Universidad del Rosario | Facultad de Economía

SERIE
DOCUMENTOS
DE TRABAJO

No. 307

Julio de 2023

Social exclusion in the lab

Mariana Blanco

Darwin Cortés

Amalia Rodríguez-Valencia

Universidad del Rosario, Facultad de Economía
Cl. 12c #4-59. Bogotá, D. C., Colombia
PBX: (031) 2970200 ext. 4163

facultadeconomia@urosario.edu.co
https://www.urosario.edu.co/Facultad-de-Economia/Inicio/
https://repository.urosario.edu.co/handle/10336/20291

Social exclusion in the lab*

Mariana Blanco[1], Darwin Cortés[†2], and Amalia Rodríguez-Valencia[3]

[1,2]Universidad del Rosario
[3]Universidad EAFIT

July 21, 2023

Abstract

We propose a laboratory experiment to understand how social exclusion affects the participants of this antisocial interaction in terms of performance and reported emotions. We adopt a widely used ostracism manipulation from psychology and take it to an experimental economics laboratory. We find that social exclusion events only affect lightly excluded participants and that this effect is explained by the emotions generated after exclusion. In addition, the victims of exclusion report reductions in the valence dimension of emotions, as do those who have the option to exclude but decide not to. The possibility of bystanders punishing potential offenders, the generalized disapproval of exclusion, and an extended approval of inclusion reduce the incidence of exclusion. This reduction comes at the cost of negative changes in the reported emotions of most participant types, but it does not translate into changes in performance in a task. Last, we find that previous exclusion increases the decision of former victims to ostracize.

1 Introduction

Bullying, broadly defined as a repeated exposition of injuries or discomfort from peers, is a primary threat to school safety and poses several undesirable long-term consequences.

*Thanks to the Alianza EFI-Colombia Científica grant, code 60185 and FP44842-220-2018.
†Corresponding author: *darwin.cortes@urosario.edu.co*

Aggression can happen either directly, in the form of words or physical contact, or indirectly, through ostracism and rumor spreading (Olweus, 1997). Bullying victims are more vulnerable to suffering from mental (depression, anxiety, and alcohol and drug abuse into adulthood) and physical (gastrointestinal distress and headaches) health issues (Lereya et al., 2015; Rivara and Le Menestrel, 2016; Olweus, 1997). Moreover, bullying increases school dropout and affects skill accumulation, and the perception of own exclusion correlates with lower achievement in math, language, and science (Berniell et al., 2016). Victimization also reduces non-cognitive skill accumulation and increases the probability of new bullying events (Sarzosa, 2021). Bullying victims in the lower decile of skill distribution are more likely to smoke, suffer from mental health diseases and depression, and have a lower chance of attending college (Sarzosa and Urzúa, 2021). Since human development and social mobility critically depend on health, cognitive, and non-cognitive skills (Heckman and Mosso, 2014), studying this phenomenon is of particular interest.

We propose a laboratory experiment to address a particular form of bullying in an experimental environment, namely social exclusion or ostracism. In particular, we want to study how social exclusion affects all the parties involved in the negative social interaction (offender, victim, and bystanders) regarding performance on a task and self-reported emotions. In our design, we depart from the study of how individual characteristics, including cognitive and non-cognitive skills, trigger ostracism (Olweus, 1997; Smith and Brain, 2000; Sarzosa and Urzúa, 2021) and focus on how exclusion in a social interaction setting generates changes in behavior. In that regard, we converse with the very sparse economic literature on the effects of bullying. In addition, we engage with the psychology literature on the effects of ostracism. We also explore the effect of bystander intervention through endorsement or punishment of the offenders' actions and the behavior of former victims once given a chance to become offenders.

