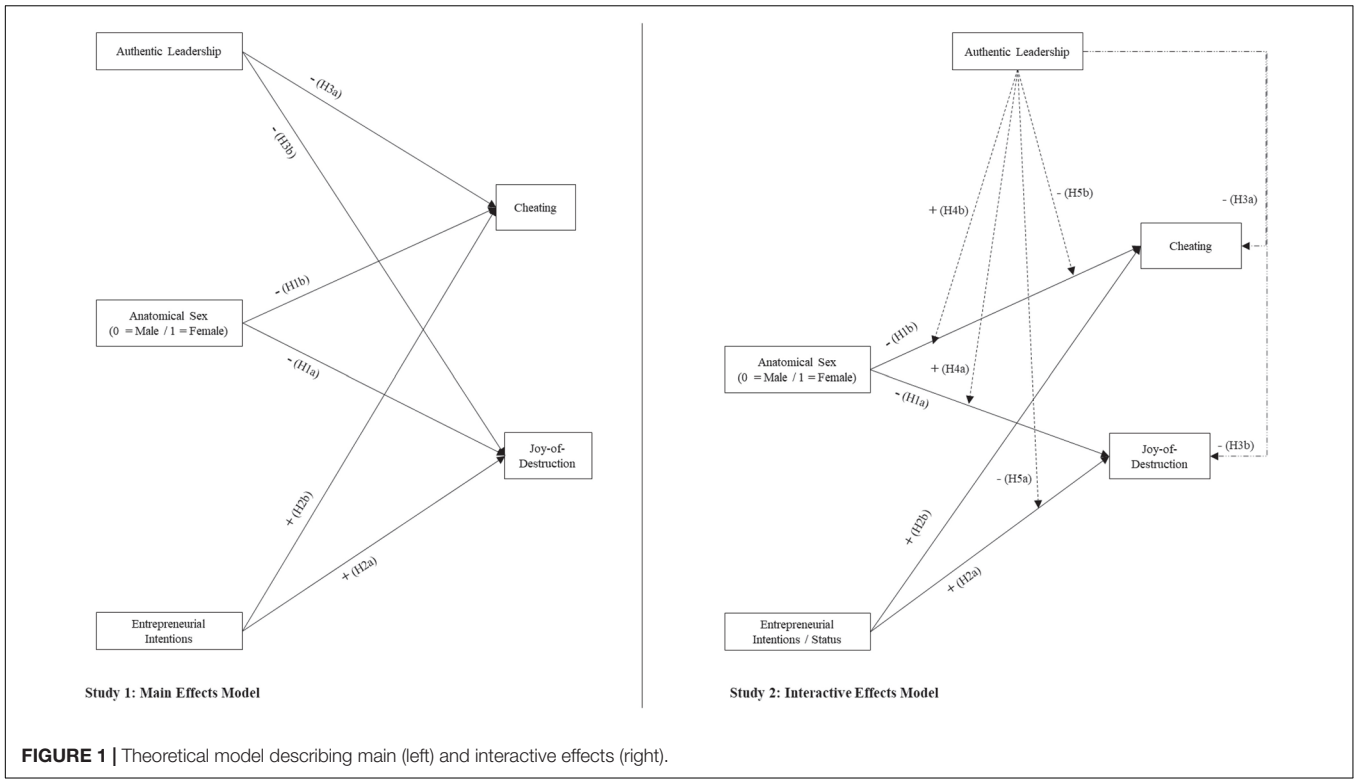

FIGURE 1 | Theoretical model describing main (left) and interactive effects (right).

with what behavioral economics defined as *self-stereotyping* (Latrofa et al., 2010). For behavioral economists, self-stereotyping occurs when an "agent perceives himself or herself as an interchangeable exemplar of a social group rather than as a unique individual" (Hernandez-Arenaz, 2020, p. 2). Social psychologists and behavioral economists seem to agree that internalized stereotypes affect an agent's economic behavior.

There are psychological risks associated with self-stereotyping. For example, self-stereotyping into a leader stereotype and embracing in-extremis its hyper-assertive prescriptions might result in seeking unethical ways to fulfill organizational goals (Ordóñez et al., 2009). Similarly, self-stereotyping into an entrepreneur role and embracing in-extremis its hyper-competitive prescriptions might result in lying to secure additional venture funding. Finally, self-stereotyping into the male gender role and embracing in-extremis the dominant and assertive behavioral prescriptions can elicit "toxic masculinity" behaviors (e.g., misogyny, homophobia, violence; Harrington, 2020).

The Female Gender Role Stereotype

In Western societies, the female role describes nurturing characteristics, such as gentleness, empathy, and support. Instead, the male gender role stereotype describes agentic characteristics, such as results-orientation and concern for advancing one's social status. The female gender role stereotype prescribes communal behaviors (e.g., concern about the well-being of others). Instead, the male gender role stereotype describes agentic behaviors (e.g., being assertive and dominant; Abele et al., 2008; Hernandez Bark et al., 2014, 2015; March et al., 2016; Hentschel et al., 2019). As mentioned above, any behavior which deviates from these stereotypical role expectations will likely elicit some form of social backlash, and particularly for women leading in organizational contexts (Gloor et al., 2018).

Adherence to gender role stereotypes can also be identified in economic games. For example, women demonstrated greater aversion toward lying for a small monetary benefit (Childs, 2012) and lower dishonesty levels than men (Friesen and Gangadharan, 2012). However, Ezquerra et al. (2018) found no gender differences in a cheating game. It follows that men (or those who identify with the male role) who self-stereotype and embrace their gender role in-extremis will more likely try to assert their dominance at work. Such a need for dominance will likely elicit antisocial economic behaviors, such as cheating and trying to harm their competition, that is, displaying *toxic masculinity*.

> *Hypothesis 1a*: Men (or those who identify with the male gender) will be more likely (a) to display antisocial behaviors aimed at harming their competition than women (or those who identify with the female gender).
>
> *Hypothesis 1b*: Men (or those who identify with the male gender role) will be more likely (a) to seek an unfair advantage by cheating than women (or those who identify with the female gender role).

The Entrepreneur Role Stereotype

The entrepreneur role stereotype collects competition-oriented traits that are seen as predictors of entrepreneurial success (need for achievement, generalized self-efficacy, individualism, risk-taking, proactive personality; Rauch and Frese, 2007;

Frese and Gielnik, 2014). Further, the entrepreneur stereotype prescribes the pursuit of wealth through the creation of new transactions (Smith et al., 2009).

Whereas competition is an inherent part of doing business, fair competition does not require entrepreneurs and business leaders to engage in antisocial economic behaviors. Yet, we claim that embracing the entrepreneur role stereotype in-extremis should elicit behaviors would appear "rational" in the traditional economic sense of the world, such as maximizing individual profits whenever possible, free-riding, and not contributing to social causes, unless when it brings an advantage for entrepreneurs. Further, another consequence of embracing in extremis the entrepreneurial role would be a higher likelihood of misrepresenting information, for example, to increase the chances to "win" further venture funding.

> *Hypothesis 2a*: Entrepreneurs (or aspiring entrepreneurs) will be more likely (a) to display antisocial behaviors aimed at harming their competition than managers (or aspiring managers).
>
> *Hypothesis 2b*: Entrepreneurs (or aspiring entrepreneurs) will be more likely to seek an unfair advantage by cheating than managers (or aspiring managers).

The Leader Role Stereotype

The leader role stereotype collects the implicit beliefs of a given social group about the "ideal" attributes that describe successful leaders. Transactional leadership is a mainstream leadership style that collects such a pattern of behaviors in western countries. Some transactional behaviors include preserving the *status quo* by rewarding with justice and actively reducing deviations from existing norms and procedures (Bass, 1985).

This transactional, behavioral pattern underlies the classic view of rational management (Zehnder et al., 2017). Again, we argue that self-stereotyping and embracing the leader role stereotype in-extremis would result in agentic behaviors aimed at increasing efficiency *at any cost* (even through unethical business practices). Stated differently, the preference for antisocial economic behaviors in business managers would evidence an in-extremis embracing of the leader role stereotype.

The above stereotypical expectations toward the leader role remain deeply rooted in Western Societies. However, the corporate scandals that led to the 2008 financial crisis challenged the perceived value of pursuing profit *at and cost* in favor of a more sustainable approach to doing business. Today, scholars care as much for "what" constitutes effective leadership as much as the "how" leaders deliver performance (Gandz et al., 2010; Monzani et al., 2016, 2019, 2021b). Authentic leadership emerged as one of many positive alternatives to the prevailing western stereotypical view of leadership (Monzani and Van Dick, 2020).

Authentic Leadership (AL) should be of interest to entrepreneurs as well. Entrepreneurs who display authentic leadership behaviors tend to feel more self-expressive when leading their ventures (Jensen and Luthans, 2006b). Further, authentic entrepreneurs elicit employee affective commitment, satisfaction, and citizenship behaviors (Jensen and Luthans, 2006a). Despite these early studies, the study of authentic entrepreneurship is in its infancy (Lewis, 2013).

One dimension of the authentic leadership style is particularly relevant for our study of antisocial economic behaviors. The dimension of "internalized moral perspective" (IMP) majorly prescribes agentic behaviors by upholding moral behaviors independently of contextual pressures (e.g., "making difficult decisions based on high standards of ethical conduct").

The ethical aspect of the IMP resonates well this some of the agentic prescriptions of the male gender role (such as being assertive). Yet, an IMP reminds leaders about the importance of adhering to existing social norms despite contextual pressures to act unethically (Monzani et al., 2015). Prior studies have shown that the more frequently leaders act coherently with their internalized moral perspective, the less likely they will engage in antisocial behaviors. Further, at least theoretically, the other three communal dimensions of AL would not prescribe antisocial behaviors. Thus, we can extend that logic into our hypotheses to claim that adopting an authentic leadership style tends to des-incentivize the display of hyper-competitive antisocial economic behaviors in favor of moral action (Hannah et al., 2011, 2014).

> *Hypothesis 3a*: As the frequency of authentic leadership behaviors increases, the likelihood of displaying antisocial economic behaviors aimed at harming their competition will decrease.
>
> *Hypothesis 3b*: As the frequency of authentic leadership behaviors increases, the likelihood of displaying antisocial economic behaviors seeking unfair advantage through cheating will decrease.

Mitigating the Effect of Stereotypical Role Expectations on Antisocial Economic Behaviors

Due to our proposed "triple bind and prejudice," female entrepreneurs should suffer a stronger prejudice than male entrepreneurs. Moreover, reconciling the stereotypical expectations toward three social roles should result in more psychological tension than their female manager counterparts. The unfortunate stereotype "Think entrepreneur – think male" (Laguía et al., 2018) captures this additional source of psychological tension. Female entrepreneurs usually struggle with limited access to venture capital, increased work-family conflict, and lack of spousal support (Das, 2000) unless they start ventures in areas congruent with stereotypical gender role expectations (e.g., social entrepreneurship; Carter et al., 2015). Increased psychological tension due to role expectations would explain why many female business students prefer a managerial position in the corporate world than starting a new venture (Jennings and Brush, 2013).

Another source of tension is the need to reconcile others' conflicting expectations of how entrepreneurs and managers should act (regardless of one's gender identity). For example, venture capitalists tend to expect entrepreneurs to be innovative by "moving fast and breaking things." Yet, the same venture capitalists expect said entrepreneurs to be efficient by "moving

slow and organizing things" (Taneja, 2019). "Organizing things" refers to developing a business strategy, making calculated decisions, and shaping norms that reduce the uncertainty inherent to any venture. Thus, to validate entrepreneurs as leaders, stakeholders demand from entrepreneurs to be visionary and managerial *at the same time* (Rowe, 2001).

To reduce the psychological tension resulting from these conflicting role expectations, individuals usually engage in "identity trade-offs" (Knapp et al., 2013). A role identity trade-off refers to following the stereotypical behavioral prescriptions of a given role (e.g., entrepreneur) to reduce the social pressure to conform to an opposing social role (e.g., gender, leader). For example, many women see in starting a new business (entrepreneur role) a functional alternative to "break free" from the societal forces that hinder their access to executive roles within corporations (leader role; Ryan and Haslam, 2007; Cook and Glass, 2014). In this way, female entrepreneurs can reduce the pressure of prevailing gender role stereotypes (female gender role), by having more latitude to balance entrepreneurial activities with their family life activities.

Authentic Leadership, Gender, and Entrepreneurial Status

From a gendered view of leadership, the authentic leadership style prescribes both agentic and communal behaviors, and thus can be classified as an androgynous style (Monzani et al., 2015). Three out of four authentic leadership dimensions to some extent overlap with the Transformational Leadership style (Banks et al., 2016), and thus prescribe communal leader behaviors (Self-awareness, Balanced Processing of Information, and Relational Transparency). More precisely, Self-awareness refers to the awareness of goals, emotions, and needs of both self *and others*. Balanced Processing of Information refers to considering *different viewpoints* before making decisions. Finally, Relational Transparency refers to establishing clear and transparent relations *with others* (Walumbwa et al., 2008).

The communal dimensions of the authentic leadership style align well with the female gender role stereotype. Such alignment could explain the findings of a recent meta-analysis, suggesting a conceptual and empirical overlap between authentic and transformational leadership when predicting several positive, growth-oriented followers outcomes (Banks et al., 2016). The overlap between transformational leadership and authentic suggests that some of the insights of RCT might as well apply to the communal dimensions of AL, and thus allows predicting potential interactive effects between gender and leader role stereotypes.

The fact that such overlap exists might have implications for entrepreneurs as well. For example, as entrepreneurs increase the frequency of their authentic leadership behaviors when running their ventures, in turn, should increase entrepreneurs' concern on how their actions impact others. Such concern should reduce the likelihood of displaying antisocial behaviors. Therefore, in this follow-up study, we propose the two additional hypotheses. The right panel of panel of **Figure 1** summarizes our additional predictions:

> *Hypothesis 4*: Authentic leadership moderates the effect of the male gender role on the likelihood of harming others' firms (H4a) and cheating (H4b). As the frequency of authentic leadership behaviors increase, men will be less likely to display said antisocial behaviors.
>
> *Hypothesis 5*: Authentic leadership moderates the effect of the entrepreneurial role on the likelihood of harming others' firms (H5a) and cheating (H5b). As the frequency of authentic leadership behaviors increase, men will be less likely to display said antisocial behaviors.

METHODS

We tested our hypotheses in two laboratory studies. Our first laboratory study (Phase 1) was conducted in a sample of French Business school students (N = 82). However, this sample had some limitations (culturally heterogeneous, aspiring leaders and entrepreneurs). To address such limitations, we conducted a follow-up study (Phase 2). During Phase 2, we re-tested our predictions in a more homogeneous sample and explored interactive effects among predictors. More precisely, we needed a societal context that valued "tradition" (i.e., reinforces the female gender role stereotype) and "benevolence" (i.e., preserving and enhancing the welfare of those with whom one is in frequent contact).

Our rationale for choosing such a societal context is that we anticipate that in societies that simultaneously embrace the universal values of "tradition" and "benevolence," the psychological tensions between conflicting role stereotypes would become more salient for female entrepreneurs than in other societies. On one side, a traditional society tends to pressure female citizens to find meaning by starting a family rather than a business.

On the other side, benevolent societies tend to value ventures that transcend the pure and single pursuit of profit. We would not expect the same level of psychological conflict in societies that score high in the universal value of benevolence and self-direction. Benevolence and self-direction do not seem to be at odds (i.e., should not create such a strong psychological tension when women occupy an entrepreneurial role).

In prior studies, Costa Rica scored 30.4% higher than Canada in the universal value of "Tradition" (M = 5.25, SD = 1.47 vs. M = 4.57, SD = 1.23 respectively). However, in the same study Costa Rica also matched the US in the value of "Benevolence" (M = 6.20, SD = 1.05 vs. M = 6.19, SD = 0.96 respectively; Schultz and Zelenzy, 1999). With such findings in mind, the Costa Rican society would be sending ambiguous signals about the value of entrepreneurship to their female citizens (or those who identify with the female gender role).

As a result of such mixed signals and ambiguity, Costa Rica seems to be a pristine context to explore how female entrepreneurs reconcile the conflicting pressure of multiple social

stereotypes. Further, it allows us to test if female entrepreneurs will display antisocial economic behaviors when leading their ventures in a society that does not value nor socially reward such antisocial economic behaviors. On these grounds, we chose to conduct the second laboratory study (Phase 2) in Costa Rica. The second laboratory study is based on a sample of Costa Rican male and female entrepreneurs, taking male and female managers as reference groups ($N = 64$).

Sample

For phase 1, our sample consisted of 82 students who attended business management courses at a French School of Business. The mean age was 22.37 years ($SD = 1.95$). A large part of our sample consisted of international students (57.3%). Thirty-five participants (42.7%) came from Mediterranean countries, Thirty (36.6%) from Asian countries, nine (11.0%) from Latin-American countries, three (3.7%) came from African nations, and three (3.7%) from Middle Eastern countries, two participants did not indicate their nationality. Twenty-one participants were male, sixty female, and one participant did not indicate his or her anatomical sex. After removing cases with missing data, our final sample for phase 1 consisted of 77 participants.

To address the limitations of Phase 1, in Phase 2 we invited traditional entrepreneurs ($N = 20$; 55.0% female) and managers from a public organization ($N = 44$; 43.2% female) to participate in our laboratory study. Entrepreneurs' age $M = 40.11$; $SD = 10.42$ and Managers' age was $M = 44.58$, $SD = 8.28$. Both the entrepreneurs ($M = 8.44$, $SD = 10.54$) and managers ($M = 18.70$, $SD = 19.13$) had several employees under their charge. 65.0% of our entrepreneurs owned a family company, and 10.5% only had high school education, 47.5% had a bachelor's degree or equivalent, and 42.1% had a post-graduate degree (e.g., MBA). Our entrepreneurial sample represented several work sectors, with financial services, planning, and communications the most numerous areas (6.3% each), followed by services, logistics, administration, and biochemical (4.7% each). 7.8% of the participants did not indicate their sector. Managers mostly supervised clerical employees.

Procedure

As participants entered the lab, the experimenter randomly assigned each participant to a cubicle. All the participants answered a self-report survey for 25 min before the laboratory task started (capturing age, anatomical sex, and entrepreneurial intentions). Immediately after, participants provided self-reports of authentic leadership and social desirability (as a consistency check).

The laboratory task consisted of several activities. First, a couple of activities collected information on our behavioral control variables. More precisely, an arithmetic exercise was used as a proxy variable of participants' cognitive ability. A risk aversion game followed our arithmetic exercise. We conducted the risk aversion game because meta-analyses revealed that individuals displaying behaviors aligned with both the female gender and managerial role stereotype declare a higher risk aversion than those individuals displaying behaviors aligned with the male gender role and the entrepreneur role stereotype (Rauch and Frese, 2007; Stewart and Roth, 2007).

In comparison, a lower risk aversion aligns better with the female gender role and the manager role stereotypes. Finally, participants undertook our two antisocial economic behavior games ("Joy-of-Destruction" and "Cheating"). After the study, all participants were debriefed about the nature of the study and received a $5 show-up fee and their respective earnings from the economic games that comprised this study's laboratory task. Participants' earning ranged from $2.97 to $11.75 ($M = $7.59; $SD = 1.85$).

For Phase 2, we employed the same procedure as in Phase 1, with a slight modification. After the risk aversion activity, we added a "one-shot" public goods game that captured participants' preference for pro-social vs. pro-individual strategizing. A pro-social strategizing aligns with the female gender role and pro-individual strategizing with the male gender role.

Further, we felt it unnecessary to assess actual leaders and entrepreneurs' cognitive ability. Instead, we collected an array of demographic characteristics. More precisely, we measured (a) Span of Control, meaning the number of employees supervised, (b) leading in a family company (dummy coded as 0 = "No"/1 = "Yes"), and (c) work tenure as leader, measured in years.

All participants provided informed consent to participate in the laboratory study and had no prior knowledge of the study's objectives. Every participant was initially endowed with the same quantity of resources (100 tokens, equal to 10 EUR), allocated to a private account, and paid off at the end of the laboratory study. Total earnings were calculated as the sum of all the earnings obtained in all the games.

Participants' earning ranged from $2.00 to $9.86 ($M = $6.28; $SD = 1.78$). At the end of the session, we offer the possibility of exchanging their monetary payoffs with souvenirs from the university, such as coffee mugs, t-shirts, caps, and pens. Most participants preferred the souvenirs to the monetary compensation.

Measures

Phase 1

Entrepreneurial intention

Given that our sample consisted of business students (and not actual entrepreneurs), we measured participants' entrepreneurial intention by asking participants about the likelihood of starting a venture after graduation. The item "How likely is that you would start a venture after you graduate?" was rated on a 5-point Likert-type scale, with values ranging from "1 = "Extremely unlikely" to "5 = "Extremely likely." Although a self-report scale of entrepreneurial intention exists (Liñán and Chen, 2009), the items that comprise the subscale of interest revealed that all items referred to the same notion. Therefore, we used a single item from Liñán and Chen's (2009) sub-scale in this laboratory study.

Authentic leadership

We asked participants to self-report the frequency of their AL behaviors using the Authentic Leadership Questionnaire (Walumbwa et al., 2008). All sixteen items were rated on 5-point

Likert-type scales, with values ranging from "1 = Not at all" to "5 = Frequently, if not always." Some examples of items are "Seeks feedback to improve interactions with others" (Self-awareness), "Says exactly what he or she means" (Relational Transparency), "Makes decisions based on his/her core beliefs" (internalized moral perspective), and "Listens carefully to different points of view before coming to conclusions" (Balanced Processing of Information). Cronbach's was $\alpha = 0.70$ for the overall scale for Phase 1, and Cronbach's α was 0.81 for Phase 2.

Antisocial economic behaviors

We captured two antisocial economic behaviors (harming others and cheating) by employing simplified versions of the "Joy of Destruction" (Abbink and Sadrieh, 2009) and "Cheating" games (Fischbacher and Föllmi-Heusi, 2013). Such games present players with a simplified version of real-stakes business decisions. Both games were scored with a binary outcome (i.e., 0 = "No," 1 = "Yes").

In the "Cheating" game (Fischbacher and Föllmi-Heusi, 2013), a player is provided with specific information about a product, and its asked to report such information to others (e.g., shareholders) for a pre-determined payoff (Dummy coded as "0"). However, when reporting said information, the player is provided with a choice, which mainly consists of misrepresenting the information provided in exchange for a higher individual payoff (dummy coded as "1"). Selecting "1" would capture a cheating behavior.

In our game, we reduced the complexity of the decision-making process by proposing a simplified version of the cheating game. Individuals were asked to report the color of a ball from an urn, knowing that the red ball reported 4 US dollars, the blue ball reported 2 US dollars, and the green ball reported no gains. Because all the balls were green in color, we measured a cheating behavior when individuals chose red and blue balls by simply asking the question of what color is the ball.

In our variation of the "Joy-of-Destruction" "game," each player is presented with the chance to harm the competition (i.e., a player randomly matched at the beginning of the game) at no additional cost to the player's firm (choosing "1 = Yes" captures a destruction preference). Such a decision has been validated as a measure of destructive behavior (see Abbink and Herrmann, 2011). In particular, we used a simplified version of the game in which individuals (Players A) were endowed with 3 US dollars and matched with an anonymous passive participant (Player B) which received 10 US dollars. The only question that Players A received was about to reduce the other's endowment in 7 US dollars. All the participants played simultaneously as Player A and B for payment effects.

Phase 2

Authentic leadership

Again, participants self-reported their authentic leadership using the ALQ (Walumbwa et al., 2008; Avolio et al., 2018). Items were rated on a 5-point Likert-type scale (1 = "Not at all" to 5 = "Frequently, if not always"). Although the ALQ has been validated for the Iberian, Spanish-speaking population (Moriano et al., 2011), linguistic differences exist between Iberian and the language spoken in Costa Rica, Latin-American Spanish. Consequently, we followed Brislin (1980) guidelines to translate the original questionnaire into Latin-American Spanish. The first author, a native Latin-American Spanish speaker, translated the original items of the ALQ scale into Latin American Spanish and required a consistency check from four Latin-American research assistants (blind to the laboratory study). Finally, the translated copy was provided to an English professional translator for re-translation into English. No linguistic differences between the original and back-translated scale emerged. In Phase 2, we used the same two games employed in Phase 1 with the same decision options and pay-out functions.

Control Variables

Phase 1

Cognitive ability

Arithmetic ability was taken as a proxy for cognitive ability (Hyde et al., 1990). Cognitive ability is a trait of successful entrepreneurs (Frese and Gielnik, 2014) and is regarded as the main predictor of performance (Schmidt and Hunter, 2003; Kanfer and Kantrowitz, 2005). Participants were asked to complete 30 simple arithmetic calculations in 30 s and were rewarded with $ 0.10 for every correct answer.

Social desirability

We used a 12 item scale of social desirability by Caprara et al. (1993). This construct captures a person's tendency to display an enhanced image of him or herself. Cronbach's Alpha was $\alpha = 0.82$ for Phase 1. In Phase 2, Cronbach's Alpha was of $\alpha = 0.85$.

Risk aversion

Despite the findings of Filippin and Crosetto (2016), which concluded that effect of gender differences on risk aversion appears in less than 10% of their review studies, taking risks is an expected stereotypical behavior of entrepreneurs. For entrepreneurs, a higher risk aversion would likely lead to social backlash or punishment. Therefore, embracing the entrepreneur role stereotype in extremis should lead to excessive risk taking.

In our game, participants were asked to choose one of three lotteries, each with a different degree of risk which was established by throwing a virtual coin. Participants were endowed with one US dollar and were asked about not playing any lottery (option A = $1), increasing payoffs and losses by 50% (option B = $0.5/$1.5), or increasing them by 100% (option C = $0/$2). After the lottery choice (among the three options), random plays determined participant's payoffs. The random nature of the lottery captures participants' inability to calculate the risk of their choice. We consider a participant to be risk adverse when he or she selected option A.

Phase 2

For phase 2, we controlled for the participants' demographic characteristics that might influence their economic behavior, as suggested by existing studies. Again. we controlled for (a) social desirability, (b) whether if the entrepreneur's venture was a family business or not (coded "0" for managers as well), (c) their span of control (number of supervised employees), and (d) work tenure as leaders (for entrepreneurs and managers).

Pro-individual vs. pro-social strategizing

As mentioned above, we used a Public Goods Game (PGG) to capture entrepreneurs' economic behavior that might evidence a preference for social entrepreneurship. One type of PGG is the Voluntary Contribution Mechanism (Keser, 2002; Chaudhuri, 2011). In such a cold-strategy public good game, participants create wealth by adopting pro-individual or pro-social contribution strategies.

The pro-individual strategy consists of capturing part of the shared pool without contributing substantially to the public good (and thus conforming to the entrepreneurial role stereotype). The pro-social strategy involves contributing substantially to the public good and trusting that others will contribute as well. A pro-social strategizing would evidence a preference for social entrepreneurship, as reflected by a higher contribution to the public good, that those with a pro-individual strategizing (a preference for traditional entrepreneurship). This measure ranged from "0" = pro-individuals strategizing up to "100" = pro-social strategizing.

Data Analysis

Phase 1

We tested our hypotheses by building a structural equation model in MPLUS 8.0. MPLUS 8.0 allows employing robust estimators, such as the Weighted Least Squares – Mean and Variance Adjusted (WSLMV). The WLSMV allows analyzing models comprising dichotomic variables, calculates well parameters estimates with relative small datasets, and adjusts for deviations of multivariate normality (Moshagen and Musch, 2014).

To assess our SEM model's fit, we employed the Satorra-Bentler scaled chi-square test and additional Goodness of Fit Indicators. The S-B chi-square test indicates a good model fit when it is non-significant (Geiser, 2011), with the caveat that the S-B chi-square test is sensitive to large sample sizes. Therefore, it is a good practice to complement the S-B chi-square test with additional goodness-of-fit indicators. Some examples are the $\chi 2$/df ratio, the Root Mean Square Error of Approximation (RMSEA), the Comparative Fit Index (CFI), and the Tucker-Lewis indicator (TLI), as well as and the Standardized-Root-Mean-Square-Residual (SRMR).

The comparative fit index (CFI) measures incremental fit whereby values higher than 0.90 and ideally above 0.95 are required to avoid incorrectly accepting miss-specified models. Similarly, the TLI is an indicator of model parsimony equivalent to the NNFI. Again, values above 0.90 (and ideally above 0.95) are preferred. CFI and TLI values close to 1 indicate that the model explains the data better than an independence model (Hu and Bentler, 1999). The RMSEA tests for approximate data fit, and it should be at least equal to 0.08 or below. Standardized-Root-Mean-Square-Residual (SRMR) provides an overall evaluation of the residuals and it is considered acceptable when it approximates the 0.08 value (Hu and Bentler, 1999).

Finally, we parceled any multi-dimensional measure in our study. Parceling refers to aggregating the respective items of a scale's dimension to reduce the overall parameters to be estimated in an SEM (see Monzani et al., 2021a; Seijts et al., 2021 for examples of parceling), and thus. Whereas this technique has received some critiques, Little et al. (2009) argued that parceling is justified when the underlying factorial structure has been previously established in the literature and when the parceled indicators respect said factorial structure. Given that both our Social desirability measure and the ALQ have been validated in multiple samples worldwide (Walumbwa et al., 2008), their parceling is justified.

Phase 2

We used SPSS 25 to conduct hierarchical binary logistic regressions. To avoid multicollinearity, we normalized scores for all our continuous independent variables before computing any interaction term. Anatomical sex was coded into 0 = "Male" and 1 = "Female," and entrepreneurial status as 0 = "Manager" and 1 = "Entrepreneur."

We mainly entered our demographic and control variables (social desirability; Family Company; Span of control; and Work tenure; pro-social strategizing; Risk Aversion and either Cheating or Joy-of-Destruction, respectively). Then, we entered our predictors (entrepreneurial intentions, biological gender, and authentic leadership scores). Finally, we included our two cross-product terms. We used Dawson (2013) Microsoft Excel templated to illustrate any non-linear interaction effects.

Given that a binary logistic regression uses a maximum likelihood approach, SPSS 25 provides goodness-of-fit indices. These indices allow assessing if (a) a model correctly classifies predicted cases into their observed categories, (b) how well the model fits the observed data, (c) and whether if each model step improves the fit of the model to the observed data in hierarchical models. First, the cutoff value to evaluate the sensitivity of a model is 50%. Higher percentage scores represent a higher sensitivity of the model. In social sciences, a test sensitivity of 50–60% is considered poor, from 60 to 70% is adequate, 70 to 80% is good, and above 80% is very good. Any percentage equal to or below 50% would mean that the model has equal or fewer chances of classifying cases correctly than a coin toss.

The second goodness of fit indicator is the omnibus test of the model. The omnibus test captures how much our model deviates from a null model (a model only with the intercept and no additional variables) or the previous step if a hierarchical regression approach is used. For this indicator, higher χ^2 scores indicate a better fit; a statistically significant χ^2 value would indicate that such deviation did not occur by chance.

The third set of goodness of fit indicators is based on the deviance statistic (–2LL), which follows a chi-square distribution. Researchers employ –2LL statistic to derive pseudo-R^2 statistics (e.g., Cox & Snell R^2 and Nagelkerke R^2). In short, these two statistics range from 0 to 1 and indicate in relative terms how well a model fits the data, which scores closer to 1 suggesting a better fit. Fourth, the Hosmer & Lemeshow test is analog to the χ^2 test used in SEM modeling. The *lower* that χ^2 score is, the better that the model fits the observed data. Finally, a non-significant p-value would indicate that the model fits the data well.

RESULTS

Phase 1 – Hypothesis Testing

Table 1 shows Means, Standard Deviations, and both Pearson's *r* (product-moment correlation) and Kendall's τ (tau) in the upper and lower diagonals, respectively. Entrepreneurial intentions were negatively related to the female anatomical sex (Kendall's $\tau = -0.22^*$). **Figure 2** shows the standardized coefficients resulting from our SEM analysis. For parsimony, only significant paths are shown. Further, whereas solid lines represent main effects, dotted lines represent either indirect effects or corrected (or latent) correlations between our constructs.

The results of our SEM analysis revealed that in overall, our model fit showed an excellent fit to the data ($\chi^2_{(33)} = 38.87$, ns; $\chi^2/df = 0.93$; RMSEA = 0.0001; CFI = 1.00; TLI = 1.00; SRMR = 0.11). In consequence, the standardized effect sizes and standard errors derived from this model are trustworthy.

A detailed inspection of **Figure 2** shows that after controlling for the effects of all the other variables in our model, our two dependent variables (participants' choices in the "Joy-of-Destruction" and Cheating games) were not correlated [$r = 0.14$ (0.19), *ns*]. Further, predictors majorly explained a statistically significant amount of variance for the "Joy-of-Destruction" game ($R^2 = 0.23$, $p < 0.05$), but not for the cheating game ($R^2 = 0.23$, $p < 0.05$).

Second, neither of our control variables (Social desirability; Arithmetic ability, nor a Risk Aversion preference) had statistically significant main effects on neither the "Joy-of-Destruction" nor the Cheating games. Instead, whereas none of our independent variables were significant predictors of participants' behavioral choices in the cheating behavior, all three independent variables were significant (and negative) predictors of participants' behavioral choices in the "Joy-of-Destruction" game. More precisely, Anatomical Sex [β = –0.30 (0.15), $p < 0.05$]; Entrepreneurial Intentions [β = –0.32 (0.12), $p < 0.01$] and Authentic leadership [β = –0.32 (0.15), $p < 0.05$] reduced the likelihood of observing an antisocial behavior aimed at harming one's competition. When taken as a whole, these results support Hypotheses 1a, 2a, and 3a but do not support Hypothesis 1b, 2b, and 3b.

Phase 1 – *Post hoc* Analyses

We conducted additional two *post hoc* analyses as per the results of our SEM model. First, we attempted to replicate the prior findings in the entrepreneurship literature regarding gender differences in entrepreneurial intentions (Jennings and Brush, 2013). To this end, we specified an additional (non-hypothesized) path between participants' Anatomical sex and their self-reported Entrepreneurial Intentions.

We expected to find a negative effect of Anatomical Sex on Entrepreneurial Intentions because female participants were dummy coded into the "1" category. Our results revealed that Anatomical Sex had a relatively strong negative effect on Entrepreneurial Intentions (β = –0.94; $p < 0.01$). This result means that in our sample, female participants tended to declare a weaker intention of starting up a business after they graduate from their business programs. Second, we attempted to integrate this non-hypothesized finding with our prior results. To this end, we tested if Anatomical Sex would have an indirect effect on participants' behaviors choices for the "Joy-of-Destruction" game (recall that we did not find a main effect of Anatomical Sex on behavioral choices on the Cheating game). By using the INDIRECT function in MPLUS 8.0, we detected a significant and positive indirect effect of Anatomical Sex on participants' behavioral choices on the "Joy-of-Destruction" game, as mediated by Entrepreneurial Intentions (β = 0.13 (0.06); $p < 0.05$). In other words, those female participants who see themselves as entrepreneurs in the future seem to embrace the ultra-competitive prescriptions of the entrepreneurial role stereotype.

Our findings of main and indirect effects with opposing signs align with our theorizing. This last result evidences the psychological tension that women declaring entrepreneurial intentions suffer. Women chose not to hurt their competition in the Joy-of-Destruction game (as prescribed by the female gender role). However, this choice was nuanced by a weaker yet positive and significant indirect effect (mediated by entrepreneurial intentions), and likely driven by participants' stereotypical views of entrepreneurship.

Phase 2 – Hypothesis Testing

Table 2 shows the means, standard deviations, Pearson's and Kendall's correlations for all variables in this study. Being an entrepreneur was strongly and positively correlated with having a family company ($\tau = 0.75^{**}$) and the time leading others ($\tau = 0.41^{**}$), but negatively correlated with Span of Control ($\tau = -0.27^*$) and self-ratings of authentic leadership ($\tau = -0.27^{**}$).

The left panel of **Table 3** shows the results of our logistic regression model predicting participants' likelihood of choosing to harm others' firms. Our model was trustworthy and had good sensitivity (73.8%), significantly deviated from the null model ($\chi^2_{(12)} = 25.79^*$) and fitted the observed data well (H&L Test = $\chi^2_{(8)} = 8.14$ ns). Three control variables were significant predictors. More precisely, Social Desirability [$B = -0.93$, (0.47); Wald's $Z = 3.98^*$], owning a family company [$B = -5.40$, (2.80); Wald's $Z = 3.72^*$] and work tenure [$B = -1.78$, (0.88); Wald's $Z = 4.14^*$] reduced the likelihood of choosing to harm others' firms ("Joy-of-Destruction").

Regarding the main effects of our independent variables, Anatomical Sex [$B = -1.81$, (0.85); Wald's $Z = 4.56^*$] was again a negative predictor of a "Joy-of-Destruction" preference, suggesting that men are more likely to choose to harm others' firms than women. Instead, our participants' Entrepreneurial status did not have a main effect [$B = 0.93$, (1.33); Wald's $Z = 0.49$ ns]. Finally, like Anatomical Sex, Authentic leadership had a negative effect [$B = -1.68$, (0.80); Wald's $Z = 4.41^*$], meaning as the frequency of authentic leadership behaviors decreased, the more likely participants were to choose to harm others' firms. Overall, these results support H1a and H3a but again do not support H2a (see **Figure 3**).

Both interaction terms predicting the "Joy-of-Destruction" preference were statistically significant. Anatomical Sex [$B = 2.44$, (1.05); Wald's $Z = 5.35^*$] and Entrepreneurial status [$B = -2.99$,

TABLE 1 | Phase 1 – Means, standard deviations, Kendall's τ and Pearson's r for all study variables.

		M	*SD*	0.1	0.2	0.3	0.4	0.5	0.6	0.7
(1)	Cognitive Ability	1.15	0.38	–	−0.18	0.20	−0.01	0.11	0.01	0.11
(2)	Risk Aversion	1.49	0.50	−0.14	–	−0.13	−0.03	−0.11	0.12	0.07
(3)	Anatomical Sex	0.74	0.44	0.10	−0.13	–	−0.20	0.06	−0.13	−0.04
(4)	Entrepreneurial Intention	3.45	1.08	0.03	0.01	−0.22*	–	0.10	−0.15	0.14
(5)	Authentic Leadership	2.96	0.33	0.10	−0.10	0.06	0.05	–	−0.20	−0.07
(6)	Antisocial Behavior-Joy of Destruction game	0.61	0.49	0.04	0.12	−0.13	−0.16	−0.18	–	0.14
(7)	Antisocial Behavior-Cheating game	0.59	0.50	0.12	0.07	−0.04	0.12	−0.07	−0.14	–

$p < 0.05$. The lower diagonal presents parametric correlations in the upper diagonal (Pearson's r) and non-parametric correlations in the lower diagonal (Kendall's tau) given that several variables are dichotomous.

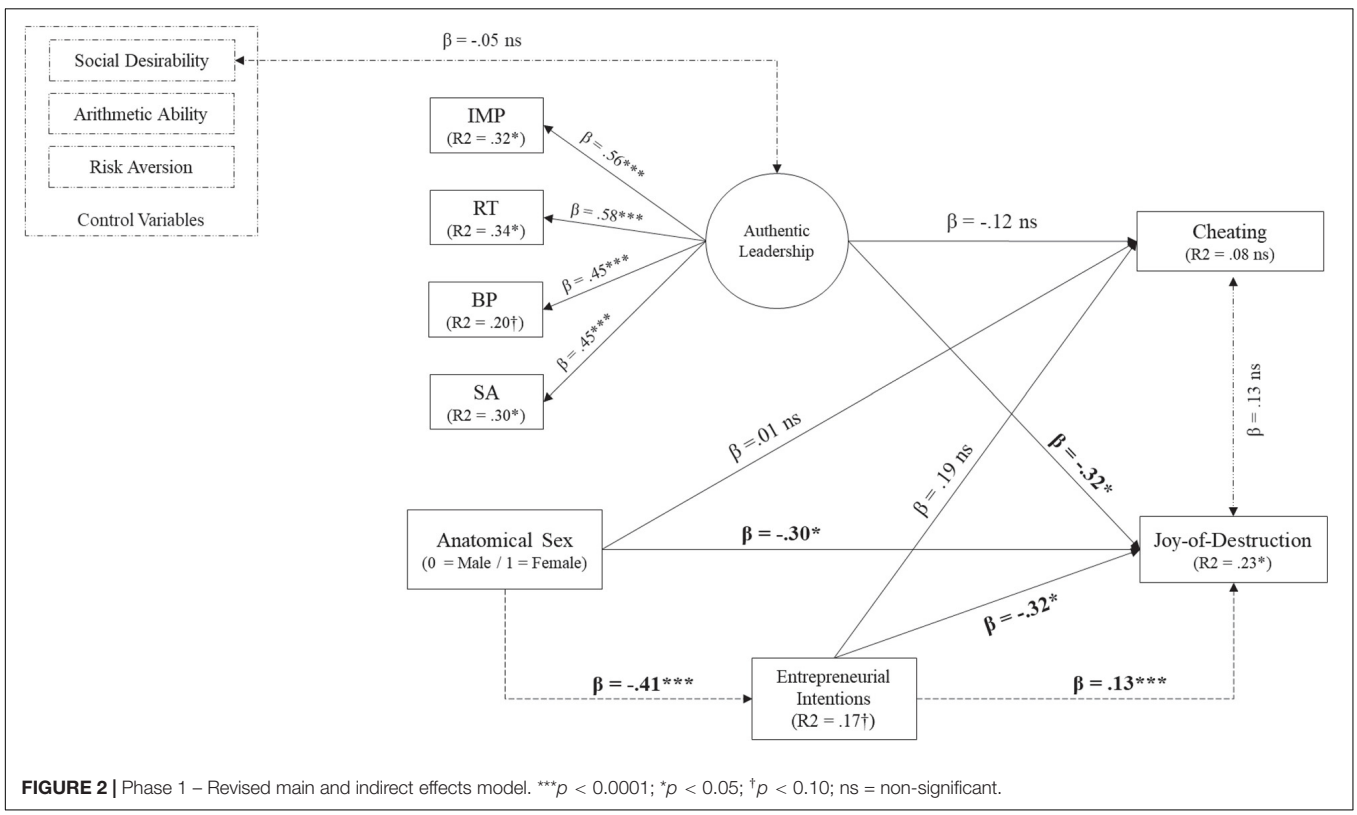

FIGURE 2 | Phase 1 – Revised main and indirect effects model. ***$p < 0.0001$; *$p < 0.05$; †$p < 0.10$; ns = non-significant.

TABLE 2 | Phase 2 – Means, standard deviations, Kendall's tau and Pearson's bivariate correlations for all study variables.

		M	*SD*	1	2	3	4	5	6	7	8	9	10
(1)	Family Company	0.20	0.40	–	−0.32*	0.50**	−0.08	−0.03	−0.01	0.75**	−0.35**	−0.17	−0.16
(2)	Span of Control	15.86	17.72	−0.35*	–	−0.22	−0.04	0.05	−0.12	−0.26*	0.12	−0.16	0.02
(3)	Work Tenure (as Leader)	7.55	7.68	0.30**	−0.14	–	−0.10	−0.08	−0.08	0.51**	−0.05	−0.27*	−0.20
(4)	Risk Aversion	1.32	0.47	−0.08	−0.05	−0.10	–	−0.16	−0.16	−0.08	0.12	−0.08	0.15
(5)	Social Desirability	0.94	0.45	−0.16	0.01	−0.13	−0.16	–	0.10	−0.12	−0.25*	0.01	−0.14
(6)	Anatomical Sex	0.47	0.50	−0.01	−0.03	−0.02	−0.16	−0.03	–	0.11	−0.15	−0.05	−0.24
(7)	Entrepreneurial Status	0.31	0.47	0.75**	−0.27*	0.41**	−0.08	−0.12	0.11	–	−0.33	−0.19	−0.17
(8)	Authentic Leadership	3.31	0.34	−0.28**	0.14	0.01	0.09	0.08	−0.10	−0.27*	–	−0.12	−0.03
(9)	Antisocial Behavior-Joy of Destruction game	0.39	0.49	−0.16	−0.04	−0.22*	−0.07	−0.05	−0.05	−0.19	−0.11	–	0.07
(10)	Antisocial Behavior-Cheating game	0.20	0.41	−0.16	−0.14	−0.25*	0.15	0.07	−0.24	−0.17	−0.05	0.07	–

***$p < 0.0001$; **$p < 0.01$; *$p < 0.05$; The lower diagonal presents parametric correlations in the upper diagonal (Pearson's r) and non-parametric correlations in the lower diagonal (Kendall's tau) given that several variables are dichotomous in nature.*

TABLE 3 | Phase 2 – Logistic Regression model predicting the likelihood of displaying two antisocial behaviors (*N* = 62).

	Joy-of-Destruction Game					Cheating Game				
	β	SE β	Wald's Z	df	Exp (β)	β	SE β	Wald's χ^2	df	Exp (β)
Constant	3.15	1.35	5.47*	1	23.26	−0.52	0.86	0.37	1	0.36
Social Desirability	−0.93	0.47	3.98*	1	0.39	−0.08	0.49	0.30	1	0.92
Family Company	−5.40	2.80	3.72*	1	0.01	−2.09	2.79	0.56	1	0.12
Span of Control (N Employees)	−1.09	0.58	3.58†	1	0.34	−0.14	0.38	0.13	1	0.87
Work Tenure	−1.78	0.88	4.14*	1	0.17	−1.32	0.87	2.30	1	0.27
"Cheating" Behavior (1 = Yes)	−1.04	0.92	1.28	1	0.35	–	–	–	–	–
"Joy-of-Destruction" Behavior (1 = Yes)	–	–	–	–	–	−0.60	0.90	0.46	1	0.54
Risk Aversion (1 = High)	−2.01	0.89	5.06*	1	0.13	−0.09	1.02	0.01	1	0.91
CPG	−0.52	0.37	2.00	1	0.59	−0.80	0.47	2.81†	1	0.12
ES – (1 = Entrepreneur)	0.93	1.33	0.49	1	2.53	0.53	1.56	0.11	1	1.69
AS (1 = Female)	−1.81	0.85	4.56*	1	0.16	−2.14	0.98	4.77*	1	0.73
Authentic Leadership	−1.68	0.80	4.41*	1	0.19	−1.90	0.95	3.97*	1	0.15
Authentic Leadership x BG.	2.44	1.05	5.35*	1	11.49	2.76	1.31	4.43*	1	15.73
Authentic Leadership x ES.	−2.99	1.51	3.89*	1	0.05	−1.01	1.77	0.33	1	0.36
Goodness-of-fit Indicators										
Correctly Classified Cases			73.8%							77.0%
Deviation from Null Model			$\chi^2_{(12)} = 25.79^{**}$							$\chi^2_{(12)} = 25.79^{**}$
Hosmer & Lemeshow Test			$\chi^2_{(8)} = 8.14$ ns							$\chi^2_{(8)} = 11.08$ ns
Pseudo R^2	–2LL =	56.78	C&S $R^2 = 0.34$		N – $R^2 = 0.46$	–2LL = 56.78		-2LL = 45.06		C&S $R^2 = 0.26$

†$p < 0.10$; *$p < 0.05$; **$p < 0.01$; *All continuous variables were standardized. AS, Anatomical Sex; ES, Entrepreneurial Status; CPG, Contribution to the Public Good; C&S – R^2, Cox and Snell pseudo R^2; N – R^2, Nagelkerke's pseudo R^2. The accepted cut-off for case classification is 50%. Scores above 70% evidence a good classification ability of the model. Similarly, A non-significant score in the Hostmer & Lemeshow test suggest a good fit of the model to the data.*

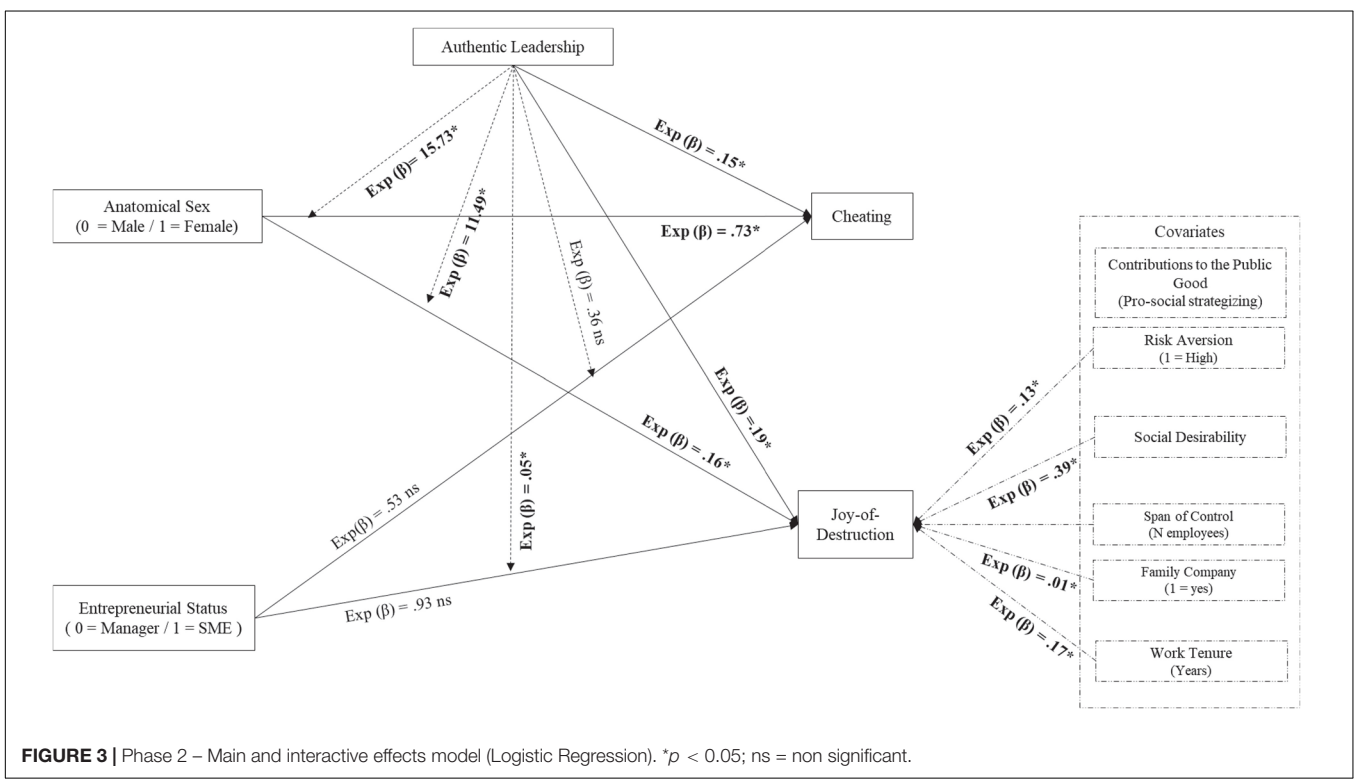

FIGURE 3 | Phase 2 – Main and interactive effects model (Logistic Regression). *$p < 0.05$; ns = non significant.

(1.51); Wald's Z = 3.89*] interacted with Authentic leadership in reducing participants' likelihood of choosing to harm others' firms. More precisely, as either men or entrepreneurs scored higher in authentic leadership, the likelihood of harming others' firms decreased. **Figure 4** illustrates these moderator effects. These results provide initial support for hypothesis 4a and 5a.

The right panel of **Table 3** shows the results of our second logistic regression predicting the likelihood of participants' cheating. Our model showed good sensitivity (77.0%), and this time, it significantly deviated from the null model ($\chi^2_{(12)} = 25.79$) and fitted the observed data well (H&L Test = $\chi^2_{(8)} = 11.08$ ns). None of our control variables were significant predictors. Instead of our independent variables, again Anatomical Sex [B = –2.14, (0.98); Wald's Z = 4.77*] was a negative predictor, suggesting that, in general, men are more likely to cheat than women. Something similar occurred for Authentic leadership [B = –1.90, (0.95); Wald's Z = 3.97*], meaning that as the frequency of authentic leadership behaviors increased, participants were less likely to cheat. Finally, our participants' entrepreneurial status did not have a main effect [B = 0.53, (1.56); Wald's Z = 0.11 ns]. Thus, our results support H1b but do not support H2b.

Figure 5 shows that only Anatomical Sex [B = 2.76, (1.31); Wald's Z = 4.43*] interacted with Authentic leadership in reducing the likelihood of participants' cheating. As men scored higher in Authentic leadership, the likelihood of participants cheating decayed. Thus, we found support for H4b but not for H5b.

DISCUSSION

Our study's main goal was to explore whether if role stereotypes drive entrepreneurs to engage in antisocial economic behaviors. More precisely, we proposed extending Eagly and Karau's (2002) RCT to the entrepreneurial arena and testing the existence of a potential entrepreneurial role stereotype, as well as a female-entrepreneurship role conflict (Laguía et al., 2018). The results of Phase 1 revealed that female business school students tend to declare weaker entrepreneurial intentions than men. Further, women are less likely to choose to harm their competitors in an economic game ("Joy-of-Destruction"). However, a *post hoc* analysis revealed that this reluctance to harm other firms is reduced when entrepreneurial intentions mediate this link.

Building on Eagly and Karau's (2002) theory, we claim that the entrepreneur role stereotype captures dominant traits and prescribes competitive behaviors that align with the male gender role stereotype. Still, we distinguish it from the assertive traits and behaviors prescribed by the leader role stereotype. Consequently, we propose two new role incongruences, namely, the gender-entrepreneur and leader-entrepreneur incongruencies.

In short, the female-entrepreneur incongruence would explain why women (or those persons that identity with the female gender role) resist occupying entrepreneurial roles (Jennings and Brush, 2013), a result we confirmed in Phase 1. Further, the female-entrepreneur incongruence would explain why those women that occupy an entrepreneurial role tend to gravitate toward communal-oriented ventures instead of pursuing ventures in more traditional sectors (Datta and Gailey, 2012).

Instead, although not tested in this work, we argue for a leader-entrepreneur role conflict that would explain why some individuals become serial entrepreneurs (rejecting the leader role prescriptions of managing a venture to its mature state), and other entrepreneurs eventually gravitate into managerial roles in other's firms (rejecting the entrepreneur role prescription of creating wealth through a new venture creation).

A core premise of this study is that self-stereotyping and embracing a role stereotype "in extremis" is a dysfunctional way of reducing these role incongruencies. Our results suggest that women or entrepreneurs who do so will likely end up displaying antisocial economic behaviors characteristic of a "toxic masculinity" mindset, given the overlap between the male gender role and both the leader and entrepreneur roles. The second premise of our work was that displaying positive leadership behaviors (regardless of one's hierarchical position) might be a better alternative to reduce psychological tension than embracing a stereotypical role in-extremis. The results of Phase 2 show that for entrepreneurs and males, high scores in authentic leadership reduced the effects of role stereotypes on antisocial behaviors.

Digging deeper into our findings, our model predicted that the more strongly than participants embraced the agentic behaviors of the male role stereotype (H1a, H1b) or the hyper-competitive prescriptions of the entrepreneur role stereotype (H2a, H2b), said participants would be more likely to prefer (and display) antisocial behaviors, such as cheating or harming others' firms to get ahead. We tested such a claim using two realistic economic games (The "Joy-of-Destruction" and the "cheating" game) that would evidence said toxic masculinity mindset.

The results of Phase 1 provide mixed support for our predictions. In line with RCT, women are less likely to display entrepreneurial intentions and engage in antisocial economic behaviors to harm their competition (H1a). This finding aligns with the prior literature on female entrepreneurship (Gupta et al., 2009; Laguía et al., 2018). Similarly, as predicted by our main effects model, as the self-reported frequency of authentic leadership behaviors increases, participants' likelihood of choosing to harm others' firms decreases (H3b). Again, this result aligns with reports in the positive leadership literature, which related authentic leadership to ethical and pro-social behaviors in work contexts (Hannah et al., 2011, 2014).

Our results show two counter-intuitive findings. The first counter-intuitive finding was that the more willing participants were to start up a firm, the less likely they were to harm others' firms (H2a). This behavior deviates from the prescription for the entrepreneur role stereotype. We invoke a sample effect as an alternative explanation for this finding. In other words, declaring entrepreneurial intentions does not equate to occupying an entrepreneurial role. Thus, participants might have made decisions in our economic games based on their implicit stereotypical role expectations about how entrepreneurs should behave without experiencing the psychological tension that results from simultaneously occupying an entrepreneur and leader role.