We follow a series of experiments in psychology that use an ostracism manipulation (Cyberball) proposed by Williams et al. (2000). The original experiment invites participants to play a computerized ball-tossing game to improve their mental visualization skills, where the (usually two) other players are computer generated and controlled. In these experiments, the participants, who are unknowingly ostracized, report lower levels of belonging, self-esteem, control, and meaningful existence (Williams and Jarvis, 2006), even when a despised outgroup member excludes them (Gonsalkorale and Williams, 2007). These results seem to be more salient for adolescents, who, after facing exclusion from the game, report significantly higher anxiety and worse mood than adults in the same situation (Sebastian et al., 2010). Exclusion also activates brain regions associated with physical pain (Eisenberger et al., 2003; Cristofori et al., 2015). These effects are also present for bystanders (Polanin et al., 2012) whose neural pain network activates when seeing a social rejection of a friend during Cyberball (Beeney et al., 2011) (empathic response also in Masten et al. (2010)). In addition, Masten et al. (2011) find that participants with empathy-related brain activity behave more prosocially towards

the victim after observing social exclusion. Riem et al. (2013) propose an adaptation of the Cyberball game to understand bystanders' behavior. In their case, the experimenter is ostracized by the computer, while the experimental subject is the bystander. Here, as well as in similar studies (Vrijhof et al. (2016)), they observe more throws towards the ostracized experimenter as a sign of prosocial compensating behavior.[1]

In our experiment, subjects connected to a virtual session participate in three parts consisting of individual and group stages. In all individual stages, subjects take part in an encryption task (Benndorf et al., 2019)) that serves as our performance measure and answer two questions about the valence and arousal dimensions of their emotions. In group stages, four randomly matched participants interact in a modified version of the ostracism manipulation. We randomly assign them to a role denoted by a letter (A, B, C, or D) that determines the actions they can undertake during the group stages where the ball-tossing game happens, making them potential victims, offenders, or bystanders. In particular, A players (potential offenders) have monetary incentives to exclude B players (potential victims) during some group stages, and C and D players (bystanders) can sometimes rate player A's actions. In addition, we have group stages with role reversal for A and B.

We find that social exclusion events do not affect potential offenders' performance; however, the decision to exclude affects the valence dimension of emotions. In particular, the potential offenders that decide to include potential victims in all rounds report a reduction in valence compared to baseline, while those that decide to exclude at most twice report feeling relatively better. These changes in emotions are consistent with our design feature, where excluding other participants increases the potential offenders' monetary payoffs. On the other hand, we do not observe consistently significant changes in the performance of potential victims of the exclusion. There is an increase in the performance of non-excluded victims and a negative change concerning baseline for those excluded at most twice. The age of the participants explains the first effect; the arousal that appears after exclusion explains the second. Victimization does negatively affect the reported valence of emotions from ostracized participants, with the worst results associated with the highest level of social exclusion. There are no results for bystanders after social exclusion regarding performance or emotions. However, here are some hints that the two bystander types (C and D) are oppositely affected by ostracism, even though they are symmetric in their possibilities in the game.

Once we include the possibility for non-costly punishment, we find that even though the net effect to an offender that excludes a victim from the game and is harshly punished by the bystanders is zero, potential offenders change their behavior in response to the signals they get from the bystanders. Generalized disapproval of exclusion favors changes towards inclusion in the game, while generalized approval of inclusion main-

[1]Regarding bystander participation, Gutierrez et al. (2018) evaluate the effectiveness of a government program aimed at increasing peer participation after bullying events.

tains that behavior. In addition, punishment negatively affects the reported valence dimension of players' emotions in the offender, victim, and the bystander closest to the victim (C), and it positively affects the emotions of D bystanders. The effects on emotions do not translate to the performance in the task for most players. When punishment is below the median, the offenders and the participants in the victim position outperform themselves compared to the baseline. We argue that this lack of expected results may be explained by the income effect that follows from an incentivized task for measuring performance that can compensate for the monetary losses due to social exclusion.

Last, we find that subjects that were excluded when they were in the role of victim tend to exclude more intensely once they are in the role of offender. The emotions that were developed when they were victims of the exclusion and personality traits do not explain this behavior. Ostracism seems to be creating negative reciprocity towards others that are not at fault for the previous exclusion. We also observe that former potential victims that decide to exclude former potential offenders see a reduction in their performance compared to those that do not exclude. However, the results increase if they were excluded as victims. The opposite happens to former potential offenders. These results are in contrast with the absence of effects identified when we analyzed social exclusion behavior before role reversal.