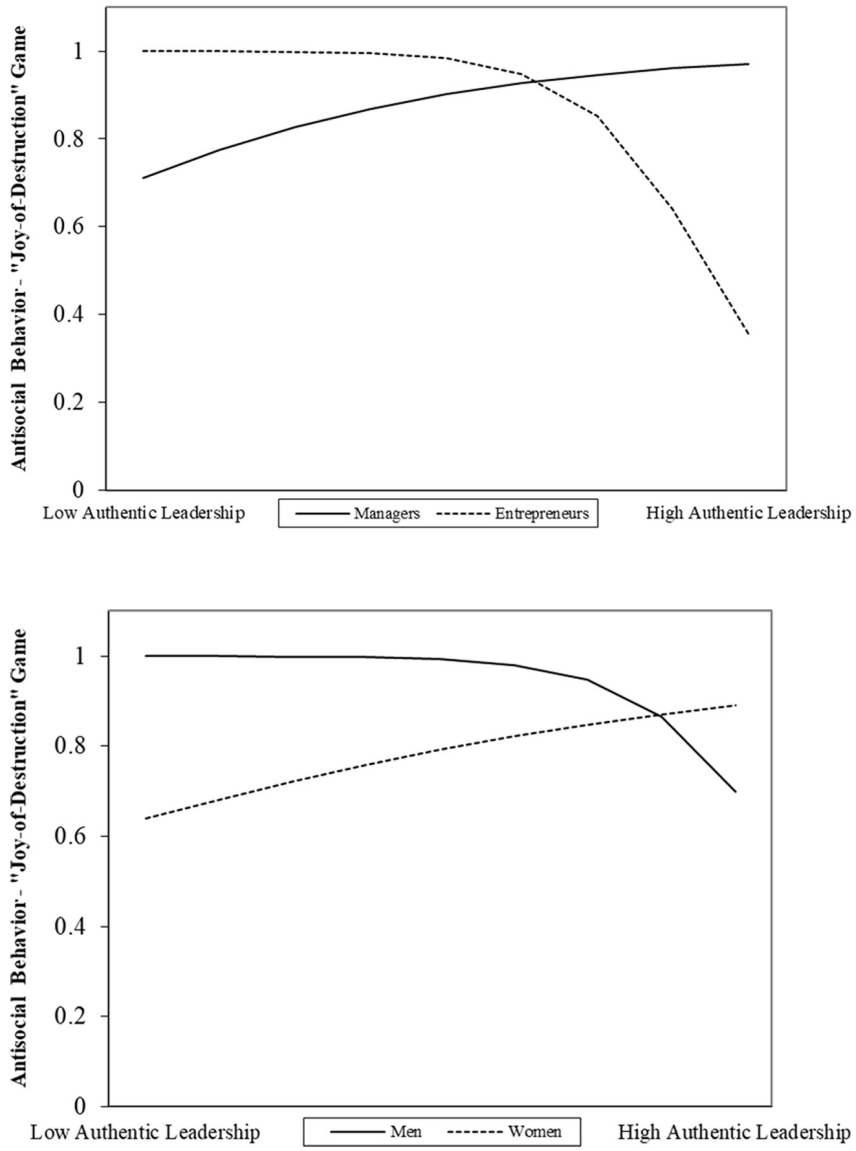

FIGURE 4 | Interactive effects of authentic leadership and entrepreneurial status (Upper) and anatomical sex (Lower) on the likelihood of an affirmative decision in the "Joy-of-Destruction" game.

The second counter-intuitive finding is that we anticipated that the more that participants would see themselves as authentic leaders, the less likely they would be to cheat, but that did not occur. A possible explanation might come from the Sendjaya et al. (2014) study, showing that as their participants' Machiavellism scores increased, the link between authentic leadership and moral action was reversed.

Again, an alternative explanation for these last findings might exist. Participants of Phase 1 comprised a heterogeneous sample of business school students and thus not "real-life" leaders. Such participants were socialized in cultures with opposing values regarding the social expectations for gender, leader, and entrepreneur role stereotypes. Thus, we decided to re-test our model in a more homogeneous sample, which ideally would comprise real entrepreneurs and managers, and conduct such a study in a western culture that embraces more traditional values than France.

Seeking to test potential mitigations for these role conflicts, in Phase 2, we adopted a gendered view of leadership. More precisely, we claimed that any given androgynous leadership style would reside at the center of the agency-communal continuum proposed by RCT. Therefore, said androgynous leadership behaviors would mitigate the toxic effect that the toxic masculinity inherent to the male gender role stereotype has on antisocial economic behaviors without triggering the double bind explained by RCT.

Following the above logic and extant research, we predicted that adopting an authentic leadership style would negatively

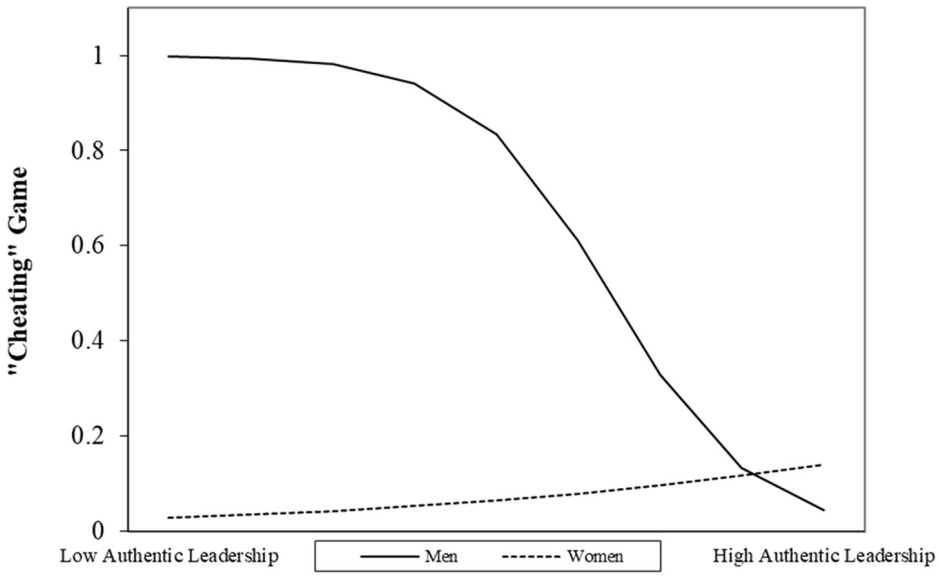

FIGURE 5 | Interactive effects of Authentic Leadership and Anatomical Sex on the likelihood of an affirmative decision in the "Cheating" Game.

relate to antisocial behaviors (H3a and H3b). Further, we predicted that adopting an authentic leadership style would enable men (H4a, H4b) and entrepreneurs (H5a, H5b) to deviate enough from the stereotypical mandates of their role stereotypes without fear of a social backlash, reducing the likelihood of observing antisocial behaviors in our games. In other words, we expected Authentic leadership to reduce the female-leader and the leader-entrepreneur role conflicts, respectively. All our hypotheses involving authentic leadership were confirmed, except one (Hypothesis 3b).

In short, our results revealed that only the effect of being anatomically male sex on the "Joy-of-Destruction" game was significant across studies. This finding evidences the negative effect of embracing stereotypical male behaviors for aspiring or actual entrepreneurs, given the inherently unethical and unsustainable nature of these practices.

Implications for Theory

Our work provides a valuable theoretical contribution to the field of leadership and the domain of female entrepreneurship. First, our work answers the call of moving beyond the study of anatomical differences in gender-focused entrepreneurship research and avoiding other forms of invisible prejudice, as purely associating female entrepreneurs with gender-congruent ventures, also unfortunately known as the "pink ghetto" (Jennings and Brush, 2013; Carter et al., 2015). Thus, the first contribution of our work is extending the RCT into the entrepreneurship arena. We contribute to RCT by proposing two additional role incongruencies, the leader-entrepreneur, and the female-entrepreneur role incongruencies.

We claim that a triple bind and prejudice derives from unpacking the male role stereotype characteristics, mainly agentic and competitive traits (Mollaret and Miraucourt, 2016). We claim that the agentic traits of the male gender role would then overlap with a leader role stereotype. Instead, the male gender competitive characteristics would overlap with the entrepreneurial role stereotype. Our model has the potential of helping male and female entrepreneurs in either traditional or social entrepreneurial roles. More precisely, we believe that our insights might help entrepreneurs resist the implicit social pressures pushing toward displaying antisocial economic behaviors.

Our theorizing is novel in claiming that the mechanism that operates against women when occupying leadership positions might also apply to entrepreneurs in general. However, this mechanism acts more strongly for female entrepreneurs. For example, in addition to being expected to be visionary and managerial at the same time (leader-entrepreneur incongruence); female entrepreneurs are also expected to be assertive and caring at these same time (gender-leader role congruence), as well as self-oriented and hyper-competitive as entrepreneurs but group-oriented and cooperative as women (gender-entrepreneur role incongruence). Such conflicting expectations can explain why women resist occupying entrepreneurial roles more accurately than focusing merely on anatomical differences.

A second theoretical contribution is that we propose an update to RCT to include the new uplifting leadership theories (Hernandez et al., 2011). Many of these uplifting theories do not fit nicely into the agency-communal continuum. We focused on authentic leadership, a new genre form of leadership that claims to be the root notion underlying positive forms of leadership for many scholars (Avolio and Gardner, 2005). Our model acknowledges and honors the RCT, at the same time proposes a more integrative gendered view of leadership, given that AL is neither fully agentic nor entirely communal (Monzani et al., 2015).

Currently, RCT focused mainly on the full-range leadership theory to describe the transactional leadership style as aligning with the agentic prescriptions of the male role stereotype and the transformational leadership style as aligning with the communal prescriptions of the female role stereotype. Thus, our work might inspire future research studies to explore how would other positive leadership styles, such as Ethical leadership (Brown et al., 2005) or Servant Leadership (Eva et al., 2019), or even Identity Leadership (Steffens et al., 2014; van Dick et al., 2018) connect with the predictions of RCT.

Finally, a third theoretical contribution of our study explains why traditional entrepreneurs tend to display antisocial economic behaviors. We focused on unfair competition or faking shareholder reports and product information as defined in microeconomic behavior studies. The importance of finding new insights on preventing the display of such antisocial economic behaviors cannot be overstated. Entrepreneurial integrity matters because whereas such antisocial economic behaviors might be functional for the short-term success of a venture, they are inherently unsustainable. So, if antisocial behaviors are institutionalized early in the life cycle of a venture, such unethical business practices will be reproduced through socialization processes as the venture matures. If unchecked, such unethical business practices will eventually erode a venture's viability (Collewaert and Fassin, 2013).

Implications for Practice

The first implication of our work is that it can inform policies aimed at fostering female entrepreneurship. We join Carter et al. (2015) call to move beyond just using anatomical sex as the sole criterion to promote female entrepreneurship. Further, we invite policy-makers to just stop simply "throwing money at women so that they can start a business" and adopt a broader perspective on gender identity. Although providing financial support to women and other minority groups is desirable and necessary, our results call for additional considerations.

Our results suggest that female entrepreneurship policies would be much more effective if said policies would incorporate provisions to reduce the gender-entrepreneur conflict and the leader-entrepreneur conflict. For example, besides providing funding and mentoring, policymakers could include provisions to build "entrepreneurial communities of practice" within their program participants to overcome the gender-entrepreneur role conflict. In such entrepreneurial communities of practice, female entrepreneurs could connect among themselves (or any who identify with the female gender role). Instead of harming their competitors, in such a safe space, female entrepreneurs could share knowledge, social support, and best practices without fear of a social backlash.

At a more meso-level, our work has implications for entrepreneurial strategizing. First, our work provides insights about how to prevent entrepreneurs from engaging in antisocial economic behaviors and indirectly how to prevent such behaviors from becoming embedded in their firms' cultures as they progress through their life cycle. In other words, our work gives a valuable first step toward the primary prevention of practices that destroyed the wealth of Theranos' shareholders.

The third practical implication is at the micro, individual level and involves the importance of positive leadership for reducing antisocial economic behavior, regardless of anatomical sex or entrepreneurial status. Creating programs to develop entrepreneurial authenticity might be useful for entrepreneurs in general and female entrepreneurs in particular. Empowering entrepreneurs to be authentic can prevent the public scandals that work against equality in entrepreneurship.

Limitations

Like any other study, our work is not without limitations. The first limitation was that we did not manipulate any of our three exogenous variables. In other words, we presented a different type of participant (male vs. female; entrepreneurs vs. leaders) with the same economic scenario, so we cannot claim to have conducted an experimental study but a laboratory study instead. Future studies should attempt to replicate our findings by comparing participants' behaviors against a more "hostile" economic context; for example, by adding a treatment condition that enhances or hinders the importance of individual contributions (e.g., punishment condition for antisocial behaviors).

The second limitation of our study is that we acknowledge two caveats regarding our samples. Strictly speaking, we did not have balanced samples in the laboratory studies comprising phase 1 and phase 2. However, future studies should attempt to replicate our work employing larger sample sizes balanced across conditions (Anatomical Sex, Entrepreneurial status, and so forth). However, we tried to attenuate this limitation using a robust estimator in our SEM model (WLSMV). Similarly, we employed additional goodness-of-fit indicators in our logistic regression models to ensure they were trustworthy.

The third limitation of our study is that we only focused on one developed country, namely France, and one emerging country, Costa Rica. Whereas Costa Rica could be seen as a paradigmatic case for Latin America, future studies should attempt to replicate our findings in a broader array of cultures, which might adopt and reward different cultural values.

Finally, our study only focused on one aspect of positive leadership, mainly authenticity. A more comprehensive study on what determines a positive entrepreneurial ethos besides authenticity would be essential (Hannah and Avolio, 2011; Crossan et al., 2017). A deeper understanding of what makes an entrepreneurial ethos might enlighten how developing entrepreneurial character can support entrepreneurs to display "ethics beyond expectations."

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Instituto Tecnológico de Costa Rica, LESSAC,

Burgundy School of Business. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

LM and GM conceived the presented idea and developed the experimental design. JMV organized the laboratory sessions in Costa Rica. GM organized the laboratory sessions in France and ran the sessions. LM and AH wrote the theory section. LM coordinated the statistical analysis with GM. All authors discussed the results and contributed to the final manuscript.

FUNDING

GM acknowledges the financial support from Burgundy School of Business, Coseil Régional de Bourgogne, supporting PARI10 and CIADEG-TEC (Costa Rica).

REFERENCES

Abbink, K., and Herrmann, B. (2011). The Moral Costs of Nastiness. *Econom. Inq.* 49, 631–633. doi: 10.1111/j.1465-7295.2010.00309.x

Abbink, K., and Sadrieh, A. (2009). The pleasure of being nasty. *Econom. Lett.* 105, 306–308. doi: 10.1016/j.econlet.2009.08.024

Abele, A. E., Uchronski, M., Suitner, C., and Wojciszke, B. (2008). Towards an operationalization of the fundamental dimensions of agency and communion: Trait content ratings in five countries considering valence and frequency of word occurrence. *Eur. J. Soc. Psychol.* 38, 1202–1217. doi: 10.1002/ejsp.575

Avolio, B. J., and Gardner, W. L. (2005). Authentic leadership development: Getting to the root of positive forms of leadership. *Leadership Q.* 16, 315–338. doi: 10.1016/j.leaqua.2005.03.001

Avolio, B. J., Wernsing, T., and Gardner, W. L. (2018). Revisiting the Development and Validation of the Authentic Leadership Questionnaire: Analytical Clarifications. *J. Manag.* 44, 399–411. doi: 10.1177/0149206317739960

Banks, G. C., McCauley, K. D., Gardner, W. L., and Guler, C. E. (2016). A meta-analytic review of authentic and transformational leadership: A test for redundancy. *The Leadership Quarterly* 27, 634–652. doi: 10.1016/j.leaqua.2016.02.006

Bass, B. M. (1985). *Leadership and performance beyond expectations*. New York, NY: Free press.

Borkowski, S. C., and Ugras, Y. J. (1998). Business Students and Ethics: A Meta-Analysis. *J. Business Ethics* 17, 1117–1127. doi: 10.1023/A:1005748725174

Brislin, R. (1980). "Translation and content analysis of oral and written material," in *Handbook of Cross-Cultural Psychology*, eds H. C. Triandis and J. W. Berry (Boston: Allyn and Bacon), 389–444.

Brown, M. E., Treviño, L. K., and Harrison, D. A. (2005). Ethical leadership: A social learning perspective for construct development and testing. *Org. Behav. Hum. Decision Proc* 97, 117–134. doi: 10.1016/j.obhdp.2005.03.002

Caprara, G. V., Barbaranelli, C., Borgogni, L., and Perugini, M. (1993). The "big five questionnaire": A new questionnaire to assess the five factor model. *Personal. Individ. Diff.* 15, 281–288. doi: 10.1016/0191-8869(93)90218-R

Carter, S., Mwaura, S., Ram, M., Trehan, K., and Jones, T. (2015). Barriers to ethnic minority and women's enterprise: Existing evidence, policy tensions and unsettled questions. *Internat. Small Bus. J.* 33, 49–69. doi: 10.1177/0266242614556823

Chaudhuri, A. (2011). Sustaining cooperation in laboratory public goods experiments: a selective survey of the literature. *Exp. Econ.* 14, 47–83. doi: 10.1007/s10683-010-9257-1

Childs, J. (2012). Gender differences in lying. *Econ. Lett.* 114, 147–149. doi: 10.1016/j.econlet.2011.10.006

Collewaert, V., and Fassin, Y. (2013). Conflicts between entrepreneurs and investors: the impact of perceived unethical behavior. *Small Business Econ.* 40, 635–649. doi: 10.1007/s11187-011-9379-7

Cook, A., and Glass, C. (2014). Above the glass ceiling: When are women and racial/ethnic minorities promoted to CEO? *Strat. Manag. J.* 35, 1080–1089. doi: 10.1002/smj.2161

Cook, K. (2020). *The Psychology of Silicon Valley: Ethical Threats and Emotional Unintelligence in the Tech Industry*. Cham: Springer, doi: 10.1007/978-3-030-27364-4

Crossan, M. M., Byrne, A., Seijts, G. H., Reno, M., Monzani, L., and Gandz, J. (2017). Toward a Framework of Leader Character in Organizations. *J. Manag. Stud.* 54, 986–1018. doi: 10.1111/joms.12254

Crossan, M. M., Mazutis, D., and Seijts, G. H. (2013). In search of virtue: The role of virtues, values and character strengths in ethical decision making. *J. Bus. Ethics* 113, 567–581. doi: 10.1007/s10551-013-1680-8

Das, M. (2000). Women Entrepreneurs from India: Problems, Motivations and Success Factors. *J. Small Business Entrepren.* 15, 67–81. doi: 10.1080/08276331.2000.10593294

Datta, P. B., and Gailey, R. (2012). Empowering Women Through Social Entrepreneurship: Case Study of a Women's Cooperative in India. *Entrepren. Theory Pract.* 36, 569–587. doi: 10.1111/j.1540-6520.2012.00505.x

Dawson, J. F. (2013). Moderation in Management Research: What, Why, When, and How. *J. Bus. Psychol.* 29, 1–19. doi: 10.1007/s10869-013-9308-7

Eagly, A. H., and Karau, S. J. (2002). Role congruity theory of prejudice toward female leaders. *Psychol. Rev.* 109, 573–598. doi: 10.1037//0033-295X.109.3.573

Eva, N., Robin, M., Sendjaya, S., van Dierendonck, D., and Liden, R. C. (2019). Servant Leadership: A systematic review and call for future research. *Leadership Q.* 30, 111–132. doi: 10.1016/j.leaqua.2018.07.004

Ezquerra, L., Kolev, G. I., and Rodriguez-Lara, I. (2018). Gender differences in cheating: Loss vs. gain framing. *Econ. Lett.* 163, 46–49. doi: 10.1016/j.econlet.2017.11.016

Filippin, A., and Crosetto, P. (2016). A Reconsideration of Gender Differences in Risk Attitudes. *Manag. Sci.* 62, 3138–3160. doi: 10.1287/mnsc.2015.2294

Fischbacher, U., and Föllmi-Heusi, F. (2013). Lies in disguise-an experimental study on cheating. *J. Eur. Econ. Assoc.* 11, 525–547. doi: 10.1111/jeea.12014

Frese, M., and Gielnik, M. M. (2014). The Psychology of Entrepreneurship. *Ann. Rev. Org. Psychol. Org. Behav.* 1, 413–438. doi: 10.1146/annurev-orgpsych-031413-091326

Friesen, L., and Gangadharan, L. (2012). Individual level evidence of dishonesty and the gender effect. *Econ. Lett.* 117, 624–626. doi: 10.1016/j.econlet.2012.08.005

Gandz, J., Crossan, M. M., Seijts, G. H., and Stephenson, C. (2010). *Leadership on Trial: A manifesto for leadership development*. London: Ivey Business School.

Geiser, C. (2011). *Datenanalyse mit Mplus. Eine anwendungsorientierte Einführung* (2nd ed.). Verlag für Sozialwissenschaften. Berlin: springer.

Gloor, J. L., Morf, M., Paustian-Underdahl, S., and Backes-Gellner, U. (2018). Fix the Game, Not the Dame: Restoring Equity in Leadership Evaluations. *J. Business Ethics* 2018:3861. doi: 10.1007/s10551-018-3861-y

Gupta, V. K., Turban, D. B., Wasti, S. A., and Sikdar, A. (2009). The Role of Gender Stereotypes in Perceptions of Entrepreneurs and Intentions to Become an Entrepreneur. *Entrepren. Theory Pract.* 33, 397–417. doi: 10.1111/j.1540-6520.2009.00296.x

Hannah, S. T., and Avolio, B. J. (2011). Leader character, ethos, and virtue: Individual and collective considerations. *Leadership Q.* 22, 989–994. doi: 10.1016/j.leaqua.2011.07.018

Hannah, S. T., Avolio, B. J., and Walumbwa, F. O. (2011). Relationships between Authentic Leadership, Moral Courage, and Ethical and Pro-Social Behaviors. *Business Ethics Q.* 21, 555–578. doi: 10.5840/beq201121436

Hannah, S. T., Avolio, B. J., and Walumbwa, F. O. (2014). Addendum to "Relationships between Authentic Leadership, Moral Courage, and Ethical and Pro-Social Behaviors. *Business Ethics Q.* 24, 277–279. doi: 10.5840/beq201453011

Harrington, D. C. (2020). What is "Toxic Masculinity" and Why Does it Matter? *Men Mascul.* 2020:1097184X2094325. doi: 10.1177/1097184X20943254

Hartley, R. E. (1959). Sex-role pressures and the socialization of the male child. *Psychol. Rep.* 5, 457–468.

Hentschel, T., Heilman, M. E., and Peus, C. V. (2019). The Multiple Dimensions of Gender Stereotypes: A Current Look at Men's and Women's Characterizations of Others and Themselves. *Front. Psychol.* 10:11. doi: 10.3389/fpsyg.2019.00011

Hernandez, M., Eberly, M. B., Avolio, B. J., and Johnson, M. D. (2011). The loci and mechanisms of leadership: Exploring a more comprehensive view of leadership theory. *Leadership Q.* 22, 1165–1185. doi: 10.1016/j.leaqua.2011.09.009

Hernandez Bark, A. S., Escartín, J., Schuh, S. C., and van Dick, R. (2015). Who Leads More and Why? A Mediation Model from Gender to Leadership Role Occupancy. *J. Business Ethics* 2015:2642. doi: 10.1007/s10551-015-2642-0

Hernandez Bark, A. S., Escartín, J., and van Dick, R. (2014). Gender and Leadership in Spain: a Systematic Review of Some Key Aspects. *Sex Roles* 70, 522–537. doi: 10.1007/s11199-014-0375-7

Hernandez-Arenaz, I. (2020). Stereotypes and tournament self-selection: A theoretical and experimental approach. *Eur. Econ. Rev.* 126:103448. doi: 10.1016/j.euroecorev.2020.103448

Hu, L., and Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Struct. Equ. Model.* 6, 1–55. doi: 10.1080/10705519909540118

Hughes, K. D., Jennings, J. E., Brush, C., Carter, S., and Welter, F. (2012). Extending Women's Entrepreneurship Research in New Directions. *Entrepren. Theory Pract.* 36, 429–442. doi: 10.1111/j.1540-6520.2012.00504.x

Hyde, J. S., Fennema, E., and Lamon, S. J. (1990). Gender differences in mathematics performance: A meta-analysis. *Psycholog. Bull.* 107, 139–155. doi: 10.1037/0033-2909.107.2.139

Jennings, J. E., and Brush, C. G. (2013). Research on Women Entrepreneurs: Challenges to (and from) the Broader Entrepreneurship Literature? *Acad. Manag. Annals* 7, 663–715. doi: 10.1080/19416520.2013.782190

Jensen, S. M., and Luthans, F. (2006b). Relationship between Entrepreneurs' Psychological Capital and Their Authentic Leadership. *J. Manag. Iss.* 18, 254–273.

Jensen, S. M., and Luthans, F. (2006a). Entrepreneurs as authentic leaders: impact on employees' attitudes. *Leadership Org. Dev. J.* 27, 646–666. doi: 10.1108/01437730610709273

Kanfer, R., and Kantrowitz, T. M. (2005). "Ability and Non-Ability Predictors of Job Performance," in *Psychological Management of Individual Performance*, ed. S. Sonnentag (Hoboken,NJ: John Wiley & Sons), 27–50. doi: 10.1002/0470013419.ch2

Keser, C. (2002). "Cooperation in Public Goods Experiments," in *Surveys in Experimental Economics SE - 5*, eds F. Bolle and M. Lehmann-Waffenschmidt (Heidelberg: Physica-Verlag HD), 71–90. doi: 10.1007/978-3-642-57458-0_5

Knapp, J. R., Smith, B. R., Kreiner, G. E., Sundaramurthy, C., and Barton, S. L. (2013). Managing Boundaries Through Identity Work. *Family Bus. Rev.* 26, 333–355. doi: 10.1177/0894486512474036

Koenig, A. M., Eagly, A. H., Mitchell, A. A., and Ristikari, T. (2011). Are leader stereotypes masculine? A meta-analysis of three research paradigms. *Psychol. Bull.* 137, 616–642. doi: 10.1037/a0023557

Laguía, A., García-Ael, C., Wach, D., and Moriano, J. A. (2018). Think entrepreneur - think male": a task and relationship scale to measure gender stereotypes in entrepreneurship. *Internat. Entrepren. Manag. J.* 2018:553. doi: 10.1007/s11365-018-0553-0

Latrofa, M., Vaes, J., Cadinu, M., and Carnaghi, A. (2010). The Cognitive Representation of Self-Stereotyping. *Personal. Soc. Psychol. Bull.* 36, 911–922. doi: 10.1177/0146167210373907

Lewis, P. (2013). The Search for an Authentic Entrepreneurial Identity: Difference and Professionalism among Women Business Owners. *Gender Work Org.* 20, 252–266. doi: 10.1111/j.1468-0432.2011.00568.x

Liñán, F., and Chen, Y. (2009). Development and Cross-Cultural Application of a Specific Instrument to Measure Entrepreneurial Intentions. *Entrepren. Theory Pract.* 33, 593–617. doi: 10.1111/j.1540-6520.2009.00318.x

Little, T. D., Cunningham, W. A., Shahar, G., Widaman, K. F., Little, T. D., Cunningham, W. A., et al. (2009). To Parcel or Not to Parcel: Exploring the Question, Weighing the Merits To Parcel or. *Struct. Equ. Model.* 55:902. doi: 10.1207/S15328007SEM0902

March, E., van Dick, R., and Hernandez Bark, A. (2016). Current prescriptions of men and women in differing occupational gender roles. *J. Gend. Stud.* 25, 681–692. doi: 10.1080/09589236.2015.1090303

Mollaret, P., and Miraucourt, D. (2016). Is job performance independent from career success? A conceptual distinction between competence and agency. *Scand. J. Psychol.* 57, 607–617. doi: 10.1111/sjop.12329

Monzani, L., Braun, S., and van Dick, R. (2016). It takes two to tango: The interactive effect of authentic leadership and organizational identification on employee silence intentions. *Germ. J. Hum. Res. Manag.* 30, 246–266.

Monzani, L., Hernandez Bark, A. S., van Dick, R., and Peiró, J. M. (2015). The Synergistic Effect of Prototypicality and Authenticity in the Relation Between Leaders' Biological Gender and Their Organizational Identification. *J. Business Ethics* 132, 737–752. doi: 10.1007/s10551-014-2335-0

Monzani, L., Knoll, M., Giessner, S., van Dick, R., and Peiró, J. M. (2019). Between a Rock and Hard Place: Combined Effects of Authentic Leadership, Organizational Identification, and Team Prototypicality on Managerial Prohibitive Voice. *Spanish J. Psychol.* 22:E2. doi: 10.1017/sjp.2019.1

Monzani, L., Seijts, G. H., and Crossan, M. M. (2021b). Character matters: The network structure of leader character and its relation to follower positive outcomes. *PLoS One* 16:e0255940. doi: 10.1371/journal.pone.0255940

Monzani, L., Escartín, J., Ceja, L., and Bakker, A. B. (2021a). Blending Mindfulness Practices and Character Strengths Increases Employee Wellbeing: A second-order meta-analysis and a follow-up Field Experiment. *Hum. Resour. Manag. J.* 2021:12360. doi: 10.1111/1748-8583.12360

Monzani, L., and Van Dick, R. (2020). "Positive Leadership in Organizations," in *Oxford Research Encyclopedia of Psychology*, ed. J. M. Peiró (Oxford: Oxford University Press), 1–37. doi: 10.1093/acrefore/9780190236557.013.814

Moriano, J. A., Molero, F., and Lévy Mangin, J.-P. (2011). Liderazgo auténtico. Concepto y validación del cuestionario ALQ en España. *Psicothema* 23, 336–341.

Moshagen, M., and Musch, J. (2014). Sample Size Requirements of the Robust Weighted Least Squares Estimator. *Methodology* 10, 60–70. doi: 10.1027/1614-2241/a000068

Ordóñez, L. D., Schweitzer, M. E., Galinsky, A. D. A. D., and Bazerman, M. H. (2009). Goals gone wild: The systematic side effects of overprescribing goal setting. *Acad. Manag. Perspect.* 23, 6–16.

Rauch, A., and Frese, M. (2007). Let's put the person back into entrepreneurship research: A meta-analysis on the relationship between business owners' personality traits, business creation, and success. *Eur. J. Work Org. Psychol.* 16, 353–385. doi: 10.1080/13594320701595438

Rowe, W. G. (2001). Creating wealth in organizations: The role of strategic leadership. *Acad. Manag. Exec.* 15, 81–94. doi: 10.5465/AME.2001.4251395

Ryan, M. K., and Haslam, S. A. (2007). The Glass Cliff: Exploring the Dynamics Surrounding the Appointment of Women to Precarious Leadership Positions. *Acad. Manag. Rev.* 32, 549–572. doi: 10.5465/amr.2007.24351856

Schmidt, F. L., and Hunter, J. E. (2003). "The Blackwell Hanbook of Principles of Organizational Behaviour," in *The Blackwell Hanbook of Principles of Organizational Behaviour*, ed. E. A. Locke (Hoboken, NJ: Blackwell Publishing Ltd), 1–14. doi: 10.1111/b.9780631215066.2003.00002.x

Schultz, P. W., and Zelenzy, L. (1999). Values As Predictors of Environmental Attitudes: Evidence For Consistency Across 14 Countries. *J. Env. Psychol.* 19, 255–265. doi: 10.1006/jevp.1999.0129

Seijts, G. H., Monzani, L., Woodley, H. J. R., and Mohan, G. (2021). The Effects of Character on the Perceived Stressfulness of Life Events and Subjective Well-Being of Undergraduate Business Students. *J. Manag. Educ.* 2021:105256292098010. doi: 10.1177/1052562920980108

Sendjaya, S., Pekerti, A., Härtel, C. E. J., Hirst, G., and Butarbutar, I. (2014). Are authentic leaders always moral? The role of machiavellianism in the relationship between authentic leadership and morality. *J. Bus. Ethics* https://doi.org/10.1007/s10551-014-2351-0

Smith, J. B., Mitchell, J. R., and Mitchell, R. K. (2009). Entrepreneurial Scripts and the New Transaction Commitment Mindset: Extending the Expert Information Processing Theory Approach to Entrepreneurial Cognition Research. *Entrepren. Theory Pract.* 33, 815–844. doi: 10.1111/j.1540-6520.2009.00328.x

Steffens, N. K., Haslam, S. A., Reicher, S. D., Platow, M. J., Fransen, K., Yang, J., et al. (2014). Leadership as social identity management: Introducing the Identity Leadership Inventory (ILI) to assess and validate a four-dimensional model. *Leadership Q.* 25, 1001–1024. doi: 10.1016/j.leaqua.2014.05.002

Stewart, W. H., and Roth, P. L. (2007). A Meta-Analysis of Achievement Motivation Differences between Entrepreneurs and Managers. *J. Small Busin. Manag.* 45, 401–421. doi: 10.1111/j.1540-627X.2007.00220.x

Taneja, H. (2019). The Era of "Move Fast and Break Things" Is Over. *Harvard Business Rev.* 2019:01.

van Dick, R., Lemoine, J. E., Steffens, N. K., Kerschreiter, R., Akfirat, S. A., Avanzi, L., et al. (2018). Identity leadership going global: Validation of the Identity Leadership Inventory across 20 countries. *J. Occup. Org. Psychol.* 91, 697–728. doi: 10.1111/joop.12223

Walumbwa, F. O., Avolio, B. J., Gardner, W. L., Wernsing, T. S., and Peterson, S. J. (2008). Authentic Leadership: Development and Validation of a Theory-Based Measure. *J. Manag.* 34, 89–126. doi: 10.1177/0149206307308913

Whitley, B. E., Nelson, A. B., and Jones, C. J. (1999). Gender Differences in Cheating Attitudes and Classroom Cheating Behavior: A Meta-Analysis. *Sex Roles* 41, 657–680. doi: 10.1023/A:1018863909149

Zehnder, C., Herz, H., and Bonardi, J.-P. (2017). A productive clash of cultures: Injecting economics into leadership research. *Leadership Q.* 28, 65–85. doi: 10.1016/j.leaqua.2016.10.004

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychology

ORIGINAL RESEARCH
published: 02 December 2021
doi: 10.3389/fpsyg.2021.745948

Effects of Inequality on Trust and Reciprocity: An Experiment With Real Effort

*Amalia Rodrigo-González[1], María Caballer-Tarazona[2] and Aurora García-Gallego[3,4]**

[1] Department of Corporate Finance, Universitat de València, Valencia, Spain, [2] Department of Applied Economics, Universitat de València, Valencia, Spain, [3] Laboratorio de Economía Experimental (LEE), Department of Economics, Universitat Jaume I, Castelló de la Plana, Spain, [4] Instituto Complutense de Análisis Económico, Universidad Complutense de Madrid, Madrid, Spain

The purpose of this paper is analyzing whether trust and reciprocity are affected by how rich the partner is or how well the partner performed several tasks with real effort. A trust game (TG) experiment is designed with three treatments. First, a baseline Treatment B in which subjects play a finitely repeated TG. Second, in a Treatment H with history, subjects know the partner's wealth level reached in the past. Third, in a Treatment E with effort the individual endowment with which the TG is played is endogenous and results from the subject's performance in three different real effort tasks (maths, cognitive and general knowledge related). The data analysis highlights the importance of past wealth levels (Treatment H) as well as endowment heterogeneity (Treatment E), on the actual levels of trust and reciprocity. Specifically, it is observed that the decision of trustors is positively affected by positive past experienced reciprocity. Moreover, trustors are sensitive to how much money the trustee accumulates each round in Treatment H, trusting more the ones that have accumulated less compared to themselves. In contrast with that, it is remarkable in Treatment E that trustors are sensitive to the endowment level of the trustees, trusting more the partners that have got a higher than own endowment, probably considering that a person that performed better in the tasks is a better partner to trust. As far as second players' behavior, as the amount received from the trustor increases it is less likely that the trustee reciprocates with higher than or with the egalitarian amount. In Treatments H and E, the probability that the trustee reciprocates with higher amount that the one received increases when inequality in endowment/accumulated earnings favors the trustor. Additional results come from analysis of personality archetypes and socio-demographic variables.

Keywords: inequality, trust, reciprocity, altruism, real-effort task, experiment

OPEN ACCESS

Edited by:
Ismael Rodriguez-Lara, University of Granada, Spain

Reviewed by:
Giuseppe Attanasi, Université Côte d'Azur, France
Hernan Daniel Bejarano, Centro de Investigación y Docencia Económicas, Mexico

***Correspondence:**
Aurora García-Gallego
mgarcia@uji.es

Specialty section:
This article was submitted to Personality and Social Psychology, a section of the journal Frontiers in Psychology

Received: *23 July 2021*
Accepted: *05 November 2021*
Published: *02 December 2021*

Citation:
Rodrigo-González A, Caballer-Tarazona M and García-Gallego A (2021) Effects of Inequality on Trust and Reciprocity: An Experiment With Real Effort. Front. Psychol. 12:745948. doi: 10.3389/fpsyg.2021.745948

INTRODUCTION

The study of human behavior in terms of trust and reciprocity is crucial for understanding the social capital creation that allows achieving goals commonly shared by societies. Experimental and behavioral economics have understood the importance of this issue and have given us a huge spectrum of results in which trust and reciprocity are the focus of the question. Specifically, numerous references analyze the dynamics of trust and reciprocity under different set-ups focusing on the effect of income inequality.

The motivation behind the role played by economic inequality in human behavior is intuitively relevant. The amount of money owned by people is naturally heterogenous, especially because it may have been originated differently and such differences seem to matter a lot. The truth is that human beings care a lot about economic heterogeneity among their peers. In particular, it seems reasonable to hypothesize that humans naturally value more the income created from their own work than the one coming from other non-work-related sources like, for example, inheritance or subsidies. In other words, people care about whether the money comes from own effort or just comes as manna from heaven, and this may affect the willingness to invest and the way of investing the money. And this effect may be stronger in the case that the investment has uncertain returns, especially returns that depend on others' decisions. This source of economic heterogeneity is considered endogenous.

A different dimension of economic heterogeneity is the one created as a result of being aware of how different my accumulated earnings are along life with respect to my peers. This source of information may be so relevant for economic behavior of people as, for example, to affect the levels of generosity, altruism or even the levels of trust on others in specific environments. In fact, social preferences are relevant in individual decision making, since they are formed by personality factors as well as by social norms. For sure one should differentiate between decisions taken in a situation where all subjects have similar wealth from the situation in which wealth differences exist. Being aware of wealth differences with my peers may wake up fraternity feelings on me and the willingness to equilibrate the imbalance by being active in giving money; or just the opposite may happen, feeling that I deserve more than the others and to make decisions that make our differences even higher. No trivial combinations and results can be found under economic inequalities, and this is the focus of our interest in this research.

The Trust Game (TG) represents a situation that is appropriate to experimentally analyze the effect of facing such economic inequality on subjects' decisions related to trust. To the best of our knowledge, our work is the first in designing a situation in which trust, reciprocity and altruism are analyzed taking into consideration those sources of economic heterogeneity: the own effort endogenous income inequality and the unequal accumulated earnings. Our design extends the TG experimental literature but tries to cover an empty space in which trust, reciprocity and altruism are analyzed under the influence of heterogeneous initial endowment generated through subjects' performance in real-effort tasks. The design also considers another source of heterogeneity, the one created by differences in accumulated earnings, which has somehow already been considered in previous literature.

Our purpose is analyzing whether trust and reciprocity are affected by how rich my partner is compared with me or how well the partner performed several tasks with real effort in contrast with my own score. A TG experiment is designed with three treatments. First, a baseline Treatment B in which subjects play a finitely repeated TG (Berg et al., 1995). Second, in a Treatment H with history, subjects know the partner's wealth level reached in the past. Third, in Treatment E with effort, the individual endowment with which the TG is played is endogenous and results from the subjects' performance in three different real effort tasks (maths, cognitive, and general knowledge related). Furthermore, our TG version allows for the trustee decision to disentangle reciprocity from altruism, since the decision is double: first, how much of the amount received to return to the trustor and, second, what part of the endowment to give to the trustor.

The data analysis highlights the importance of heterogeneity in earnings levels (Treatment H) as well as in initial endowment (Treatment E) in the last two periods on the actual levels of trust and reciprocity. Specifically, it is observed that the decision of trustors is positively affected by positive past experienced reciprocity. Moreover, trustors are sensitive to how much money the trustee accumulates each round in Treatment H, trusting less the ones that have more compared to themselves. In contrast, it is remarkable the fact that in Treatment E trustors are sensitive to the endowment level of the trustees, trusting more the partners that have got a higher than own endowment, probably considering that a person that performed better in the tasks is a better partner to trust.

As far as the trustee is concerned, his role aims at reducing the wealth gap existing between the two players. Specifically, we take the egalitarian strategy as a reference, meaning that the trustee sends back to the trustor an amount such that his earnings equalize those of the trustor. Three reciprocity levels are taken into consideration: first, second and third levels of reciprocity stand for sending back to the trustor, respectively, a lower, equal and higher amount. In this sense, data reveal that it is more likely that the trustee reciprocates with higher or equal to the egalitarian amount as the trustor decreases the amount sent in the first place. In Treatment H/Treatment E the probability that the trustee reciprocates with higher/equal than/to the egalitarian amount increases when inequality in accumulated earnings/endowment favors the trustor. Previous results provide several tentative explanations to this behavior. For instance, Attanasi et al. (2019) justify the increase in reciprocity in the face of low trust levels as an incentive to raise trust levels in the future. Furthermore, Khalmetski et al. (2015) and Balafoutas and Fornwagner (2017) use a dictator game and found that guilt aversion plays a role in second movers' decision and, because of that, correlation between transfer and expectations can be negative.

In our design the decision of the trustee is double, so that we can measure not only the reciprocity level, but we also measure the level of altruism when the trustee decide how much of his own endowment to send to the trustor. Results show that the probability of being altruistic for a trustee is independent of the reciprocity decision but it depends positively on the trustor decision as well on his advantage (disadvantage) in endowment (cumulated earnings) with respect to those of the trustor.

The Equality Equivalence Test (EET) has been used in order to classify subjects by personality archetypes. It is worth mentioning that trustors classified as inequality-lovers present significant differences with respect to those classified as altruists. In general, trustees classified as altruists in the EET are trustees that more likely will choose to reciprocate with the egalitarian strategy.

The remainder of the paper is organized as follows. Section "Related Literature" reviews the related literature on trust experiments. Section "Materials and Methods" lists our main research questions and also gives a detailed description of the experimental design. The results are presented in section "Results." Section "Econometric Analysis" concludes.

RELATED LITERATURE

The seminal work by Berg et al. (1995), has been widely used in experimental economics to study trust and reciprocity behaviors. Many authors made some variations on the Berg's TG in order to stand out other factors involved in cooperative behavior. For instance, it has been found that factors such as experimental protocols and geographical variations or gender, among others, have an effect on trust levels. For example, Johnson and Mislin (2011) find that minor variations in the design protocol (i.e., payment criteria, rate of return or population characteristics) can imply significant changes in share behavior. Their findings suggest that subjects trust less if they are paid randomly and if they play with a simulated counterpart instead of a human. Moreover, trustworthiness decreases when the rate of return is 2 (instead of 3) and when the experiment was run with students. In the same line, in Bornhorst et al. (2010) participants choose their partner to play a TG with some information about each other's age, gender, nationality and number of siblings. At the beginning of the sessions, authors find differences among participants' decisions from northern and southern countries in terms of share amounts and type of partner chosen. However, over the course of the game, those cultural differences become blurred. This research evidences that, in spite of the different individual characteristics, trust breeds trust and allows to identify where to find trustworthiness.

One of the aspects which has recently attracted the researchers' attention is the effect of heterogeneity on trust and reciprocity behaviors. On one hand, the non-experimental literature has long since coincided with the negative effect of individual characteristics heterogeneity on trust levels and cooperative behaviors (Putnam, 2000). On the other hand, recent experimental literature has focused on how wealth heterogeneity implies variations on levels of trust and reciprocity.

Most of the experimental studies agree on the idea that wealth inequality and generalized trust correlate negatively (Gallego, 2016), although aspects such as the availability of information or the direction of inequality regarding the other players should be considered (Andreoni et al., 2017; Xiao and Bicchieri, 2010; Bejarano et al., 2018).

For instance, Lei and Vesely (2010) explore how the inequality in the endowment activates favoritism among the members of the same group, through the TG and the dictator game. They find that this favoritism within the group remains even if there is no longer inequality, concluding that favoritism is activated within members of the same group only in cases in which trustors are classified as rich. However, this favoritism effect decreases but does not disappear when playing under an equitable endowment. Effects of group membership on trust were explored also by Smith (2011a). He found that information about the identity of the other player had positive in-group and negative out-group effects on trust. However, the in-group effect was small and statistically insignificant, while the out-group effect was larger and statistically significant.

The role of an unequal endowment was also explored by Smith (2011b) and Brülhart and Usunier (2012). These authors produce lab-induced players with high and low endowment, and observe which are the behavioral dynamics in the four combinations or profiles of couples. While Brülhart and Usunier (2012) did not find a different behavior when individuals play with rich players or poor players, Smith (2011b) found that subjects with low endowment paired with high endowment subjects showed more trust than subjects in other pairs; in addition, their trust was reciprocated with higher trustworthiness. In the same line, Ciriolo (2007) finds that an unequal distribution of show-up fees may eventually reduce the incentive to cooperate of both players.

Other authors have focused their research on the trustee's behavior. For example, Xiao and Bicchieri (2010) study inequality aversion when the trustee has lower endowment than the trustor. In this case, it is observed that the trustee's reciprocity decreases significantly, and the authors associate this effect with inequality aversion. In this line, Rodriguez-Lara (2018) also focus on the trustee's strategy, but no evidence of inequality aversion is found. They find that in a context of heterogeneous endowment, reciprocity decreases, but not necessarily because of inequality aversion. Conversely, Bejarano et al. (2021a) create an inequality endowment through negative shocks, resulting in a situation where trustors are poorer than trustees. Within this context, the authors observe that inequality increases the levels of both, trust and reciprocity.

Regarding the effect of inequality, information availability seems to be the key point for some authors. For instance, Anderson et al. (2006) conclude that the effect of inequality on trust, in terms of both sign and significance, depends on whether the show-up payments are awarded publicly or privately. In other words, when the induced inequality of payments is awarded privately, the levels of trust decrease; however, when payments are awarded publicly, differences on trust levels are not observed. Inequality has not the same effect on all players, though. Heap et al. (2013) study the inequality effect in a non-market and in a market setting (trust and labor market games, respectively) and found that when it is common knowledge, inequality has a negative effect on trust. In addition, trust in a market setting appears generally more sensitive to the introduction of inequality than in the non-market setting. That is, the wage levels (trust) are on average lower when there is inequality.

Blanco and Dalton (2019) combine a Dictator Game lab experiment with information about the real income stratum of each participant. A positive relation between donations and wealth is shown to be due to the fact that for rich people the experimental endowment has lower real value. They find that the motivation to donate is similar across strata, where the generosity act is explained mainly by warm-glow rather than pure altruism.

Especially inspiring is the work of Greiner et al. (2012) that explore the effect on trust of endogenous as well as exogenous inequality. Authors consider as endogenous inequality

the heterogeneity generated along the decisions made in the TG during 20 rounds. The study concludes that with heterogeneous endowment, trust levels remain more stable than in a context of egalitarian endowment. The levels of trust are initially higher in a treatment with equal initial endowments, but these levels of trust decrease over time as the accumulated earnings generated in the game become more heterogeneous. In a treatment with unequal endowments, trust is initially lower than in treatment with equal endowments but the levels remain more stable in comparison with the case of equal endowments.

The design of Fehr et al. (2020) exogenously induces unjust economic inequality after performing a real-effort task, but the payment is not related with the effort nor with the performance in the task. Results show a decline in levels of trust and reciprocity on the extent to which this is deemed fair by participants.

In Bejarano et al. (2018, 2021b), authors analyze the inequality effect on trust and reciprocity both in a context of endowment heterogeneity and inequality generated by random shocks. They find that first-movers send less to second-movers only when the inequality results from a random shock. Moreover, second-movers return less when they are endowed less than the first-mover, regardless of whether the difference in endowments was initially given or occurred after a random shock.

With the exception of Greiner et al. (2012), most of the previous studies referred are devoted to studying the effect of exogenous heterogeneity on trust or reciprocity. In line with Greiner et al. (2012), the present work considers both endogenous and exogenous wealth heterogeneity in the analysis of their effects on trust, reciprocity and altruism levels. Furthermore, our analysis endogenously creates income heterogeneity, generated by a set of real-effort tasks carried out before playing the TG. Higher earnings derived from real-effort task is commonly associated with higher effort, and this has an effect on cooperation decision. As Fehr (2018), Fehr et al. (2020) suggest, the fairness in the income-generating process matters.

Additionally, more recent experimental literature on trust focuses on categorizing individuals based on personal characteristics or motivations. Some of these works have used post-experimental questionnaires with questions to correlate psychological or cognitive characteristics with behaviors observed in the game. This is the case of Corgnet et al.'s (2016) work which combines the decisions in the TG with the results in the Cognitive Reflection Test (CRT). The authors find a positive relationship between cognitive reflection and trusting behavior. In this line, in Bellucci et al. (2019), participants carry out the RSFC test (resting-state functional connectivity) after playing the TG because they were interested in observing the relation between the decisions in TG and results in RSFC. Also, Espín et al. (2016) analyze the relation between the decisions in the TG game and individuals' social motivation. Their research applies both the Dictator Game and a dual-role Ultimatum Game to identify individuals' social preferences for altruism, spitefulness, egalitarianism, and efficiency. They find considerable heterogeneity in the TG decisions' motivation. Furthermore, in Attanasi et al. (2013, 2019) authors use pre and post-experimental questionnaires to correlate players' characteristics with their decisions in the TG. The purpose of introducing the questionnaires was to classify trustees as guilt averse or selfish. Trust increases with guilt sensitivity and the reputation effect is very strong.

Other authors who want to investigate motivations of trust, try to isolate the behavior that really indicates a trust decision. A good example of this is Chetty et al. (2020). They focus on identifying risk-trust relationships by using a risk-preference task. They conclude that attitudes to risk may partly confound the measurement of trust. In this line, Cox et al. (2016) uses different treatments of the investment game to categorize individuals according to their social preference, and then, analyzing their decisions on the TG, in order to isolate effects as vulnerability or inequality aversion from trust.

With this background in mind, the present paper analyses the effect of income inequality on trust and reciprocity. The income heterogeneity comes from two different sources. On one hand, it is endogenously generated through the TG-repeated decisions. On the other, the inequality comes from a heterogeneous endowment generated from real-effort tasks performed before playing the TG. Different experimental treatments are designed in order to isolate these effects. Finally, inspired by the work of Cox (2004) and Anderson et al. (2006), our design allows to disentangle the second-player decision in the TG between reciprocity and altruism.

MATERIALS AND METHODS

This section describes in detail the experimental design and also motivates the research questions.

Experimental Design

The experiment is divided in three treatments, all having in common that participants play the Trust Game. Therefore, before describing the details of each treatment, the version of the Trust Game implemented in our treatments is exposed.

Our Trust Game

Following the version of Berg et al. (1995), a trustor (sender) and a trustee (receiver) are endowed with the same amount of money E. The trustor decides which part (in absolute value) $x \in (0, E)$ of the endowment to send to an anonymous trustee. The amount x is then multiplied by $n = 3$ in the trustee's hands. After the trustor's decision is observed, the trustee decides about two (absolute) amounts to return to the trustor[1]:

1. Amount $y_1 \in (0, 3x)$ to return to the trustor.
2. Amount $y_2 \in (0, \mathrm{E})$ to send to the trustee.

Consequently, the final payoff for the trustor is $\pi_{or} = E - x + y_1 + y_2$, and that of the trustee equals $\pi_{ee} = 3x - y_1 + E - y_2$. **Figure 1** shows the extensive form of the TG version just described.

This game has a unique subgame perfect Nash equilibrium in (*"no trust," "no reciprocate"*) and therefore, neither trust

[1]The trustee observes two boxes on the screen, both preceded by the corresponding question: (1) How much of the amount received do you want to return to the trustor? (2) How much from your endowment do you want to send to the trustor?

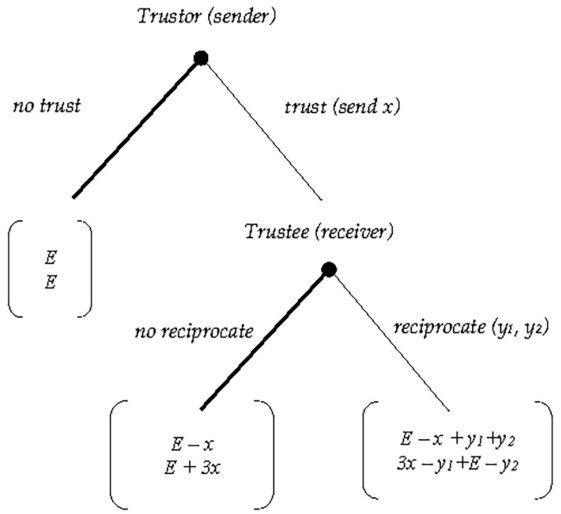

FIGURE 1 | Extensive form of the one-shot TG with fixed endowment.

nor reciprocity is a possible result under the assumptions of rationality and selfishness of both players.

Our subjects played this TG repeatedly during 12 rounds. Each round each subject was randomly matched with a different participant in the same session. Each session had 5 groups of 8 people each, so that each group can be considered an independent observation in our analysis.

Treatments

The three treatments of our experiment are the following (see **Table 1**):

Treatment B. Baseline treatment in which subjects play the TG during 12 periods with fixed initial endowment, random matching and fixed roles. At the end of each round, each player receives feedback on own payoffs and accumulated payoffs in that specific round. No feedback about the partner's earnings is given at all.

Treatment H–Treatment with *History*. It is the same as the baseline with subjects receiving at the end of each period, feedback about own as well as the partner's total earnings accumulated in the past.

Treatment E–Treatment with *Effort*. This treatment differs from the other treatments in that the initial endowment is endogenous. In this treatment, subjects play first an Effort Task (with three sub-tasks). The endowment of each subject depends on the performance of the subject in the three tasks. In particular, we established a linear relation between the endowment and the final score so that a certain level of heterogeneity was assured.

Experimental Session

Two sessions were run of each treatment. Each experimental session included different stages, most of them common to all treatments. **Table 2** shows in detail the stages of a session:

Stage 0. Real-effort tasks (only for Treatment E)

TABLE 1 | Experimental treatments.

Treatment	Endowment	Sessions	Subjects	Females
Baseline -Treatment B	50 ExCUs	2	80	47.50%
History -Treatment H	50 ExCUs	2	80	41.25%
Effort -Treatment E	[10, 100] ExCUs	2	80	48.75%

Subjects performed three individual tasks. The first is related with visual search and consisted in counting ones; the second was of cognitive nature and subjects had to sum 3-digit numbers; the third was miscellaneous and consisted in answering multiple choice questions on general knowledge. In the following we describe each task in more detail:

– *Task 1. Counting number of ones* (Mohnen et al., 2008; Abeler et al., 2011): In a sequential way, each computer screen showed to the subject a 6×6 matrix with randomly ordered 0 and 1s. The subject had to count and write the number of 1s, with no feedback about whether the answer was correct. The participants solved as many matrices as possible during 3 min.

– *Task 2. Summing 3-digit numbers* (Niederle and Vesterlund, 2007): In a sequential way, subjects had to add four 3-digit numbers, without getting any feedback about whether the answer was correct. The participants summed as many series as possible within a time period of 3 min.

– *Task 3. General Knowledge* miscellaneous quiz questions about society, history, geography, maths etc., general knowledge that everyone may acquire through formal education (Coane and Umanath, 2021). The task was programmed with a maximum of 50 questions. Each subject had to answer as many questions as possible during 2.5 min.[2]

Stages 1 and 3. The Equality Equivalence test (EET-pre and EET-post)

The EET (also known as EE-test) was developed by Kerschbamer (2015) to elicit a subject's distributional preference type. It is based on two panels with 5 binary choices that affect both own payoff and other's payoff (see **Table 3**). In the first panel (benevolence behind), decisions are made between receiving the same payoff as the other or a lower one (disadvantageous inequality, x-list). In the second panel (benevolence ahead), decisions are made between receiving the same or a higher payoff as the other (advantageous inequality, y-list). The structure of the test is such that, in order to fulfill the m-monotonicity property, a rational subject decides to switch her decision from equality to inequality once at most.