The rest of the paper is organized as follows. In Section 2, we describe the experimental design and procedures. Then, in Section 3 we move to detail our hypotheses. In Section 4, we present the results. In Section 5, we discuss results and breakdown our conclusions.

2 Experimental Design and procedures

We propose an experimental design that involves three parts. All parts are composed of individual and group stages. Subjects participate in a Real Effort Task (RET) during the individual stages (six RET in total). This task, which gives us our performance measure, entails encoding three-letter words into numbers (Benndorf et al., 2019) (see Figure A.1 in the Appendix). In some of these individual stages, we also collect self-reported emotions using the Self-Assessment Manikin techniques from Bradley and Lang (1994) (see the diagrams in Figure A.2 in the Appendix). The emotional state includes valence and arousal measures, which are re-escalated from negative to positive for analysis purposes.[2] Group stages consist of a virtual tossing-the-ball game that follows the ostracism manipulation by Williams et al. (2000), and that lasts for five rounds. Contrary to Williams et al. (2000), participants in our Cyberball game are all connected to the virtual laboratory and interact with three other players also connected to the session. There are five group stages with particular rules that are common

[2]The nine scale figures are re-escalated from -4 to +4.

knowledge for all subjects right before the beginning of each. We also include a sample of questions from the Big Five Inventory (John et al., 1991) adapted for Colombia (Salgado et al., 2016) at the beginning of the experiment.[3] After Part 3, we include an incentivized questionnaire to elicit the social norm of exclusion (Krupka and Weber, 2013) and a sociodemographic survey.

Part 1 begins with an individual stage (RET_0) that we consider a trial stage, followed by the first group stage. In this first group stage, the computer randomly assigns participants to the letter A, B, C, or D. This letter determines the actions a participant can undertake in group stages and remains the same for all group stages in parts 1 and 2. During the first group stage in Part 1, we introduce the Cyberball game and the binding rule of tossing the ball only to the player on the right. This rule applies for the five rounds. A round starts when player A tosses the ball to player B and ends when player D tosses back the ball to player A (see Figure 1). We pay participants 5 Experimental Currency Units (ECUs) to toss the ball and 5 ECUs to receive the ball. This game dynamic yields a payment of 10 ECUs to each participant per round (50 ECUs during the group stage of Part 1). Once the five rounds of the first group stage end, we ask participants to report their emotions ($Emotions_1$) and to take part in a new individual stage (RET_1), which are our baseline measures for emotions and performance, respectively. In both individual stages of Part 1, participants earn 1 ECU for each correct encryption. The sum of the 50 ECUs from the group stage and the individual earnings made with the two RET from Part 1 make up for part of the gains from the activity.

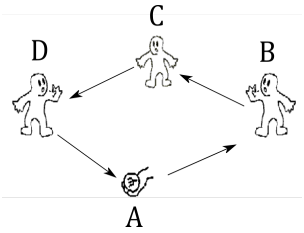


Figure 1: Group Stage from player A's point of view

Notes: The group stages consist of a virtual tossing-the-ball game that lasts for five rounds. A round begins when player A tosses the ball and ends after player D tosses the ball back to player A. Different group stages allow for different possibilities for the players.

Part 2 is composed of two group stages, each followed by an emotions questionnaire ($Emotions_2$ and $Emotions_3$, respectively) and performance tasks (RET_2 and RET_3,

[3]We include two questions per dimension of the Big Five: for extraversion, we include "Is reserved" and "Is outgoing, sociable"; for agreeableness, we ask if "Is generally trusting" and "Tends to find fault with others"; for conscientiousness, we inquire if "Tends to be lazy" and "Does a thorough job"; for neuroticism, we ask if "Is relaxed, handles stress well" and "Gets nervous easily"; and for openness to experience we include "Has few artistic interests" and "Has an active imagination". With the answers to these questions we build an index using a Confirmatory Factor Analysis (CFA).