This test reveals how benevolent the subject is in the domains of disadvantageous and advantageous inequality. We use this test in order to control for social archetypes. Computing the (*x-score*, *y-score*) as described in Holzmeister and Kerschbamer (2019, p. 219), we are able to identify four behavioral archetypes[3]:

[2]It was unlikely that the subject could answer the 50 questions before the time was over.

[3]A positive (negative) *x-score* corresponds to benevolent (malevolent) behavior in the domain of disadvantageous inequality, whereas a positive (negative) *y-score* corresponds to benevolent (malevolent) behavior in the domain of advantageous inequality. Both scores vary from −2.5 to +2.5 (given that there are 5 questions

TABLE 2 | Structure of an experimental session.

Stage	Decision making
0	Real-effort tasks (only in Treatment E)
1st	EET-pre
2nd	Trust Game
3th	EET-post
4th	Socio-demographic questionnaire
5th	Questionnaire on trust and reciprocity

altruist (b, b), spiteful (m, m), inequality loving (b, m), and inequality adverse (m, b).

In all treatments, this test was performed by the subjects before (EET-pre) and after (EET-post) playing the TG. Our motivation is that the decisions made in the TG may affect the distributional preference of the subjects. In Treatment E we informed the subjects that this task had no relation whatsoever with part "zero" of the session–the real-effort tasks. In each of the two performances of this test, each subject is randomly matched with another anonymous participant in the room.

Stages 4 and 5. Questionnaire

In the final part of the session, subjects had to answer a questionnaire that was divided in two parts.[4] In the first part, questions related to socio-demographic issues like gender, age, studies, job, and housing were formulated. In the second part, the questions focused on personality traits related to trust and trustworthiness (Evans and Revelle, 2008),[5] negative and positive reciprocity (Caliendo et al., 2012, 2014),[6] and empathy (Spreng et al., 2009).[7]

Participants

The sessions were run within the time period November 2018 - November 2019 in the Laboratorio de Economía Experimental (LEE), at the Universitat Jaume I in Castellón (Spain). The experiment was programmed in z-Tree (Fischbacher, 2007). Participants were all students from several degrees (engineering, health science, humanities, social sciences, etc.) taught at that University and were recruited using ORSEE (Greiner, 2015). Two sessions of 40 subjects per treatment were run, with a total of 240 participants (80 per treatment). Each session lasted around 90 min and average payoffs were 14 euros per subject.

Research Questions

The central objective of our study is to analyze the effect of economic inequality in human decisions related to trust, reciprocity and altruism. The results by Berg et al. (1995) constitute our reference's point in the formulation of our research questions. Throughout the paper the absolute amount of money the trustor sends to the trustee is denoted as "trust level", and the absolute amount the trustee sends back to the trustor from the money received (initial endowment) is denoted as "reciprocity level" ("altruism level"). Four are the main research questions of our design:

RQ1. The decisions of the trustor in the TG are expected to show that the level of trust observed in one period depends on the reciprocity experienced in the last round, and this relation has a positive sign.

This is a result expected in any TG, independently of the treatment. In fact, it is assumed that one of the motivations of the trustor for sending a positive amount to the trustee is her expectations about receiving some amount back from him.[8]

in each test), being the relevant magnitude the number of times in which the subject chooses RIGHT(*x-score*)/LEFT(*y-score*). Therefore, the correspondence of each archetype and score interval is: altruist ($x > 0, y > 0$); inequality loving ($x > 0$, $y < 0$); spiteful ($x < 0, y < 0$); inequality adverse ($x < 0, y > 0$).

[4]The questions are available from the authors upon request.

[5]Evans and Revelle (2008) use "The propensity to trust survey (PTS)" and find evidence that trust and trustworthiness are compound personality traits, and that PTS scales are preferable to general Big Five measures for predicting trusting behavior.

[6]Caliendo et al. (2012, 2014) use the German Socio-Economic Panel (SOEP).

[7]Spreng et al. (2009) use the Toronto Empathy Questionnaire (TEQ).

[8]Other authors like Attanasi et al. (2019) underline the relevance of the reputation effect on players' decisions. According to them, reputation is a significant

TABLE 3 | Equality Equivalence Test (EET).

LEFT				RIGHT	
You receive	**Another person receives**			**You receive**	**Another person receives**
Benevolence behind					
3.2	5.2	LEFT	RIGHT	4	4
3.6	5.2	LEFT	RIGHT	4	4
4	5.2	LEFT	RIGHT	4	4
4.4	5.2	LEFT	RIGHT	4	4
4.8	5.2	LEFT	RIGHT	4	4
Benevolence ahead					
3.2	2.8	LEFT	RIGHT	4	4
3.6	2.8	LEFT	RIGHT	4	4
4	2.8	LEFT	RIGHT	4	4
4.4	2.8	LEFT	RIGHT	4	4
4.8	2.8	LEFT	RIGHT	4	4

RQ2. Compared with the baseline, in the treatment with heterogeneous endowment (Treatment E) the trustor sends, on average, lower amounts to the trustee. However, no general effect is expected on the trustee's behavior.

This research question motivates our Treatment E. In fact, deciding on how much money to send to an anonymous partner may be affected by the origin of the initial endowment. Specifically, if the endowment comes from performing several tasks, this fact is expected to play a significant role in trustor's decisions in comparison with the situation in which the endowment comes as manna from heaven. However, this endowment heterogeneity is not expected to play a role in trustees' behavior, since the reciprocity level is considered to be purely affected by the decision of the corresponding trustor.

RQ3. In Treatment E where subjects play with unequal initial endowment, to have higher endowment positively affects trust and reciprocity levels.

Playing Treatment E results in "endogenous inequality," since the endowment depends on the performance of the subject in three real-effort tasks. We speculate that the origin of the inequality may have an effect on trustors and, specifically, we believe that having a higher endowment makes the trustor/trustee more likely to send a higher amount to the trustee/trustor.

RQ4. In Treatment H, to be the one with higher cumulated earnings positively affects trust and reciprocity levels.

Treatments H and E may result in economic inequalities among the subjects. Specifically, playing Treatment H results in economic inequality given that -except for the first period- subjects may end up with different cumulated earnings. The same argument described for RQ3 holds here in the sense that, for a trustor/trustee, being the one with higher accumulated earnings makes it more likely to send a higher amount to the trustee/trustor. Individual experiences, characteristics and situations can influence trust levels. Then, a situation of economic advantage/disadvantage can condition trust and reciprocity decisions. Alesina and La Ferrara (2002), in a broad analysis of individual and community characteristics that influence how much people trust each other, point to the economic unsuccessfulness in terms of income as one of the main factors that reduce trust levels. Similarly, we can expect that individuals with more resources tend to trust and reciprocate more (Yan and Miao, 2007). Thus, both in RQ3 and RQ4 we expect a positive effect of economic advantage on trust and reciprocity.

RESULTS

In this section we present the results of our data analysis. First, we summarize the non-parametric analysis, carrying out a general perspective of the data obtained and offer some preliminary insights. After that, we show the adjustment of an econometric model for the behavior of both types of players, trustor and trustee, in which some related variables identified during the experiment are included.

Real-Effort Tasks (Only in Treatment E)

The score in a task is the sum of total correct answers. The global score in each task is computed by the sum of the three scores weighted by the value of a correct answer. Because the difficulty level is heterogenous[9] among tasks, task 1 is taken as the reference task. The tasks requiring higher effort were given a higher weight. Weights of 25, 40, and 35% were applied for task 1, task 2 and task 3, respectively. Thus, a correct answer is task 1 has a value of 1 point, of 1.6 points in task 2, and of 1.4 in task 3. Mistakes were allowed in the three tasks but incorrect answers were not considered for the final score. The total score for each subject was therefore calculated as:

$$\text{Score} = (1 \times N_1) + (1.6 \times N_2) + (1.4 \times N_3)$$

Where N_i is the number of correct answers in task i.

On average, making a ranking in the negative domain, task 3-general knowledge questions was performed the worst, with an average error rate of 32.26%. Task 2-summing four 3-digits numbers was the second in the ranking, reaching 30.53% of incorrect answers. Task 1-counting ones was, as expected, the best performed, with 11.86% rate of average error (see **Table 4**).

The system then had to calculate the initial endowment of each subject in order to start the part dedicated to playing the TG. The endowment of each participant after performing the tasks was calculated in such a way that differences in performance could guarantee enough heterogeneity among endowments in the total population. More specifically, all endowments were within a closed interval in which 10 ExCUs was the minimum value and 100 ExCUs the maximum. Specifically, the endowment of each subject was calculated as $E_i = [10 + (100–10)$ score/max score$]$.

Final Questionnaire's Results

The second part of the final questionnaire consists in answering questions about personality traits related to trust, trustworthiness, negative and positive reciprocity, and empathy. **Table 5** reports some statistics about this data analysis. Specifically, an equal weighted index is computed on the 4-point Likert items of questions corresponding to each category. The categories are: trust, trustworthiness, reciprocity and empathy. Looking for any gender effect, significant differences are found only in negative reciprocity and empathy categories. According to the rank-sum M-W test, males (females) show a higher (lower) score than females (males) with a probability of 0.620 (0.626) in negative reciprocity (empathy).

Equality Equivalence Test Performance

This test allows us to identify four archetypes in our data sample: "inequality loving," "spiteful," "inequality adverse," and "altruist."

motivation that positively affects share decisions under a partner matching context, unlike our random matching design where the reputation effect is blurred.

[9]Specifically, the weights were chosen considering that task 1 is the easiest and task 2 is the more difficult for an average person. Even though not directly, the criteria we followed are related with the concepts of control and routine processing indicated in Goldhammer et al. (2014). Task 1 is considered the easiest, since it just requires visual speed; task 3 requires a more routine processing; task 2 includes more difficulty, since it demands for more control processing.

TABLE 4 | Average error rates, global score, and final endowment in the real-effort tasks, by gender.

	Task 1 25%	Task 2 40%	Task 3 35%	Global Score	Endowment in ExCUs	Obs.
Females	10.37% (0.12)	30.43% (0.22)	35.56% (0.15)	29.17 (9.81)	51.51 (12.24)	39
Males	13.28% (0.14)	30.63% (0.23)	29.31% (0.14)	35.69 (10.64)	61.67 (17.51)	41
All	11.86% (0.13)	30.53% (0.22)	32.26% (0.15)	32.51 (11.25)	56.72 (16.74)	80

Std. dev. in parenthesis. Rates of standard deviation error are expressed in decimal numbers.

TABLE 5 | Summary of personality traits, by gender.

Index (%)	Interpersonal trust	Intrapersonal trustworthiness	Positive reciprocity	Negative reciprocity	Empathy	Obs.
Females	2.74 (0.36)	3.15 (0.31)	3.46 (0.49)	1.81 (0.57)	3.25 (0.30)	110
Males	2.74 (0.33)	3.11 (0.38)	3.49 (0.49)	2.07 (0.64)	3.10 (0.32)	130
All	2.74 (0.35)	3.13 (0.35)	3.48 (0.49)	1.95 (0.63)	3.17 (0.32)	240
Ranksum M-W test	z = 0.454	z = −0.664	z = 0.520	z = 3.276	z = −3.382	
	p = 0.6500	p = 0.5067	p = 0.6032	p = 0.0011	p = 0.0007	
Males have a higher score than Females with probability	0.475	0.517	0.519	0.620	0.374	
Sign test	Low Median < 3	High Median > 3	High Median > 3	Low Median < 3	High Median > 3	
Females	p = 0.0000	p = 0.0000	p = 0.0000	p = 0.0000	p = 0.0000	
Males	p = 0.0000	p = 0.0008	p = 0.0000	p = 0.0000	p = 0.0060	

Average values and standard deviation in parenthesis. Index computed as an average of items in each category.

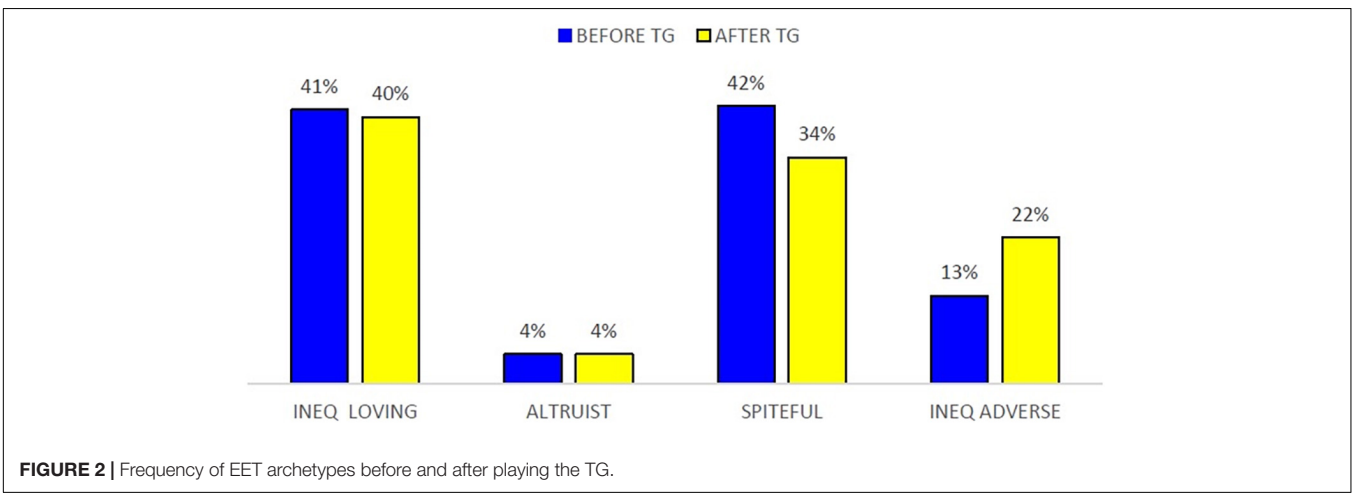

FIGURE 2 | Frequency of EET archetypes before and after playing the TG.

A possible TG-effect[10] on participants' individual choices in the EET is analyzed. Such effect is represented as a change in the percentage of participants assigned to each archetype according to their individual choices. **Figure 2** shows the percentage of each archetype before and after playing the TG. It is observed that the archetypes "spiteful" and "inequality adverse" experience a significant change after playing the TG: a higher number of "inequality adverse" participants are found after playing the TG, whereas the "spiteful" type notably decreases. This can be observed in **Figure 3**, showing the presence of archetypes in each treatment before and after playing the TG.

Differentiating by treatment (**Figure 3**), in Treatment B it is detected a general TG-effect on individual choices in EET with major contributions to symmetric payoffs by "inequality loving" and "spiteful" archetypes. It is recorded an increase of 15.79 percentage points in inequality loving participants, and this fact is clearly explained by the decrease in spiteful participants. Observe that in Treatment H the "inequality loving" and "inequality averse" the archetypes contributing to the TG effect most. Specifically, the participants classified as "inequality loving" fall 18.5 percentage points after playing the TG, and the ones

[10]The TG-effect on EET is measured through a contrast in the global sample (exact symmetry test, p = 0.0198; Stuart-Maxwell (SM) test for marginal homogeneity, chi2 = 7.57, df = 3, p = 0.0558). Such an effect exists in Treatment B (exact symmetry test, p = 0.0430; SM-test for marginal homogeneity, chi2 = 8.17, df = 3, p = 0.0426), and in Treatment H (exact symmetry test, p = 0.0608; SM-test, chi2 = 8.47, df = 3, p = 0.0372). No TG-effect on EET is found in TE (exact symmetry test, p = 0.3579; SM-test, chi2 = 2.06, df = 3, p = 0.5592).

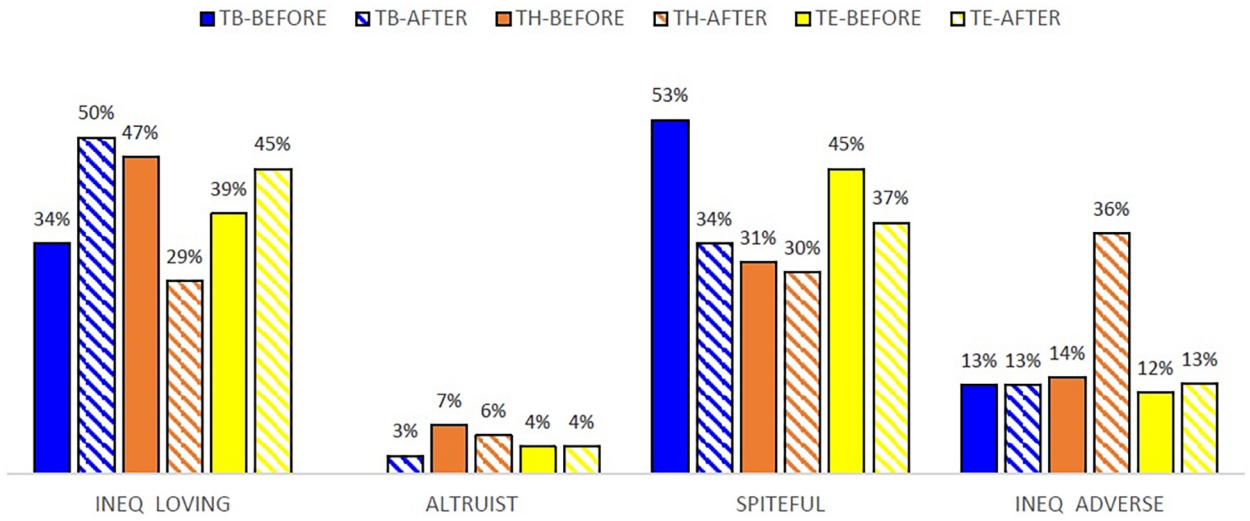

FIGURE 3 | Frequency of EET archetypes before and after the TG, per treatment.

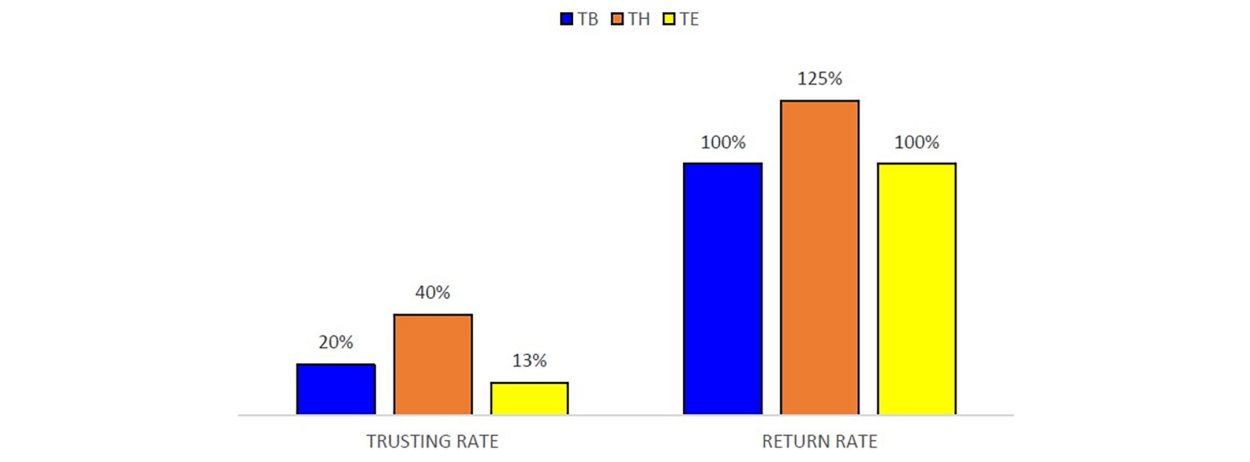

FIGURE 4 | Median decisions in the TG, per treatment.

classified as "spiteful" fall 1.4 percentage points. In contrast, the "inequality adverse" increase 21.40 percent points. In Treatment E, the TG effect on individual choices is negligible. In fact, the increase of 6.6 percentage points in inequality loving is offset by the decrease of 8 percentage points in spiteful participants.

The above evidence allows confirming that providing participants with information about the partner's cumulated earnings during the finite repeated TG has a significant effect on the EET choices. In fact, Treatment H shows this effect especially intense on the players who were labeled as "inequality lovers" in the pre-TG then converted in "inequality adverse" in the post-TG. Our interpretation is that participants' social preferences reflected on the EET are sensitive to the -maybe negative- experience playing the TG. Interestingly, no TG-effect is found on EET when participants earn their initial endowment with their own effort in Treatment E.

Trust, Reciprocity and Altruism: A Non-parametric Analysis

This subsection presents the results from a non-parametric analysis implemented on trust, reciprocity and altruism decisions.

Trust

In order to make the three treatments comparable, the decision of trust is measured as the percentage of the initial endowment, what we call *trusting rate* (see **Figure 4**).

A first general result is that significant differences among treatments are found with respect to trust, implying that the decisions of the trustor are endowment as well as cumulated earnings dependent. Observe in **Figure 4** that in median values, the trustor sends 20% of the endowment in Treatment B, 40% in Treatment H, and 13% in Treatment E. That is, in comparison with the baseline, trust is significantly lower when the initial

TABLE 6 | Testing treatment and gender effects on TG decisions.

	Treatment differences			Gender differences (males vs. females)		
Trusting rate	$z_{BH} = -3.820$ $p = 0.0001$	$z_{BE} = 4.743$ $p = 0.0000$	$z_{HE} = 8.512$ $p = 0.0000$	$z_B = 6.109$ $p = 0.0000$	$z_H = 2.663$ $p = 0.0078$	$z_E = -0.477$ $p = 0.6332$
Return rate	$z_{BH} = -2.267$ $p = 0.0234$	$z_{BE} = -0.157$ $p = 0.8750$	$z_{HE} = 1.985$ $p = 0.0471$	$z_B = 0.801$ $p = 0.4234$	$z_H = -1.846$ $p = 0.0648$	$z_E = 2.311$ $p = 0.0208$
Reciprocity	$z_{BH} = -3.622$ $p = 0.0003$	$z_{BE} = -0.780$ $p = 0.4354$	$z_{HE} = 2.745$ $p = 0.0060$	$z_B = 0.165$ $p = 0.8692$	$z_H = -0.330$ $p = 0.7417$	$z_E = 1.163$ $p = 0.2447$
Altruism	$z_{BH} = -0.207$ $p = 0.8360$	$z_{BE} = 3.619$ $p = 0.0001$	$z_{HE} = 3.207$ $p = 0.0013$	$z_B = 0.440$ $p = 0.6602$	$z_H = -3.499$ $p = 0.0005$	$z_E = -1.051$ $p = 0.2932$

The test applied is the Wilcoxon rank-sum test for two independent samples.

endowment is endogenously determined through real-effort tasks, but significantly higher in the case the trustor knows the cumulated earnings of the corresponding trustee before deciding in the next period (see **Table 6**).

Focusing on Treatment H, trustors with higher cumulated earnings than their partners, send a significantly higher (in median) percentage of their endowment to the trustee compared to trustors with equal or lower cumulating earnings than their partners. The opposite result is obtained when extrapolating to Treatment E, i.e., the trustor rate (in median) is lower for trustors with higher endowment than their (see **Table 7**).

The same **Table 6** shows the comparison among treatments of the gender of the trustor. Observe that in Treatment B and Treatment H females send, in median, significantly lower amounts than males.[11] This is very much in line with several previous results in the literature on trust (Buchan et al., 2008; Dittrich, 2015).

Reciprocity and Altruism

It has been already mentioned in section "Materials and Methods" that our design includes two decisions for the trustee: a reciprocity decision that accounts for the amount sent back to the trustor from the total amount received; as well as an altruism decision that accounts for the amount sent to the trustor from the initial endowment. The analysis of reciprocity and altruism are measured using the *return rate*, defined as the total amount sent by the trustee divided by the amount sent by the trustor.

Figure 4 shows the return rate per treatment. In median values, the return rate is 100% in Treatment B and Treatment E, indicating that the trustor sends and receives the same amount. Furthermore, in the treatments with heterogeneity, Treatments H and E, the return rate is not significantly different independently on the advantage/disadvantage that it may exist in cumulated earnings (Treatment H) or endowment (Treatment E) with respect to the partner (see **Table 7**).

Figure 5 presents the trustees' reciprocity and altruism decision separately, in average percentage. It is found that the reciprocity decision is higher in Treatment H than in the other two treatments, on average as well as in median values. Furthermore, Treatments B and E do not show significant differences concerning reciprocity. This may indicate that

TABLE 7 | Testing the effect of earnings/endowment inequality on TG decisions.

	Treatment H	Treatment E
Trusting rate	WC signed-rank test: $z = -3.113$, $p = 0.0019$ Left-sided sign test: $p = 0.0003$	WC signed-rank test: $z = 2.534$, $p = 0.0113$ Right-sided sign test: $p = 0.0171$
Return rate	WC signed-rank test: $z = 1.023$, $p = 0.3065$ Two-sided sign test: $p = 0.6089$.	WC signed-rank test: $z = -0.521$, $p = 0.6022$ Two-sided sign test: $p = 0.3997$
Recipro-city	WC signed-rank test: $z = 0.691$, $p = 0.4898$ Two-sided sign test: $p = 0.8974$	WC signed-rank test: $z = -2.376$, $p = 0.0175$ Left-sided sign test: $p = 0.0059$
Altruism	WC signed-rank test: $z = 2.115$, $p = 0.0344$ Right-sided sign test: $p = 0.0135$	WC signed-rank test: $z = 1.595$, $p = 0.1106$ Two-sided sign test: $p = 0.3560$

The tests are Wilcoxon signed-rank test for two dependent samples and sign test. The difference between two groups of subjects is tested: a group with those with no advantage with respect to the partner, and another group with advantage. The groups change in each period, and therefore samples are not independent.

reciprocity is not primarily determined by inequality on initial endowment, but by inequality built as the game is played and players are aware of the information about the other's cumulated earnings. It seems therefore that the reciprocity decision is cumulated earnings-dependent.

Regarding altruism decisions, treatment significant differences are found only between Treatment E and the other two treatments, indicating that inequality in the initial endowment is relevant as far as the altruism decision in the TG is concerned (see **Table 6**). Looking specifically at each treatment, in Treatment E it is observed that the reciprocity decision is significantly higher when trustees have superior endowment than their partners compared to trustees with equal or inferior endowment. On the contrary, in Treatment H the altruism decision of trustees with equal or lower cumulated earnings than their partner is significantly higher than that of trustees with higher cumulated earnings compared to their partner (see **Table 7**).

Contrary to the role of the trustor, the significant gender differences with respect to the role of the trustee are found in Treatment E. Females send back, in median, significantly lower amounts than males. In Treatment H the same effect is found but

[11] See also Table A in the **Supplementary Material file** for details on trustors' decisions by gender. Furthermore, no significant gender differences are found in Treatment E.

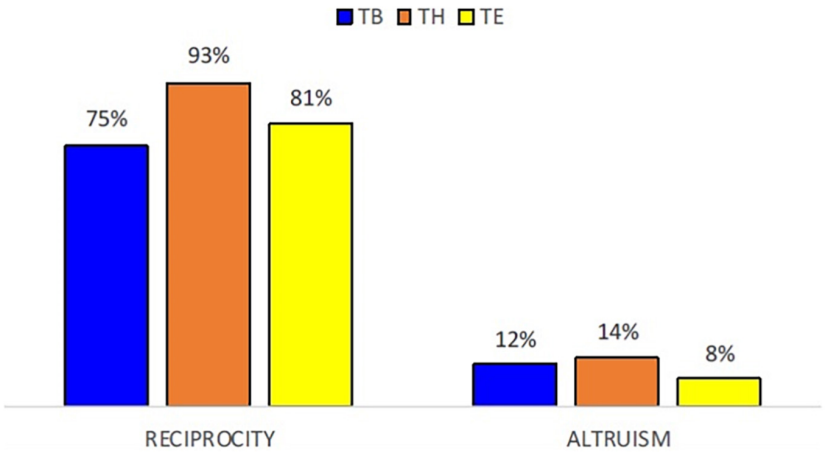

FIGURE 5 | Average levels of reciprocity and altruism in the TG, per treatment.

the significance is weak. A between treatments analysis shows that females are found to be significantly more altruistic in Treatment H than in the other two treatments.

ECONOMETRIC ANALYSIS

In this section we use multivariate regression models to enrich the previous non-parametrical analysis with additional interesting results non-captured by a non-parametric analysis.[12] **Table 8** contains the definitions of both dependent and independent variables. Specifically, the trustor's decision is modeled by using a multivariate linear regression model; the trustee's double decision is estimated through two probability models. On one hand, the reciprocity decision is modeled by a 3-level ordered logit model that estimates the probability of each reciprocity level. On the other hand, the altruism decision is modeled by a binary logit model. All models are included in **Table 9**. Observe in the table that the dependent variables are "trusting rate" (column 1), "reciprocity level" (columns 2–5), and "altruism level" (columns 6–7). For probability models, marginal effects are also shown (columns 3–5, and 7). Concerning ETT data, the results in the table have been calculated under the hypothesis that the m-monotonicity property in the decisions of the EET-pre, is fulfilled.[13]

Trustors' Behavior

Table 9 shows our estimation for the trustors' behavior: a lineal model by GLS with random-effects and cluster–robust standard errors for panels nested within groups.

We find a positive and significant relationship between the trustor's decision in the current period *t* and the trustee's decision not only in the previous period (*t-1*), but also the previous to the previous period (*t-2*). Our first result summarizes this general finding:

> ***Result 1*** Independently on the treatment, the trustor's decision each round is influenced by her recent interaction with the correspondent trustee. In each specific round, the higher the amount returned by the trustee in the previous (up to two) period(s), the higher is the trust transmitted by the trustor, thus sending a higher amount.

However, looking further in the treatments with inequality (Treatments H and E) we search the possible that the difference between own and the partner's initial endowment may have on the trustor's decision, the evidence splits up. First, with respect to TE, when the trustor's endowment is greater than the trustee's (Em – Eo > 0), the corresponding regression coefficient presents a significant negative sign (−0.09%, $p = 0.001$), which is aligned with the non-parametric evidence commented in the previous subsection. However, in the opposite case (Eo – Em > 0), the coefficient is significant and positive (0.12%, $p = 0.000$). Therefore, our second result states that:

> ***Result 2*** When the trustor's endowment is higher (lower) than the trustee's, the trust level is negatively (positively) affected, sending less (more) money to the partner.

Thus, our RQ3 is partially confirmed, since it holds only for the case in with the trustor has lower endowment than the trustee.

Second, looking at Treatment H, also mixed are the results obtained when looking at the effect of the differences between own and the other's accumulated earnings. In particular, observe in **Table 9** (column 1) the corresponding regression coefficient is negative and no significant (−0.02%, $p = 0.442$) when the cumulated earnings of the trustor are higher than those of the trustee (Gm-Go > 0). Moreover, when the contrary happens, i.e., Go – Gm > 0, the coefficient is negative (−0.03%) and

[12] An ex-post power analysis has been conducted using Stata with power set at 0.80 and probability at 0.05. It is obtained that the sample size necessary for statistically significant differences at 5% between baseline and treatment should be at least $N = 239$ observations.

[13] Results from the EET-post are not included in this analysis, given that the number of inconsistencies in the post experiment test was enough to un-equilibrate our number of observations and, therefore, the interest in the comparability. For the purpose of the econometric analysis, we believe that the spirit of the test is more genuinely captured by the decisions taken in the EET-pre.

TABLE 8 | Definition of variables.

Dependent variables
Trusting rate: Amount sent by the trustor/Endowment (*x*/E)
3-level reciprocity variable: {L1, L2, L3}{L1, L2, L3}
– L1: Reciprocity amount (y_1) is smaller than the egalitarian amount
– L2: Reciprocity amount (y_1) equals the egalitarian amount
– L3: Reciprocity amount (y_1) is higher than the egalitarian amount
Egalitarian amount:
– In Treatment B: $y_1 = 2x$
– In Treatment H: *Max* {(Gm–Go)/2 + 2*x*, 0} ≤ 3*x*; Gm denotes own (m stands for myself) cumulated earnings and Go is the other's cumulated earnings
– In Treatment E: *Max* {(Em–Eo)/2 + 2*x*, 0} ≤ 3*x*; Em denotes own (m stands for myself) initial endowment and Eo is the other's endowment
Altruism: Binary variable (taking value 0 if the amount sent from the endowment (y_2) = 0; or 1 if the amount sent is (y_2) > 0)
Independent variables
Trustor amount (x): amount sent by the trustor
Reciprocity amount (y_1): amount returned by the trustee from the total amount (3x) received from the trustoraAltru
Total returned amount lag = 1: total amount sent by the trustee in period t-1
Total returned amount lag = 2: total amount sent by the trustee in period t-2
Economic inequality:
– Max{Em – Eo, 0}: Own (m stands for myself) initial endowment (Em) is higher than the other's (Eo)
– Max{Eo – Em, 0}: The other's initial endowment (Eo) is higher than mine (Em)
– Max{Gm – Go, 0}: Own (m stands for myself) cumulated earnings (Gm) are higher than the other's (Go)
– Max{Go – Gm, 0}: The other's cumulated earnings (Go) are higher than mine (Gm)
Gender: dummy variable (0-Male, 1-Female)
EET-types: dummy variable (Spiteful, Inequality-lovers, Inequality-averse, Altruist)
Treatments: dummy variable (Treatments B, H and E)
Personality related questions: 4-point Likert scale (1 strongly disagree, 2 disagree, 3 agree, 4 strongly agree)
– I always act fairly with others. (Trustworthiness)
– If you deal with strangers it is better to be careful before trusting them. (Trust)
– I go out of my way to help someone who was previously nice to me. (Reciprocity)
– I think most people lie to take advantage of others. (Negative Trust)
– I would never evade my taxes. (Trustworthiness)
– If someone offends me, I will offend them. (Negative Reciprocity)

statistically significant (p = 0.000). This implies that RQ2 is not confirmed. Here our third result:

> ***Result 3*** When the trustor cumulated earnings are higher than those of the trustee, this advantage has a significant negative effect on the trust decision, sending a lower amount to the trustee. No significant effect is found otherwise.

Additionally, to catch a possible treatment effect, we include dummy variables in our analysis, where Treatment B is taken as the reference treatment. Only Treatment E is found to be statistically significant. In other words, in Treatment E the trustor sends an amount (10% lower) that is significantly different (see **Table 9**, column 1) from that of the trustor in Treatment B. On the contrary, no statistical differences are found between trustors' decisions in Treatments H and B. This is related to our second research question, and allows us to state that:

> ***Result 4*** Trust is found to be significantly lower in Treatment E than in Treatment B. No other differences between treatments are found with respect to the trust decisions.

Finally, our analysis on the influence of personality traits on the trustors' decision finds statistically significant positive differences between altruist trustors and inequality loving ones.

Trustees' Reciprocity

We estimate random-effects ordered logistic regression with cluster–robust standard errors for panels nested within groups. For the analysis of the trustees' decisions, we have created three dependent variables (levels L1, L2, L3), associated with the egalitarian strategy,[14] that take values 1, 2, 3, respectively, indicating that the trustee returns an amount lower than (L1), equal to (L2) or higher than (L3) the egalitarian amount, respectively. **Table 9** reports the coefficients and marginal effects.

Regarding the relationship between the amount sent by the trustor and the amount returned by the trustee, we find a negative and significant coefficient in the regression which implies that, in general terms, the higher the amount sent by the trustor, the lower the (total) amount sent back by the trustee. Taking as a reference the egalitarian amount and differentiating by levels, we

[14] A trustee that follows this strategy chooses the amount y1 such that the payoffs of both players are equal that round. This implies that trustor's payoff $\pi_{or} = E\text{-}x+y_1$ has to be equal to the trustee's $\pi_{ee} = 3x-y_1 + E$. In TB, the egalitarian amount is: $y_1 = 2x$. In Treatment H, the egalitarian amount is balanced by the earnings inequality: $y_1 = 2x + (\pi_{ee}-\pi_{or})/2$. In Treatment E, the egalitarian amount considers the endowment inequality, i.e., equal payoffs imply that $E_{or}-x+y_1 = 3x-y_1 + E_{ee}$; therefore, the egalitarian amount in Treatment E is $y_1 = 2x + (E_{ee}-E_{or})/2$. In all cases, the amount sent by the trustee has to fulfill the non-transfer restriction along rounds and, therefore, $y_1 \in [0, 3x]$.

TABLE 9 | Econometric models for trustors (1) and trustees (2–7).

	1	2	3	4	5	6	7
	Trusting rate	Recipr. Level	L1	L2	L3	Altruism	
	COEF	COEF	ME	ME	ME	COEF	ME
Trustor amount (x)		−0.2446***	0.0237***	−0.0156***	−0.0081***	0.0495***	0.0054***
		(0.0712)	(0.0050)	(0.0043)	(0.0011)	(0.0171)	(0.0019)
Reciprocity amount (y)						−0.0141	−0.0016
						(0.0133)	(0.0015)
Total returned amount lag = 1	0.0041***						
	(0.0007)						
Total returned amount lag = 2	0.0033***						
	(0.0004)						
Max{Em – Eo, 0}	−0.0009***	−0.0535***	0.0053***	−0.0034***	−0.0018***	0.0301***	0.0033***
	(0.0002)	(0.0112)	(0.0012)	(0.0007)	(0.0006)	(0.0044)	(0.0005)
Max{Eo – Em, 0}	0.0012***	0.0897***	−0.0087***	0.0057***	0.0030***	−0.0113	−0.0012
	(0.0001)	(0.0230)	(0.0018)	(0.0015)	(0.0005)	(0.0129)	(0.0014)
Max{Gm – Go, 0}	−0.0002	0.0014*	−0.0001*	0.0001*	0.00005*	−0.0021***	−0.0002***
	(0.0002)	(0.0007)	(0.0001)	(0.0000)	(0.0000)	(0.0008)	(0.0001)
Max{Go – Gm, 0}	−0.0003***	0.2415***	−0.0234***	0.0154***	0.0080***	0.0363	0.0040
	(0.0001)	(0.0568)	(0.0038)	(0.0034)	(0.0010)	(0.0258)	(0.0029)
Female	−0.0290	−0.0579	0.0056	−0.0037	−0.0019	0.5126	0.0566
	(0.0333)	(0.1564)	(0.0153)	(0.0100)	(0.0053)	(0.7047)	(0.0765)
Inequality loving	0.0850***	0.0963	−0.0095	0.0063	0.0032	−0.5867	−0.0647
	(0.0298)	(0.1392)	(0.0137)	(0.0090)	(0.0048)	(0.6199)	(0.0701)
Inequality averse	0.0073	−0.5331	0.0488***	−0.0340***	−0.0149***	0.0965	0.0107
	(0.0374)	(0.1273)	(0.0124)	(0.0084)	(0.0049)	(0.7786)	(0.0863)
Altruist	0.0563	1.3755**	−0.1430**	0.0783***	0.0647	−0.6933	−0.0763
	(0.0542)	(0.6915)	(0.0664)	(0.0283)	(0.0398)	(0.2079)	(0.2229)
Treatment H	0.0505	0.0260	−0.0024	0.0017	0.0008	0.3650	0.0401
	(0.0455)	(0.3729	(0.0348)	(0.0239)	(0.0109)	(0.8688)	(0.0939)
Treatment E	−0.1084***	0.4831	−0.0469	0.0307	0.0162	0.1939	0.0212
	(0.0401)	(0.3967)	(0.0386)	(0.0248)	(0.0141)	(0.7660)	(0.0833)
I always act fairly with others. (Trustworthiness)	0.0775***						
	(0.0203)						
If you deal with strangers it is better to be careful before trusting them. (Trust)	−0.0181						
	(0.0220)						
I go out of my way to help someone who was previously nice to me. (Reciprocity)		0.2122	−0.0206	0.01353	0.0070*		
		(0.1419)	(0.0132)	(0.0091)	(0.0042)		
I think most people lie to take advantage of others. (Neg.Trust)						−0.9156***	−0.1008***
						(0.3336)	(0.0393)
I would never evade my taxes. (Trustworthiness)						0.2083	0.0229
						(0.4205)	(0.0456)
If someone offends me, I will offend them. (Neg. Reciprocity)						−0.9748***	−0.1073***
						(0.2299)	(0.0233)
Constant	0.0164					2.3411	
	(0.0852)					(2519)	
Cutoff point for L1		0.8391*					
		(0.5002)					
Cutoff point for L2		3.5255***					
		(0.6919)					

(Continued)

TABLE 9 | (Continued)

	1	2	3	4	5	6	7
σ_u^2 (panel-level variance)		0.0576					
		(0.0958)					
σ_u (panel-level deviation)	0.0783					2.8251	
						(0.3270)	
σ_e (error term deviation)	0.1673						
$\rho = \sigma_u^2/\sigma_u^2 + \sigma_e^2$	0.1796					0.7081	
						(0.0478)	
R^2 (overall)	0.6359						
Log pseudolikelihood		−465.92				−495.98	
Wald χ^2	687.61***	7244.36***				4331.13***	
Number of observations	920	1152				1152	
Groups	92	96				96	

All regressions are estimated with random-effects and cluster–robust standard errors for panels nested within groups. The trust model is estimated as a linear regression with GLS estimation. COEF indicates regression coefficient, and ME indicates marginal effect. Standard errors in parentheses. Independent variables are defined in ***Table 8****. *, **, *** indicate statistical significance at the 10, 5, and 1% levels, respectively.*

find a positive and significant marginal effect at level L1; that is, an increase in the amount sent by the trustor makes more likely that the trustee returns an amount lower than the egalitarian (2.37%, $p = 0.000$). On the contrary, the marginal effects associated to the decision at levels L2 and L3 are significant but negative: an increase in the amount sent by the trustor makes less likely that the trustee returns the egalitarian (−1.56%, $p = 0.000$) or higher than the egalitarian (−0.81%, $p = 0.001$) amount. Therefore, it may be concluded that:

Result 5 The probability of reciprocating with a higher or equal (lower) than the egalitarian amount decreases (increases) with the trust rate.

With respect to Treatment E, it is important to highlight that when the endowment of the trustee is higher (lower) than that of the trustor, the marginal effect is positive and significant (0.53%, $p = 0.000$) (negative and significant: −0.87%, $p = 0.000$), increasing (decreasing) the probability of returning a lower than the egalitarian amount. Also significant but the opposite is observed at levels L2 and L3, where the estimated probability decreases (increases) by 0.34% (0.57%) and 0.18% (0.3%), respectively, when the trustee has an initial endowment higher (lower) than that of the trustor. In other words, the initial endowment inequality has an effect on L2 or L3 reciprocity decisions that is the opposite to the inequality sense. The opposite is found in L1. Summarizing:

Result 6 Reciprocity is affected by the endowment inequality in the TG. Specifically, the probability that the trustee reciprocates with an amount equal or higher than the egalitarian increases (decreases) when he is the one with the lower (higher) endowment.

Observe that our Result 6 contradicts the second part of RQ2. That is, the trustee's decisions are affected not only by the amount received from the trustor but also by the endowment heterogeneity. Previous literature suggests that the trustee's decisions in the TG are affected by his psychological characteristics (Attanasi and Nagel, 2008; Andrighetto et al., 2015; Di Bartolomeo and Papa, 2016).

Somehow the opposite occurs in Treatment H. Specifically, in the case in which the trustee's cumulated earnings are higher (lower) than those of the trustor, the marginal effect is negative and significant (−0.01%, $p = 0.052$) (negative and significant: −2.34%, $p = 0.000$), decreasing the probability of returning amounts lower than the egalitarian. For levels L2 and L3 we observe the opposite: the estimated probability significantly increases, respectively, by 0.01% ($p = 0.054$) and 0.005% ($p = 0.062$) when the trustee's cumulated earnings are higher than those of the trustor. A significant increase is also estimated when the trustee's cumulated earnings are lower than those of the trustor in L2 (1.54%, $p = 0.000$) and L3 (0.8%, $p = 0.000$). Consequently, the cumulated earnings inequality exhibits a positive (negative) effect on the probability of taking an egalitarian or superior (inferior) reciprocity decision.

Result 7 Inequality in accumulated earnings affects the reciprocity decision in the TG. In particular, the probability of reciprocating with an amount lower than the egalitarian increases (decreases) only when the trustee's accumulated earnings are higher (lower) than those of the trustor. The probability of reciprocating with an amount equal or higher than the egalitarian increases independently on who is richer/poorer.

Two trustees' personality archetypes are found to be statistically significant with respect to the egalitarian strategy: the altruist and the inequality-adverse. Specifically, the altruist is more likely than any other archetype to reciprocate with the egalitarian. Surprisingly, the inequality-adverse is significantly less likely to do that.

Finally, we have estimated the probability of reciprocating for each reciprocity level: 77.1% (in L1) 17.3% (in L2) and 5.6% (in L3). It is not surprising that, given that our sample of trustees is highly represented by selfish and inequality-lovers, the more likely decision has been to reciprocate with an amount that is lower than the egalitarian.

Trustees' Altruism

We estimate random-effects binary logit regression with cluster–robust standard errors for panels nested within groups. The dependent variable (see **Table 8**) takes value 1 indicating that the trustee sends a positive amount from his own initial endowment, and value 0 otherwise. Observe **Table 9** for the coefficients and marginal effects.

First, a general positive and significant relationship is found between trust and altruism decisions, with a positive and significant marginal effect (0.0054, $p = 0.004$), indicating a positive effect on the probability of being altruistic in our design. In fact, it is also observed that how much the trustee gives in the altruism decision does not depend from how much he reciprocates. Therefore:

> ***Result 8*** Independently on the treatment, the altruism decision of the trustee is positively related to the trust decision. Moreover, the altruism decision does not depend on the reciprocity decision, but exclusively on the trustor's decision.

Second, in Treatment E the initial endowment inequality plays a significant positive role in the probability of sending a positive amount in the altruism decision only when the trustee's endowment is higher than that of the trustor's (0.0033, $p = 0.000$). For Treatment H the opposite result is found, that is, when the trustee's cumulated earnings are higher than those of the trustor, there is a significant negative effect on the probability of the trustee adopting an altruistic decision (-0.0002, $p = 0.007$). As a result:

> ***Result 9*** In the altruism stage, it is more (less) likely that the trustee sends a positive amount when he has a higher initial endowment (cumulated earnings) than his corresponding trustor. No significant marginal effect is found otherwise.

Finally, our analysis on the influence of personality traits on the trustees' decision on altruism finds that psychological variables related to inter-personal trust and reciprocity[15] exhibit the highest marginal effects on the probability of sending money in the altruism decision. However, the EET archetypes show no statistical significance.

DISCUSSION AND MAIN CONCLUSION

Our motivation for this paper has been to study the importance of economic heterogeneity in the decision of how much to trust and reciprocate. Our hypothesis is that individuals assign a different value to income resulting from their own effort than to income received without dedicating energy to it in the case of inheritance or a subsidy. In other words, endogenous economic heterogeneity plays a role in trust and reciprocity behavior. Moreover, we have put this endogenous source of inequality in contrast with an exogenous source of economic inequality, that is the case in which individuals have accumulated different amount of money over time. We also have hypothesized that being aware of such economic heterogeneity can affect the trust and/or reciprocity levels of individuals.

To the best of our knowledge, our work is the first in designing a situation in which trust, reciprocity and altruism are analyzed taking into consideration those sources of economic heterogeneity: endogenous through real-effort and exogenous as result of different accumulated earnings. With an experiment in which a finitely repeated version of the TG is at the core of the design, this paper has analyzed whether the levels of trust, reciprocity and altruism are affected by heterogeneity on accumulated earnings and/or initial endowment. Two treatments in our design allow for testing how the fact of knowing how rich the partner is or how well the partner performed several tasks with real effort may really affect the trust, reciprocity or altruism levels in a TG environment. On one hand, a Treatment H in which the cumulated earnings are common knowledge at the end of each round, has made the subjects aware of any income heterogeneity throughout the TG. On the other hand, an alternative Treatment E has introduced three initial real-effort tasks which generate endowment heterogeneity kept fixed at the beginning of each round in the TG. A clear treatment effect has been found confirming that the trust level is endowment as well as cumulated earnings dependent.

A baseline Treatment B with no endowment heterogeneity is taken as a reference. In the three treatments, the trustee takes two decisions: reciprocity and altruism decision. A general result that does not depend on the treatment is that trust decisions are aligned with recent past experience. More specifically, except for the first period, the amount sent to the trustee in one period positively depends on the amount returned from the corresponding trustee in the two previous periods. Interestingly, a positive experience received from the trustee has a positive effect in the attitude of the trustor toward the new trustee(s) in the next round(s). Such experience effect obtained under random matching protocol somehow extends previous results obtained by Attanasi et al. (2019) under partner matching.

In general terms, the literature agrees on the fact that wealth inequality reduces incentives to cooperate (Ciriolo, 2007; Heap et al., 2013; Gallego, 2016). Specifically, Bejarano et al. (2018) show that inequality reduces trust levels only when this inequality is generated by random shocks. Our results show that inequality significantly reduces trust in Treatment E. Also, from the trustee's perspective, the endowment inequality generated in Treatment E affects negatively the levels of reciprocity that are higher or equal to the egalitarian strategy, the one that assures that trustor and trustee enjoy the same payoffs. Previous literature provides evidence of negative effects on reciprocity under different contexts. For instance, Pelligra et al. (2020) find that trustors' expectations do not always have a positive effect on trustees' decision, that is, trustors' expectations expressed as request not always increase reciprocity. Furthermore, Balafoutas and Fornwagner (2017), using a Dictator Game, find that the guilt aversion only affects decisions up to a certain level of recipient expectations. As a result, RQ2 is confirmed.

The present work has addressed also the effect of the inequality direction -which player is the richest-on the levels of trust, reciprocity and altruism. Our results on this diverge from those

[15] Negative-reciprocity and negative-trust (see **Table 9**).

of Ciriolo (2007); Brülhart and Usunier (2012), and Rodriguez-Lara (2018) who find a non-significantly different behavior when facing poorer/richer partners compared to the case of equality. More in line with Xiao and Bicchieri (2010) and Smith (2011b), our analysis results in significant effects of inequality direction on trust and reciprocity levels, especially in the treatment with effort. Surprisingly, whenever the trustor's endowment is higher (lower) than the trustee's, trust levels decrease (increase). It seems that having performed better than the partner in the tasks affects trust in a negative way, maybe because the trustor anticipates that the trustee will be less willing to send money back. Furthermore, the altruism decision does not depend on the reciprocity decision, but exclusively on the trustor's decision.

From the perspective of the trustee, the contrary effect of inequality direction is found to be significant with respect to the altruism level in Treatments E and H. Specifically, when the initial endowment is higher than that of the trustor, the trustee behaves more altruistically, sending a higher part of his initial endowment. The contrary effect is found in Treatment H, therefore confirming that the effect of deserving the endowment with effort has a positive effect in the intention of being altruistic, thus decreasing the inequality between partners. Somehow comparable is the result of Engel (2011) in his metanalysis of the Dictator Game where he finds that when the dictator has to earn the pie, or the recipient has his own endowment, generosity significantly decreases. Although not in the altruism' decision, we also find a negative effect of effort on the trust levels.

Interestingly, addressing the relationship between trust and reciprocity decisions with some personality archetypes, authors like Espín et al. (2016) and Bellucci et al. (2019) find considerable heterogeneity in the TG decisions as well as in social preferences or motivation. Regarding the four archetypes identified with the EET-pre, it is surprising that although altruist and inequality-lover trustors present positive significant differences, the altruists and the spiteful trustors do not. On the trustees' type, the inequality-averse presents a negative marginal effect, decreasing the probability of taking the egalitarian strategy, contrary to expected. The altruistic archetype does show a positive marginal effect on the reciprocity decision, using more likely the egalitarian strategy.

In summary, the data analysis highlights the importance of past accumulated earnings levels (Treatment H) as well as endowment heterogeneity (Treatment E) on the actual levels of trust, reciprocity and altruism. Specifically, it is observed that the decision of trustors is positively affected by positive past experienced reciprocity. Moreover, trustors are sensitive to how much money the trustee accumulates each round, trusting more the ones that have less compared to themselves. The salient result in Treatment E is that trustors are sensitive to the endowment level of the trustees, trusting more the partners that have got a higher than own endowment, probably considering that a person that performed better in the tasks is a better partner to trust.

A gender analysis of our data, although without significant differences to remark, confirms a result previously found in the literature (Buchan et al., 2008; Dittrich, 2015): women trust in median less than males, although this gender effect vanishes when the endowment is the result of own effort in real tasks, where the significant gender difference is found in the role of the trustee, being females the ones that reciprocate less than males in Treatment E. One could say that the importance that women give to getting the endowment with own effort is stronger if they play the role of trustees rather than the trustor.

Of course, our results go in line with any policy measures that focus on minimizing economic inequality, since its importance goes beyond unexpected limits that affect social welfare.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were refereed and approved by the Laboratorio de Economía Experimental, Universitat Jaume I, Castellón, Spain. The participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

AR-G contributed to the experimental design, programmed the software of the experiment, and coordinated the data analysis and made a first draft of the manuscript. MC-T built the review of the literature and was also involved in the design and in the writing of the final version of the manuscript. AG-G coordinated the experimental design and run the sessions in the lab and was deeply involved in the writing of the final version of the manuscript. All authors were involved in the writing of the final version of the manuscript and approved the submitted version.

FUNDING

AG-G is grateful to the financial support of the Spanish *Ministerio de Ciencia, Innovación y Universidades* (grant RTI2018-096927-B-100), Universitat Jaume I (grant UJI-B2018-76/77), and Generalitat Valenciana (grant AICO/2021/005).

ACKNOWLEDGMENTS

Technical support by Manuel Guerrero Martos during the experimental sessions is gratefully acknowledged.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2021.745948/full#supplementary-material

REFERENCES

Abeler, J., Falk, A., Goette, L., and Huffman, D. (2011). Reference points and effort provision. *Am. Econ. Rev.* 101, 470–492. doi: 10.1257/aer.101.2.470

Alesina, A., and La Ferrara, E. (2002). Who trusts others?. *J. Public Econ.* 85, 207–234. doi: 10.1016/S0047-2727(01)00084-6

Anderson, L. R., Mellor, J. M., and Milyo, J. (2006). Induced heterogeneity in trust experiments. *Exp. Econ.* 9, 223–235. doi: 10.1007/s10683-006-9124-2

Andreoni, J., Nikiforakis, N., and Stoop, J. (2017). *Are the Rich More Selfish Than The Poor, Or Do They Just Have More Money? A Natural Field Experiment*. Cambridge: National Bureau of Economic Research. doi: 10.3386/w23229

Andrighetto, G., Grieco, D., and Tummolini, L. (2015). Perceived legitimacy of normative expectations motivates compliance with social norms when nobody is watching. *Front. Psychol.* 6:1413. doi: 10.3389/fpsyg.2015.01413

Attanasi, G., Battigalli, G., Manzoni, E., and Nagel, R. (2019). Belief-dependent preferences and reputation: experimental analysis of a repeated trust game. *J. Econ. Behav. Organ.* 167, 341–360.

Attanasi, G., Battigalli, G., and Nagel, R. (2013). *Disclosure of Belief-Dependent Preferences in the Trust Game*. Milano: Bocconi University.

Attanasi, G., and Nagel, R. (2008). "A survey of psychological games: theoretical findings and experimental evidence," in *Games, Rationality and Behaviour, Essays in Behavioural Game Theory and Experiments*, Eds A. Innocenti, and P. Sbriglia, (London: Palgrave Macmillan), 204–232.

Balafoutas, L., and Fornwagner, H. (2017). The limits of guilt. *J. Econ. Sci. Assoc.* 3, 137–148.

Bejarano, H., Gillet, J., and Rodriguez-Lara, I. (2021a). Trust and trustworthiness after negative random shocks. *J. Econ. Psychol.* 86:102422

Bejarano, H., Gillet, J., and Rodriguez-Lara, I. (2021b). When the rich do (not) trust the (newly) rich: experimental evidence on the effects of positive random shocks in the trust game. *OSF* [Preprint]. doi: 10.31219/osf.io/wmejt

Bejarano, H., Gillet, J., and Rodriguez-Lara, I. (2018). Do negative random shocks affect trust and trustworthiness? *South. Econ. J.* 85, 563–579. doi: 10.1002/soej.12302

Bellucci, G., Hahn, T., Deshpande, G., and Krueger, F. (2019). Functional connectivity of specific resting-state networks predicts trust and reciprocity in the trust game. *Cogn. Affect. Behav. Neurosci.* 19, 165–176. doi: 10.3758/s13415-018-00654-3

Berg, J., Dickhaut, J., and McCabe, K. (1995). Trust, reciprocity, and social history. *Games Econ. Behav.* 10, 122–142. doi: 10.1006/game.1995.1027

Blanco, M., and Dalton, P. (2019). *Generosity and Wealth: Experimental Evidence from Bogotá Stratification*. Bogotá: Universidad del Rosario.

Bornhorst, F., Ichino, A., Kirchkamp, O., Schlag, K. H., and Winter, E. (2010). Similarities and differences when building trust: the role of cultures. *Exp. Econ.* 13, 260–283.

Brülhart, M., and Usunier, J. C. (2012). Does the trust game measure trust? *Econ. Lett.* 115, 20–23. doi: 10.1016/j.econlet.2011.11.039

Buchan, N. R., Croson, R. T., and Solnick, S. (2008). Trust and gender: an examination of behavior and beliefs in the Investment Game. *J. Econ. Behav. Organ.* 68, 466–476. doi: 10.1016/j.jebo.2007.10.006

Caliendo, M., Fossen, F., and Kritikos, A. S. (2014). Personality characteristics and the decisions to become and stay self-employed. *Small Bus. Econ.* 42, 787–814.

Caliendo, M., Fossen, F. M., and Kritikos, A. S. (2012). Trust, positive reciprocity, and negative reciprocity: do these traits impact entrepreneurial dynamics? *J. Econ. Psychol.* 33, 394–409. doi: 10.1016/j.joep.2011.01.005

Chetty, R., Hofmeyr, A., Kincaid, H., and Monroe, B. (2020). The Trust Game does not (only) measure trust: the risk-trust confound revisited. *J. Behav. Exp. Econ.* 90:101520. doi: 10.1016/j.socec.2020.101520

Ciriolo, E. (2007). Inequity aversion and trustees' reciprocity in the trust game. *Eur. J. Polit. Econ.* 23, 1007–1024. doi: 10.1016/j.ejpoleco.2006.01.001

Coane, J. H., and Umanath, S. (2021). A database of general knowledge question performance in older adults. *Behav. Res. Methods* 53, 415–429. doi: 10.3758/s13428-020-01493-2

Corgnet, B., Espín, A. M., Hernán-González, R., Kujal, P., and Rassenti, S. (2016). To trust, or not to trust: cognitive reflection in trust games. *J. Behav. Exp. Econ.* 64, 20–27. doi: 10.1016/j.socec.2015.09.008

Cox, J. C. (2004). How to identify trust and reciprocity. *Games Econ. Behav.* 46, 260–281. doi: 10.1016/S0899-8256(03)00119-2

Cox, J. C., Kerschbamer, R., and Neururer, D. (2016). What is trustworthiness and what drives it? *Games Econ. Behav.* 98, 197–218. doi: 10.1016/j.geb.2016.05.008

Di Bartolomeo, G., and Papa, S. (2016). Trust and reciprocity: extensions and robustness of triadic design. *Exp. Econ.* 19, 100–115.

Dittrich, M. (2015). Gender differences in trust and reciprocity: evidence from a large-scale experiment with heterogeneous subjects. *Appl. Econ.* 4736, 3825–3838. doi: 10.1080/00036846.2015.1019036

Engel, C. (2011). Dictator games: a meta study. *Exp. Econ.* 14, 583–610.

Espín, A. M., Exadaktylos, F., and Neyse, L. (2016). Heterogeneous motives in the trust game: a tale of two roles. *Front. Psychol.* 7:728. doi: 10.3389/fpsyg.2016.00728

Evans, A. M., and Revelle, W. (2008). Survey and behavioral measurements of interpersonal trust. *J. Res. Pers.* 42, 1585–1593.

Fehr, D. (2018). Is increasing inequality harmful? Experimental evidence. *Games Econ Behav.* 107, 123–134. doi: 10.1016/j.geb.2017.11.001

Fehr, D., Rau, H., Trautmann, S. T., and Xu, Y. (2020). Inequality, fairness and social capital. *Eur. Econ. Rev.* 129:103566. doi: 10.1016/j.euroecorev.2020.103566

Fischbacher, U. (2007). z-Tree: zurich toolbox for ready-made economic experiments. *Exp. Econ.* 10, 171–178. doi: 10.1007/s10683-006-9159-4

Gallego, A. (2016). Inequality and the erosion of trust among the poor: experimental evidence. *Soc. Econ. Rev.* 14, 443–460. doi: 10.1093/ser/mww010

Goldhammer, F., Naumann, J., Stelter, A., Tóth, K., Rölke, H., and Klieme, E. (2014). The time on task effect in reading and problem solving is moderated by task difficulty and skill: insights from a computer-based large-scale assessment. *J. Educ. Psychol.* 106, 608–626. doi: 10.1037/a0034716

Greiner, B. (2015). Subject pool recruitment procedures: organizing experiments with ORSEE. *J. Econ. Sci. Assoc.* 1, 114–125. doi: 10.1007/s40881-015-0004-4

Greiner, B., Ockenfels, A., and Werner, P. (2012). The dynamic interplay of inequality and trust—an experimental study. *J. Econ. Behav. Organ.* 81, 355–365. doi: 10.1016/j.jebo.2011.11.004

Heap, S. P. H., Tan, J. H., and Zizzo, D. J. (2013). Trust, inequality and the market. *Theory Decis.* 74, 311–333. doi: 10.1007/s11238-011-9287-y

Holzmeister, F., and Kerschbamer, R. (2019). oTree: the equality equivalence test. *J. Behav. Exp. Finance* 22, 214–222. doi: 10.1016/j.jbef.2019.04.001

Johnson, N. D., and Mislin, A. A. (2011). Trust games: a meta-analysis. *J. Econ. Psychol.* 32, 865–889. doi: 10.1016/j.joep.2011.05.007

Kerschbamer, R. (2015). The geometry of distributional preferences and a non-parametric identification approach: the equality equivalence test. *Eur. Econ. Rev.* 76, 85–103. doi: 10.1016/j.euroecorev.2015.01.008

Khalmetski, K., Ockenfels, A., and Werner, P. (2015). Surprising gifts: theory and laboratory evidence. *J. Econ. Theory* 159, 163–208.

Lei, V., and Vesely, F. (2010). In-group versus out-group trust: the impact of income inequality. *South. Econ. J.* 76, 1049–1063.

Mohnen, A., Pokorny, K., and Sliwka, D. (2008). Transparency, inequity aversion, and the dynamics of peer pressure in teams: theory and evidence. *J. Labor Econ.* 26, 693–720.

Niederle, M., and Vesterlund, L. (2007). Do women shy away from competition? Do men compete too much? *Q. J. Econ.* 122, 1067–1101. doi: 10.1162/qjec.122.3.1067

Pelligra, V., Reggiani, T., and Zizzo, D. J. (2020). Responding to (un)reasonable requests by an authority. *Theory Decis.* 89, 287–311.

Putnam, R. (2000). *Bowling Alone: The Collapse And Revival Of American Community*. New York: Simon and Schuster paperbacks.

Rodriguez-Lara, I. (2018). No evidence of inequality aversion in the investment game. *PLoS One* 13:e0204392. doi: 10.1371/journal.pone.0204392

Smith, A. (2011a). Identifying in-group and out-group effects in the trust game. *BE J. Econ. Anal. Policy* 11, 1–13. doi: 10.2202/1935-1682.2878

Smith, A. (2011b). Income inequality in the trust game. *Econ. Lett.* 111, 54–56. doi: 10.1016/j.econlet.2011.01.008

Spreng, R. N., McKinnon, M. C., Mar, R. A., and Levine, B. (2009). The Toronto Empathy Questionnaire: scale development and initial validation of a factor-analytic solution to multiple empathy measures. *J. Person. Assess.* 91, 62–71. doi: 10.1080/00223890802484381

Xiao, E., and Bicchieri, C. (2010). When equality trumps reciprocity. *J. Econ. Psychol.* 31, 456–470. doi: 10.1016/j.joep.2010.02.001

Yan, J., and Miao, L. (2007). "Effects of endowments on reciprocal behaviors," in *2007 International Conference on Wireless Communications, Networking and Mobile Computing*, (Manhattan: IEEE), 4591–4594.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychology

ORIGINAL RESEARCH
published: 10 December 2021
doi: 10.3389/fpsyg.2021.728115

Gender Differences in Individual Dishonesty Profiles

Adrián Muñoz García[1], Beatriz Gil-Gómez de Liaño[1,2,3] and David Pascual-Ezama[4]*

[1] Department of Methodology and Social Psychology, Universidad Autónoma de Madrid, Madrid, Spain, [2] Department of Experimental Psychology, Cognitive Processes, and Speech Therapy, Universidad Complutense de Madrid, Madrid, Spain, [3] Center for Biomedical Technology, Universidad Politécnica de Madrid, Madrid, Spain, [4] Accounting and Financial Administration Department, Universidad Complutense de Madrid, Madrid, Spain

Dishonesty has an enormous impact on all aspects of our society. It causes huge financial losses annually, so efforts to understand dishonest behavior have increased. However, one of the main questions yet to be answered is whether dishonesty varies according to gender. Do men behave more dishonestly than women? Although the literature points to a yes, there is still no consensus on the matter. We examined gender differences in dishonesty in a large sample (N = 2,452) using a model recently developed by Pascual-Ezama et al. It is a variation of the classic *die-under-the-cup* task. It enabled us to identify individual dishonesty profiles and look for gender differences between them. The results show that the men were more prone to behave dishonestly than women with small rewards, who seem satisfied without maximizing the potential reward. However, the differences vanished when there was no reward. The men also showed more radical dishonest behavior than the women. The results also suggest that gender differences might be shaped by factors other than gender.

Keywords: dishonesty, gender differences, dishonesty classification, die task, experimental

OPEN ACCESS

Edited by:
Ismael Rodriguez-Lara, University of Granada, Spain

Reviewed by:
Jason Childs, University of Regina, Canada
Niklas Wallmeier, University of Hamburg, Germany
Simon Dato, EBS University of Business and Law, Germany

****Correspondence:***
Adrián Muñoz García
adrian.munnoz.garcia@gmail.com

Specialty section:
This article was submitted to Personality and Social Psychology, a section of the journal Frontiers in Psychology

Received: *20 June 2021*
Accepted: *17 November 2021*
Published: *10 December 2021*

Citation:
Muñoz García A, Gil-Gómez de Liaño B and Pascual-Ezama D (2021) Gender Differences in Individual Dishonesty Profiles. Front. Psychol. 12:728115. doi: 10.3389/fpsyg.2021.728115

INTRODUCTION

Whether we like it or not, dishonesty seems to be inherent in the human condition. Unfortunately, dishonest behavior is a daily occurrence at every level of life: at work, at home, at school, and in various social settings. It is so common that 93% of the 2,624 participants in an extensive poll in 2004 reported different types of daily dishonest behaviors (Kalish, 2004). However, despite the everyday nature of dishonesty and its social acceptance in certain cultures, it has an enormous impact on economies (e.g., Mazar and Ariely, 2006) such that annual losses were once estimated to have reached around $52 billion in workplaces in the United States alone (Weber et al., 2003). It also affects social policy, education, and personal wellbeing (e.g., Christensen and Wright, 2018; Lee et al., 2020). It is therefore not difficult to see why dishonesty research has grown rapidly in recent years. The complex nature of dishonesty, which is sensitive to external and internal factors in human interactions, means that we lack a comprehensive general model of dishonest behavior. Jacobsen et al. (2018) conducted a review of dishonesty research, offering an insightful guide to dishonest behavior. However, although they described some of the major advances in dishonesty research to date, they also raised critical questions that need to be addressed: Why do we cheat? What scenarios elicit dishonesty? Who is more prone to dishonesty? What factors drive dishonesty? It may be suggested that gender is one, although empirical studies are not conclusive. The present

study aimed to shed some light on the matter. Are there gender differences in dishonesty? If so, how are they manifested?

Several pre-1990s studies (e.g., Eisen, 1972) showed that, in general, men seemed to show higher levels of dishonesty than women. Ward and Beck (1990) argued that this difference might have resulted from women's propensity to follow social rules, as the sex-role socialization theory suggested. More importantly, the authors suggested that women also cheated when they were allowed to do so. Using the die-under-the-cup task (wherein participants must roll a die and, depending on the outcome reported, they can gain higher or lower rewards), Fosgaard et al. (2013) observed that the women reached the same level of dishonesty as the men when they were reminded that they could cheat. These results implied that the men were somehow more aware of the chance to lie than women. Indeed, more recent studies have found no differences between males and females in terms of dishonest behavior (e.g., Ezquerra et al., 2018; Siniver, 2021).

Others (Gino et al., 2013) showed that females cheated more than men in certain tasks. Using a math-based task, the authors argued that the women may have cheated more to compensate for the general belief that women perform worse in maths. So, despite the bulk of studies claiming that men are more likely to engage in dishonest behavior (e.g., Capraro, 2017), the findings are contradictory. The evidence thus far suggests that the factors driving gender differences have yet to be elucidated (see Rosenbaum et al., 2014 for a review).

Some factors have already been presented. For instance, as mentioned above, Gino et al. (2013) suggested that the belief that women are worse at maths tasks may have explained why they cheated more. Thus, *perceived competence* seems to be related to the proneness of their dishonest behavior. According to Maggian and Montinari (2017), high-performing competitive females are more likely to be dishonest. Competition has been discussed as a factor mediating gender differences. Schwieren and Weichselbaumer (2010) reported an increase in women's cheating within a competitive setup compared with a non-competitive one, whereas men's remained stable across both. However, this does not mean that men are not also influenced by competition. Nieken and Dato (2016) ran a task in which participants, paired with anonymous peers, only received rewards when they reported better outcomes. The males claimed better outcomes than the women and were thus likely to have cheated more. In that instance, the presence of a direct peer/competitor seemed to make the men cheat more than the women. Muehlheusser et al. (2015) claimed that men in groups were more likely to cheat than females in groups. Interestingly, when decisions had to be made on an individual basis, differences between the genders disappeared, as Muehlheusser et al. (2015) also discovered. Houser et al. (2016) presented evidence that parents were more honest in front of their daughters than in front of their sons. Erat and Gneezy (2012) concluded that women were more likely to tell an altruistic *white lie* (i.e., a lie that benefits the counterpart even if it entails a slight loss to oneself) but were less likely than men to engage in a Pareto white lie (i.e., a lie that benefits both parties). Finally, planning seems to be a factor involved in gender-dishonesty interactions. Chowdhury et al. (2021) argued that men lie more than women when an unexpected opportunity arises to do so.

However, a factor that has not been tested yet is based on the type of dishonest behavior elicited *per se*. Most previous studies analyzed aggregated data, but they did not determine which, and under what conditions, individual subjects were cheating or lying. We took a novel approach in our study. We looked at dishonesty at a personal level to determine the nature of the particular dishonest behavior to compare reported versus real outcomes and thus establish direct comparisons between men and women. Based on the die-under-the-cup task, which was first proposed by Fischbacher and Föllmi-Heusi (2013), and following the Pascual-Ezama et al. (2020) paradigm, we asked participants to roll a virtual die using their mobile devices (cellphones, tablets, or similar). We controlled for gender (see Dufwenberg and Gneezy, 2005 or Shalvi et al., 2011 for similar designs). The die-under-the-cup task involves participants rolling a die in private to earn a reward. The reward depends on the outcome they report; they can deceive either to earn the reward or to increase the outcome reward. To measure dishonesty individually, Pascual-Ezama et al. (2020) proposed a variation of the task that allows the researcher to discover the real distribution of the rolls. We explain the procedure in more detail in the section "Materials and Methods." Using this new approach, Pascual-Ezama et al. (2020) presented a new classification for individual dishonesty profiles. In addition to the lucky individuals who obtained the highest reward by chance, they found other behavioral profiles for those less fortunate. There were two types of honest people: "unlucky honest," who had no reward; and "lucky honest," who had a reward and claimed their winnings from having rolled the die. Excluding honest and lucky people, there were three different types of dishonest participants: the "cheating-non-liars" were those who reported a real-outcome, but cheated rolling the die several times until they reached the desired reward, contrary to the rules (they could only roll the die once); the "liars," who directly lied and claimed a reward they did not deserve when rolling the die; and the "radically dishonest," who did not even roll the die but claimed the maximum reward. Within each of these three categories, some maximized the reward, and others did not. The study aimed to determine whether similar profiles would be found for men and women. If so, how were they distributed within them? Did any potential gender differences change according to the profile? The results showed differences between men and women only within some of them.

EXPERIMENT

Materials and Methods

Participants

To guarantee sufficient analytical power, we decided to run the experiment with a significantly sized sample of more than 2,000 participants (Fox et al., 2009; García-García et al., 2013). The 2,452 individuals (1,286 males and 1,166 females) were recruited by Amazon Mechanical Turk, and they received $1.50 for

turning up and an opportunity to earn up to $0.50 performance-based bonus in the first part of the experiment. One hundred-and-twenty-six participants (76 men and 50 women) did not complete the task correctly (i.e., they did not complete the MTurk process with the MTurk code), so they were eliminated. Another 324 participants (212 men and 112 women) were excluded in accordance with Pascual-Ezama et al.'s (2020) criterion[1]. 29 participants (17 male and 12 female) report less than they obtain. We consider these participants as "incoherent"; the rest did not use the suggested website, so we could not get sufficient information from them. Respect the "incoherent" participants, it could be a mistake when they report, it is possible they do not understand the instructions correctly, or we can suppose any logical reason. Perhaps they have extreme social image concerns, and they believe that someone reporting a five is seen as "most likely dishonest." Then they might consider reporting a 4. In that way, they only give up a small amount of money but gain a lot in the social image dimension. In any case, as we do not have information about the real reason, the number of excluded participants is minimal, and even if they had lied, they would not have done it with dishonest intentions but for self-image. Therefore, we decided to eliminate them. Concerning the rest of excluded participants, we cannot classify them as we do not have enough information. However, we have analyzed the distribution of reported outcomes to rule out a possible selection effect (see section "Results"). The final sample comprised 2,002 participants, of whom 1,004 were women and 998 were men; the average age was 34 (SD = 13) and 36 (SD = 11), respectively.

TABLE 1 | Dishonesty classification.

Behavior	Label	Classification
Roll the die – obtain 5 – report 5	LUCKY	Lucky
Roll the die – obtain 1 to 4 – report the same outcome obtained	HONEST	Lucky honest
Roll the die – obtain 6 – report 6		Unlucky honest
Roll the die – obtain 6 – roll several times until other outcome and report it	CHEATERS NON-LIARS	Sub-maximizing cheaters non-liars
Roll the die – obtain an outcome different than 5 – repeat until 5 – report 5		Maximizing cheaters non-liars
Roll the die – obtain an outcome – report a higher outcome, but less than 5	LIARS	Sub-maximizing liars
Roll the die – obtain an outcome different than 5 – report 5		Maximizing liars
Do not roll the die at all – report < 5	RADICALS	Sub-maximizing radically dishonest
Do not roll the die at all – report 5		Maximizing radically dishonest

Adapted from Pascual-Ezama et al. (2020).

Procedure

The experiment was conducted using the MTurk platform, and the participants were paid after submitting their report[2]. The experiment consisted of an adaptation of the die-under-the-cup task proposed by Fischbacher and Föllmi-Heusi (2013), using the new paradigm proposed by Pascual-Ezama et al. (2020). Participants were asked to roll the die on www.rollandflip.com or a similar website using their cell phone. They can use our suggested website to roll the die[3] or any other website to roll a die, but importantly they use their cellphone, so the perception of no-supervision is high. They would receive no bonus if they rolled a 6, following Fischbacher and Föllmi-Heusi's (2013) rewards system. They could therefore choose not only to be (dis)honest but also to adapt their (dis)honesty according to different levels, from the maximum to minimum rewards (see **Table 1**). Each participant received the same simple and short instructions: "First, ensure you have a smartphone, a tablet, or another electronic device with internet access. You have to roll a die, and you can earn money depending on your roll result: if you roll a 1, you will receive $0.10. If you roll a 2, you will receive $0.20. If you roll a 3, you will receive $0.30. If you roll a 4, you will receive $0.40. If you roll a 5, you will receive $0.50. If you roll a 6, you will receive nothing. Now, please proceed to the following website: https://www.rollandflip.com/ (or another similar site), select the "roll the die" option, and roll the die once." The critical manipulation here was to link the real outcome and the reported one for a given person. We had access to the rollandflip.com database to match the rolls individually, controlling the exact moment every participant performed the task. Although we could consider that deception occurs place since participants maintain the perception of impunity while the researchers are monitoring their behavior, this procedure used by Pascual-Ezama et al. (2020) is essential to classify the different behavioral profiles. Therefore, we could determine the precise number of rolls and the real outcome distribution and link them with the reported ones for each participant. Most of the participants chose to use the rollandflip.com website, allowing us to connect their real and reported outcomes to study honest and dishonest behavior in detail. The website www.rollandflip.com was created by researchers to record real outcomes from rolling a die or flipping a coin. We were able to record the real results, IP address, timestamp, the reported results, and the time the participants took to complete the task. Therefore, we were able to link data from https://rollandflip.com with https://behavioralexperiments.com to classify participants' real behavior.

RESULTS

The most relevant results are presented in the following three subsections. First, we show the typical population-level analysis

[1]Individuals who gave an immediate response (less than 5 s after receiving computerized instructions) without using www.rollandflip.com (see the following section) were classified as radically dishonest because they claimed the high money outcome without flipping a coin or rolling a die on some other website (5 s was not enough time to go to an alternative website and/or initiate a coin toss or a die roll). Individuals who submitted a report that took more than 5 s and also did not go to our chosen website were eliminated from the final sample.

[2]https://behavioralexperiments.com

[3]www.rollandflip.com

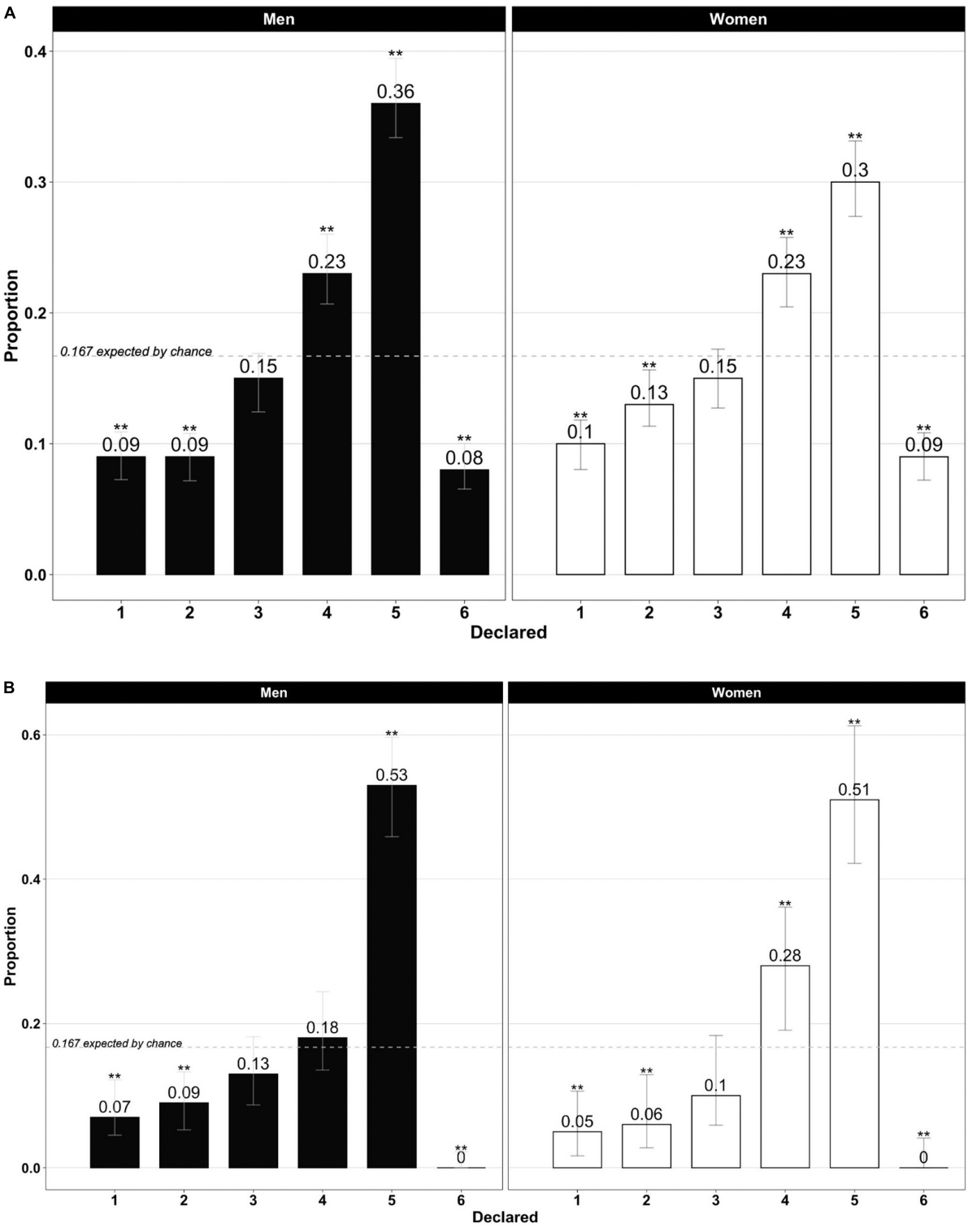

FIGURE 1 | (A) Declared die outcome (men vs. women). Proportion test confidence interval at 95% is represented in gray. Asterisks above the confidence interval mean significant differences between the observed and expected distribution by chance. $^{**}p < 0.01$; $^{*}p < 0.05$. **(B)** Declared die outcome (men vs. women in excluded participants). Proportion test confidence interval at 95% is represented in gray. Asterisks above the confidence interval mean significant differences between the observed and expected distribution by chance. $^{**}p < 0.01$; $^{*}p < 0.05$.

as aggregated data from the reported results as if we did not have the real outcomes to make direct comparisons with previous studies in the field. Then, we show the individual-level analyses comparing reported and real outcomes. Finally, we group the subcategories of the (dis)honest classification into higher categories (by the dichotomy dishonest/honest, nature, and gradient); in each case, the men are compared with the women.

Population-Level Analysis

We examined whether the reported outcome distribution for the males and females differed from a uniform distribution, as is the case in classical inferred tasks aggregated analyses. A Kolmogorov–Smirnov (KS) test for one sample showed that both sample distributions differed significantly from the expected uniform distribution ($p < 0.001$), which indicated that both the men and women did not report the real outcome at the first die-roll; that is, they lied or cheated. Then, we tested for each die outcome to see whether the proportions differed from what would be expected by chance. As we can see in **Figure 1**, high reward proportions were significantly higher than expected by chance ("4" and "5" outcomes; i.e., above the expected 16.7% by chance, as shown in the dashed lines). Low reward proportions were significantly lower than the 16.7% percentage expected by chance ("1" and "2" outcomes). Also, "6-no reward" was significantly lower than expected by chance for both the men and women. The "in-between 3" outcome fell somewhere between 14 and 15% for the men and women; however, in this case, the proportions were not significantly lower than the chance level. The KS test showed only marginally differences between the men and women ($p = 0.08$), which indicated that they cheated almost similarly. The main difference appears when analyzing results for maximizing "5" outcomes, that is, the maximum reward, although both are significantly above the chance level, men maximized the reward more than women (36% vs. 30% in outcome 5; $\chi^2 = 8.37$, $p < 0.01$; see **Figure 1** again). There were also differences for those declaring "2": the women reported significantly more "2s" ($\chi^2 = 9.96$, $p < 0.01$).

Concerning the participants eliminated for not using the proposed website and, therefore, not having information to classify them as honest or dishonest, they maintain a similar distribution (see **Figure 1**) to the rest of the participants, thus ruling out a sample selection effect in terms of distribution. Perhaps we could highlight that they could be a more dishonest sample. They have a higher (and unusually high) number of the maximum prize, any of them report the non-reward output (impossible from a statistical point of view), and one-third of the participants spend less than 30 s, a short time to search for a website to roll the die or to search for a physical die and respond to the experiment. In any case, we do not have enough information to classify them, so eliminating them is the most correct and conservative.

Individual-Level Analysis

The population-level analysis revealed dishonest behavior but did not discriminate between different dishonest profiles. Although we can infer more maximizing cheating among the men than the women (the results from outcome "5"), the aggregated results did not show any gender differences when the general cheating behaviors were compared. Individual-level analyses made it possible to provide a more fine-grained picture of different forms of dishonesty, and this helped us to detect the potential gender differences that had been glimpsed in the population-level results.

First, when we surveyed the participants who were eliminated because they did not follow the rules, we observed a greater number of men; of the 324 excluded participants, 212 were men and 112 were women; $\chi^2 = 19.442$, $p < 0.01$). In reality, they were eliminated conservatively. Beyond having exceeded a certain time and not using the recommended website, we did not know how they behaved. However, there was a high probability that a large proportion were radicals who took longer than the time limit we considered appropriate to classify them as such. This result was therefore logical and supported the finding that the men were more radical than the women.

Second, the real distribution did not differ from the uniform expected distribution, and nor did the proportions (see **Figure 2**). This is important because it shows that the theoretical distribution existed in reality, so we could take the previous deviant declared distribution as proof of a pattern of general dishonesty (see **Supplementary Figure 1**). We did not find any difference between genders in the proportion of real roll. Therefore, statistically speaking, the men and women started from the same conditions.

Once we had checked the statistical assumptions that allowed us to compare men and women, we linked each participant's real outcomes with the reported outcomes following Pascual-Ezama et al. (2020) paradigm. There were no differences in the rate of "lucky" participants by gender ($\chi^2 = 0.05$, $p = 0.83$). There were also no differences between the men and women when there was no reward (i.e., they achieve an outcome of "6." We then calculated the percentage of men and women reporting a different outcome than which they obtained in the first roll of the die (thus, the rate of dishonest individuals divided according to gender). The results showed that 52.2% of the men and 41.0% of the women were dishonest. The statistical analysis of relative risk (RR) revealed that the men were more dishonest than the women; in particular, the men were 1.25 times more likely to be dishonest [RR = 1.24, 95% CI (1.13, 1.35), $\chi^2 = 22.6$, $p < 0.001$].

We calculated the percentages again, but for each profile, as described in Pascual-Ezama et al. (2020), to obtain a more detailed and at the same time broader picture of the nature of the dishonest behavior of the men and women. In **Table 1**, we describe each of those profiles, and in **Figure 3**, we can see the percentage of participants at each profile divided by gender (besides the "lucky," who, as we explained above, were removed from the individual analysis since they did not provide sufficient information for the study).

As **Figure 2** shows, there were two significant results. First, the percentage of "radically dishonest" (i.e., those who did not even roll the die) was higher for the men, both for non-maximizers [RR = 1.83, 95% CI (1.37, 2.44), $\chi^2 = 17.12$, $p < 0.001$] and maximizers [RR = 1.83, 95% CI (1.34, 2.49), $\chi^2 = 14.53$, $p < 0.001$]. That is, regardless of maximizing or not, the men were more "radically dishonest" (see **Figure 2**). Second, for the

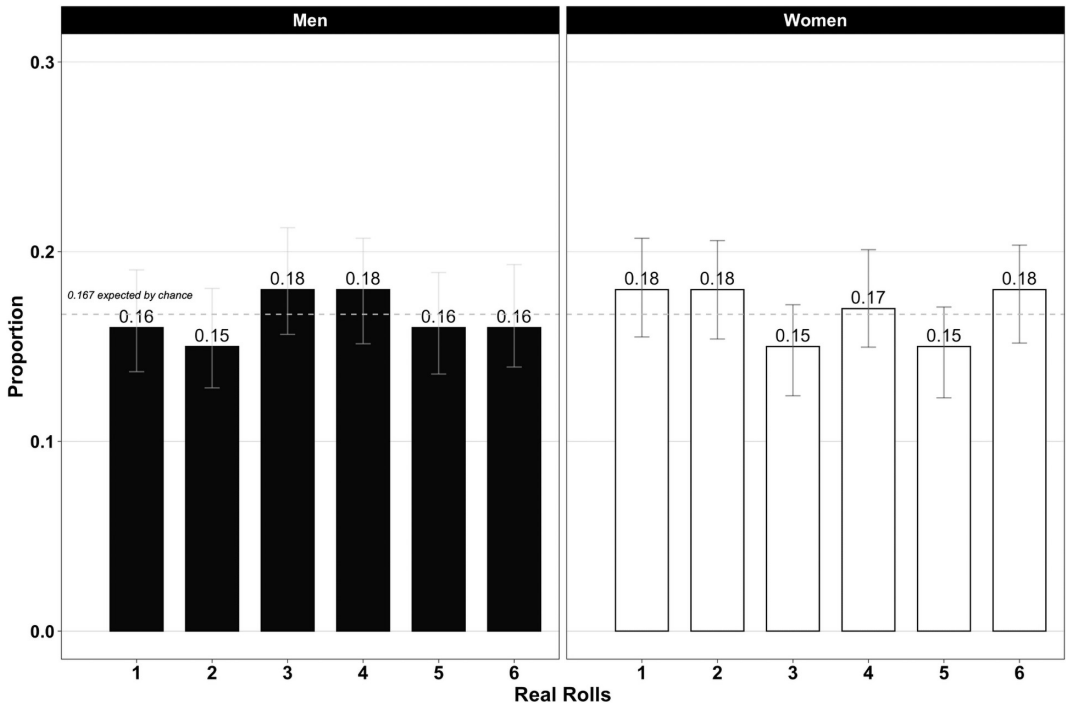

FIGURE 2 | Real die outcome (men vs. women). Proportion test confidence interval at 95% is represented in gray.

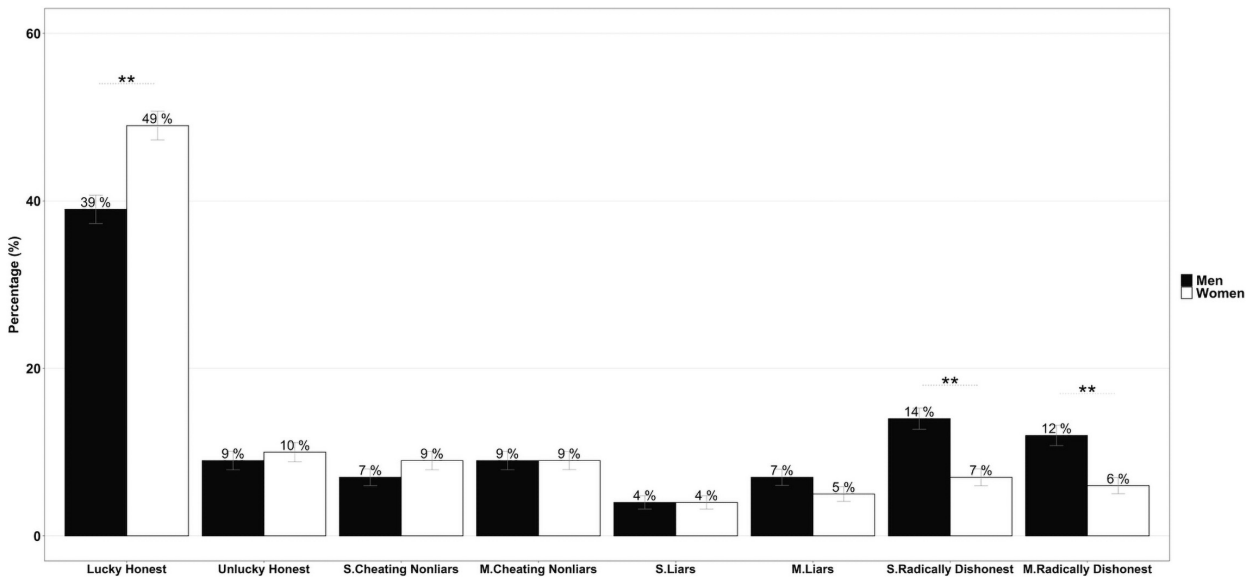

FIGURE 3 | Percentage of men (black) and women (white) for each (dis)honesty profile. Note that lucky people achieving an outcome "5" were excluded from the analysis. In contrast, honest people included here are those achieving an outcome other than "5"; S., sub-maximizing and M., maximizing. The number above bars indicates the percentage; asterisks indicate significance in pairwise proportion comparisons: $^{**}p < 0.01$.

honest people, the differences between the men and women were only apparent among the "lucky honest." There was a significantly higher proportion of "lucky honest" women than men [RR = 1.27, 95% CI (1.14, 1.42), $\chi^2 = 19.42$, $p < 0.001$]. These results suggest that when obtaining a "minimum" reward ("1−4"), the women seemed to be sufficiently satisfied to behave honestly. We hypothesized that the men needed a higher reward to act honestly. To test this, we analyzed the frequency of outcomes (1, 2, 3, or 4) for the "lucky honest" people according to gender (**Figure 4**). Although the outcome "4" was the most

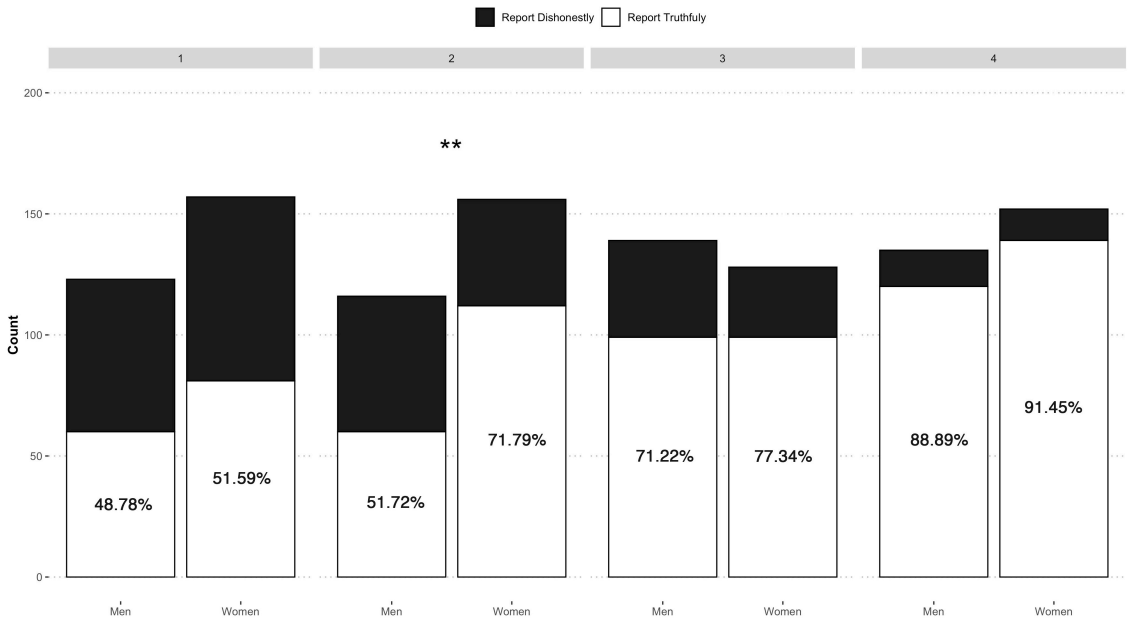

FIGURE 4 | Lucky-honest declared die outcome (men vs. women). Asterisks indicate significance in pairwise proportion comparisons: $^{**}p < 0.01$.

often declared among both groups, the same occurs for the rest of outcomes. Regardless of the number of times the participants (men and women) got each of the outcomes, women honestly reported a higher percentage. The trend is unanimous for the different outcomes and significant for outcome 2 ($\chi^2 = 11.53$, $p < 0.01$). This result could be because as there are more male radicals, more women are throwing the die, and therefore, we could find a more significant number of lucky honest women. However, in this case, we should also find a higher number of women in the other groups, and this is not the case. It could also be the case that the proportion of female rolls is higher in the outcomes with prizes, and therefore more women will accept smaller outcomes. However, as we can see in **Figure 2**, this has not occurred either. Therefore, the results supported our hypothesis that the women were probably more satisfied with smaller rewards.

If this was indeed the case, there should also have been differences between the men and women in the sub-maximizing cheater category. In particular, there should have been a greater proportion of women cheating with outcomes "2" or "3," with men tending to wait for "4" outcomes to cheat more frequently. As **Figure 5** shows, this was the case: there are significantly more cheating men waiting for a "4" outcome (0.68), compared with women (0.48). It should be remembered that in this case, the participants cheated by rolling the die several times until they obtained the desired outcome/reward ($\chi^2 = 5.37$, $p = 0.02$). Although the differences in outcome "3" were not significant ($\chi^2 = 1.31$, $p = 0.25$), the tendency was again more apparent among the men, who were more inclined to be satisfied with a "4" outcome. The women were more frequently satisfied with lower outcomes.

Category-Level Analysis

We merged the dis(honest) labels into three categories, again based on Pascual-Ezama et al. (2020; see **Table 1**). First, as defined in **Figure 3**, the honest people comprised those considered lucky and unlucky. Second, we took into account the nature of the dishonest behavior regardless of the gradient of dishonesty (i.e., whether the behavior was maximized or not). Second, we combine cheater non-liars (i.e., those who rolled the die several times until they obtained the desired value); liars (i.e., those who reported a different outcome to the one obtained from rolling the die); and radicals (i.e., those who did not even roll the die and reported the desired outcome to win the reward). The third group comprised, according to the gradient of dishonesty, those who maximized their dishonest behavior (reporting the outcome "5"), namely, the *maximizers*; and those who decided to report a different outcome from "1" to "4" (which probably fitted the minimum outcome they considered before claiming a reward), namely, the *sub-maximizers*. In **Table 2**, we can see the proportion of women and men who occupied each of those merged profiles, as well as the chi-square tests that illustrated the significant differences between them.

Analyzing differences by profile, we can see that, first, the difference between men and women in terms of the percentage of individuals exhibiting honest or dishonest behavior was significant: the women were more honest. Second, depending on the nature of the dishonest behavior, there were again differences between the men and women (see the third chi-square test in **Table 2**). Still, this only applied to the radicals: the men were 1.83 times more radical than the women [RR = 1.83, 95% CI (1.50, 2.23), $\chi^2 = 36.2$, $p < 0.001$]. No gender differences were apparent in the proportion of cheaters non-liars or liars. Finally, there were

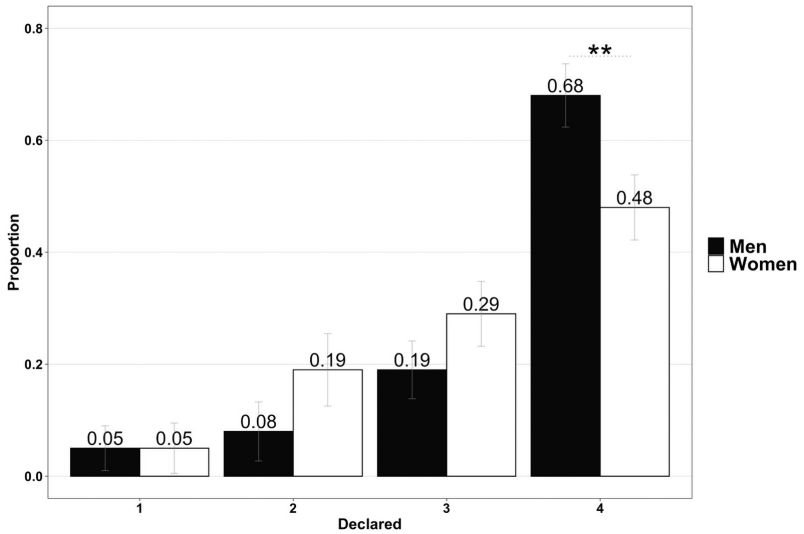

FIGURE 5 | Sub-maximizing cheating non-liars die outcomes obtained after several rolls (men vs. women). Error bars are represented in gray. Asterisks indicate significance in pairwise proportion comparisons: $^{**}p < 0.01$.

TABLE 2 | Proportion of men and women with different (dis)honesty profiles.

		Women	Men	Chi-squared test
Honest		59%	48%	$\chi^2_{1,N=1753} = 22.69,\ p < 0.0001$
Dishonest		41%	52%	
By gradient	Sub-maximizers	50%	47%	$\chi^2_{1,N=814} = 0.55,\ p = 0.46$
	Maximizers	50%	53%	
	Cheaters	44%	31%	$\chi^2_{2,N=814} = 19.22,\ p < 0.0001$
By nature	Liars	22%	20%	
	Radicals	34%	49%	
Total (*n*)		877	876	

Chi-squared test reports independence between honest and dishonest participants; sub-maximizers and maximizers; and liars, cheaters, and radicals with a 95% confidence level.

no differences between men and women regarding the gradient of dishonesty, whether the rewards were maximized or not. While these results were not significant, there were differences in outcomes in the earlier analysis.

CONCLUSION AND DISCUSSION

Although various studies have pointed toward a difference in dishonest behavior between men and women (often showing men as more prone to behave dishonestly), the nature of this difference has not been studied in detail. Certain factors, such as "perceived competence," "individual" versus "grouped" dishonesty, competition, or planning dishonest behavior have revealed a more diverse picture showing—in some cases—no gender disparities and even situations in which women are more dishonest than men. The present study aimed to explore those potential gender differences in dishonesty in more detail by using the Pascual-Ezama et al. (2020) paradigm and ensuring that the participants did not know they were being observed (see Fries et al., 2021). Under this paradigm, we were able to study the nature of different types of dishonest behavior (e.g., cheating, lying, and radically dishonest actions) and its gradient (i.e., maximizing or otherwise). We were also able to depict dishonesty at an aggregated level, as previous studies have done, but more importantly, at an individual level. We examined how the participants behaved by collecting and comparing self- and real reports using the die-under-the-cup online task. The results showed that women were more honest than men in general, but depending on the nature of the dishonest behavior, they could behave similarly or in distinctive ways by graduating their actions.

In particular, we observed that the women were more likely to be honest for lower rewards, while the men needed higher rewards to maintain honest behavior. The women seemed to be satisfied enough at lower rewards, which led them to decide not to cheat for higher ones (see **Figures 2**, **3**). The men tend to maximize rewards, even at non-maximizing levels (that is, achieving outcomes "3" and "4"). Even when cheating, the women tended to be satisfied with lower rewards than the men, indicating that they seemed to be more satisfied even when they were actually cheating (see **Figure 4**). Other studies found that men over-reported higher results than women (e.g., Nieken and Dato, 2016; Grosch and Rau, 2017; Abeler et al., 2019; Benistant et al., 2021). Our results revealed that there were more radically dishonest men than women [according to the Pascual-Ezama et al. (2020) classification], which again supported the idea that the women were more satisfied with lower rewards.

What is also significant is that we replicated Pascual-Ezama et al.'s (2020) dishonesty profiles using a sample of around 2,500 participants. Moreover, we replicated the traditional general finding that men are more dishonest than women, even at the aggregate level. By using Pascual-Ezama et al.'s (2020) new model, we could go beyond individual levels of analysis to observe both real and reported outcomes. Under these circumstances, although differences between men and women were apparent, there were

no differences when there were no rewards: The proportion of the unlucky honest was statistically the same for both the men and women. In other words, the men's levels of honesty did not differ from the women's when they were not winning anything. This result accords with the literature. Our results add a significant nuance to the standard interpretations of differences between men and women regarding dishonesty. Our null results and previous results suggest that gender differences are reliant on the reward factor; differences in rewards reveal differences in gender dishonesty. Although more research is needed, our rewards were small enough to generate results that were different than when higher rewards were available and higher differences obtained between the different outcomes in the die-under-the-cup task. It seems that rewards can modulate gender differences in dishonesty, in that men may be prepared to be more radical in their quest for higher rewards, and women are more satisfied with lower rewards.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Comité de Ética de la Investigación UAM. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

AM and DP-E developed the study concept and performed the testing and data collection. AM drafted the manuscript. BG-G provided critical revisions. All authors contributed to the data analysis, interpretation, and study design, and approved the final version of the manuscript for submission.

FUNDING

This study was made possible thanks to the funding received by the Comunidad Autónoma de Madrid award IND2019/SOC-17283 Ph.D. doctoral research fellowship 2020 to AM.

ACKNOWLEDGMENTS

We thank Aurora Gallego for her useful comments in the junior seminar organized by the rede3c.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2021.728115/full#supplementary-material

Supplementary Figure 1 | Real and declared rolls. Real rolls in the horizontal axis and declared roll in the vertical axis.

Supplementary Figure 2 | Average report for different profiles (Maximizers report always the maximum outcome "5").

REFERENCES

Abeler, J., Nosenzo, D., and Raymond, C. (2019). Preferences for truth-telling. *Econometrica* 87, 1115–1153. doi: 10.3982/ECTA14673

Benistant, J., Galeotti, F., and Villeval, M. C. (2021). *The Distinct Impact of Information and Incentives on Cheating (No. 14014)*. Bonn: Institute of Labor Economics (IZA).

Capraro, V. (2017). Gender differences in lying in sender-receiver games: a meta-analysis. *arXiv* [preprint] arXiv:1703.03739

Chowdhury, S. M., Jeon, J. Y., Kim, C., and Kim, S. H. (2021). Gender differences in repeated dishonest behavior: experimental evidence. *Games* 12:44. doi: 10.3390/g12020044

Christensen, R. K., and Wright, B. E. (2018). Public service motivation and ethical behavior: evidence from three experiments. *J. Behav. Public Adm.* 1, 1–8. doi: 10.30636/jbpa.11.18

Dufwenberg, M., and Gneezy, U. (2005). "Gender and coordination," in *Experimental Business Research*, eds R. Zwick and A. Rapoport (Boston, MA: Springer), 253–262. doi: 10.1007/0-387-24244-9_11

Eisen, M. (1972). Characteristic self-esteem, sex, and resistance to temptation. *J. Pers. Soc. Psychol.* 24, 68–72. doi: 10.1037/h0033387

Erat, S., and Gneezy, U. (2012). White lies. *Manag. Sci.* 58, 723–733. doi: 10.1287/mnsc.1110.1449

Ezquerra, L., Kolev, G. I., and Rodriguez-Lara, I. (2018). Gender differences in cheating: loss vs. gain framing. *Econ. Lett.* 163, 46–49.

Fischbacher, U., and Föllmi-Heusi, F. (2013). Lies in disguise – an experimental study on cheating. *J. Eur. Econ. Assoc.* 11, 525–547. doi: 10.1111/jeea.12014

Fosgaard, T. R., Hansen, L. G., and Piovesan, M. (2013). Separating Will from Grace: an experiment on conformity and awareness in cheating. *J. Econ. Behav. Org.* 93, 279–284. doi: 10.1016/j.jebo.2013.03.027

Fox, N., Hunn, A., and Mathers, N. (2009). *Sampling and Sample Size Calculation*. The National Institutes for Health Research Design Service for the East Midlands / Yorkshire and the Humber. Available online at: https://www.bdct.nhs.uk/wp-content/uploads/2019/04/Sampling-and-Sample-Size-Calculation.pdf (accessed September, 2021).

Fries, T., Gneezy, U., Kajackaite, A., and Parra, D. (2021). Observability and lying. *J. Econ. Behav. Org.* 189, 132–149. doi: 10.1016/j.jebo.2021.06.038

García-García, J. A., Reding-Bernal, A., and López-Alvarenga, J. C. (2013). Cálculo del tamaño de la muestra en investigación en educación médica. *Investig. Educ. Méd.* 2, 217–224. doi: 10.1016/S2007-5057(13)72715-7

Gino, F., Krupka, E. L., and Weber, R. A. (2013). License to cheat: voluntary regulation and ethical behavior. *Manag. Sci.* 59, 2187–2203. doi: 10.1287/mnsc.1120.1699

Grosch, K., and Rau, H. A. (2017). Gender differences in honesty: the role of social value orientation. *J. Economic Psychol.* 62, 258–267. doi: 10.1016/j.joep.2017.07.008

Houser, D., List, J. A., Piovesan, M., Samek, A., and Winter, J. (2016). Dishonesty: from parents to children. *Eur. Econ. Rev.* 82, 242–254. doi: 10.1016/zj.euroecorev.2015.11.003

Jacobsen, C., Fosgaard, T. R., and Pascual-Ezama, D. (2018). Why do we lie? A practical guide to the dishonesty literature. *J. Econ. Surveys* 32, 357–387. doi: 10.1111/joes.12204

Kalish, N. (2004). How honest are you? *Readers Digest* 164, 114–119.

Lee, S. D., Kuncel, N. R., and Gau, J. (2020). Personality, attitude, and demographic correlates of academic dishonesty: a meta-analysis. *Psychol. Bull.* 146, 1042–1058. doi: 10.1037/bul0000300

Maggian, V., and Montinari, N. (2017). The spillover effects of gender quotas on dishonesty. *Econ. Lett.* 159, 33–36. doi: 10.1016/j.econlet.2017.06.045

Mazar, N., and Ariely, D. (2006). Dishonesty in everyday life and its policy implications. *J. Public Policy Mark.* 25, 117–126.

Muehlheusser, G., Roider, A., and Wallmeier, N. (2015). Gender differences in honesty: groups versus individuals. *Econ. Lett.* 128, 25–29. doi: 10.1016/j.econlet.2014.12.019

Nieken, P., and Dato, S. (2016). "Compensation and honesty: gender differences in lying," in *Proceedings of the Beiträge zur Jahrestagung des Vereins für Socialpolitik 2016: Demographischer Wandel - Session: Organizational Design, No. A23-V3*, (Kiel: ZBW – Deutsche Zentralbibliothek für Wirtschaftswissenschaften, Leibniz-Informationszentrum Wirtschaft).

Pascual-Ezama, D., Prelec, D., Muñoz, A., and Gil-Gomez de Liano, B. (2020). Cheaters, liars, or both? A new classification of dishonesty profiles. *Psychol. Sci.* 31, 1097–1106.

Rosenbaum, S. M., Billinger, S., and Stieglitz, N. (2014). Let's be honest: a review of experimental evidence of honesty and truth-telling. *J. Econ. Psychol.* 45, 181–196. doi: 10.1016/j.joep.2014.10.002

Schwieren, C., and Weichselbaumer, D. (2010). Does competition enhance performance or cheating? A laboratory experiment. *J. Econ. Psychol.* 31, 241–253. doi: 10.1016/j.joep.2009.02.005

Shalvi, S., Dana, J., Handgraaf, M. J. J., and De Dreu, C. K. W. (2011). Justified ethicality: observing desired counterfactuals modifies ethical perceptions and behavior. *Org. Behav. Hum. Decis. Process.* 115, 181–190. doi: 10.1016/j.obhdp.2011.02.001

Siniver, E. (2021). Do happy people cheat less? A field experiment on dishonesty. *J. Behav. Exp. Econ.* 91:101658. doi: 10.1016/j.socec.2020.101658

Ward, D. A., and Beck, W. L. (1990). Gender and dishonesty. *J. Soc. Psychol.* 130, 333–339. doi: 10.1080/00224545.1990.9924589

Weber, J., Kurke, L. B., and Pentico, D. W. (2003). Why do employees steal? *Bus. Soc.* 42, 359–380. doi: 10.1177/0007650303257301

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Advantages of publishing in Frontiers

OPEN ACCESS
Articles are free to read for greatest visibility and readership

FAST PUBLICATION
Around 90 days from submission to decision

HIGH QUALITY PEER-REVIEW
Rigorous, collaborative, and constructive peer-review

TRANSPARENT PEER-REVIEW
Editors and reviewers acknowledged by name on published articles

Frontiers
Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org
Contact us: frontiersin.org/about/contact

REPRODUCIBILITY OF RESEARCH
Support open data and methods to enhance research reproducibility

DIGITAL PUBLISHING
Articles designed for optimal readership across devices

FOLLOW US
@frontiersin

IMPACT METRICS
Advanced article metrics track visibility across digital media

EXTENSIVE PROMOTION
Marketing and promotion of impactful research

LOOP RESEARCH NETWORK
Our network increases your article's readership