respectively) (see Figure 2). During the first group stage of Part 2, participants are part of the same group as in Part 1. Here, we introduce a monetary incentive for player A to exclude player B by tossing the ball directly to player C. If player A decides to follow the rule defined in Part 1, she earns the same 10 ECUs per round. However, if she decides to toss the ball directly to player C, she earns 15 instead of 5 ECUs for the toss and 5 ECUs for receiving the ball, for 20 ECUs per round. This game rule makes player A our potential offender in the game, player B the potential victim, and players C and D the (passive) bystanders. Note that if player A excludes player B from the game, player B cannot toss the ball to player C, and thus, she obtains 0 ECUs in that round. During the second group stage from Part 2, the group composition changes. In particular, all B, C, and D players remain part of their former group. However, the computer assigns player A to a new group. To facilitate this reassignment, we create super-groups of eight participants, among which group changes occur. We use these super-groups in all subsequent group stages for changes in group composition.

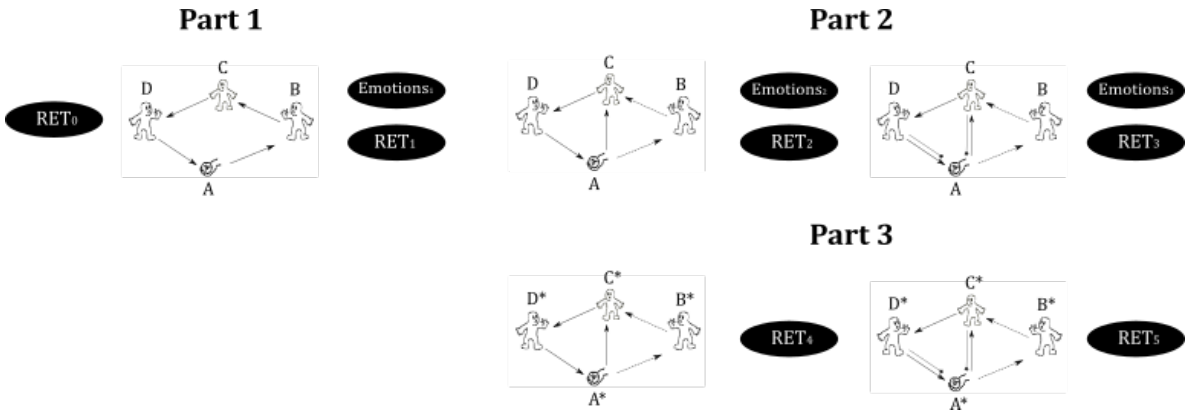


Figure 2: Experimental Design

Notes: Our experiment has three parts consisting of individual and group stages. The experiment begins with a RET that makes up for a trial stage, followed by a group stage that introduces the game dynamic of tossing the ball to the player on the right. After five rounds of a tossing-the-ball game, we collect our baseline measures for performance (RET_1) and emotions ($Emotions_1$). Part two has two group and individual stages. During the first individual stage from Part 2, we create monetary incentives for player A to deviate from the game rule and exclude participant B from the game. After 5 rounds, we collect measures for performance (RET_2) and emotions ($Emotions_2$). The second group stage introduces the possibility of punishment from C and D to A. Once five rounds are done, we collect new measures for performance (RET_3) and emotions ($Emotions_3$). Part 3 follows the same rules and dynamic as Part 2, but there is role reversal: former A players become B players, and vice versa. C and D players may or may not change roles among themselves. After the group stages in Part 3, we collect performance measures (RET_4 and RET_5, respectively).

Once we reassign A players to new groups, we again give player A incentives to exclude player B. During this stage, however, players C and D (the bystanders) must rate player A's actions using a 4-point Likert scale once she makes her decision. These ratings, which range from "Strongly Disagree" to "Strongly Agree," affect player A's

earnings. For each "Strongly Disagree" or "Disagree" rating, player A faces a reduction of 5 and 3 ECUs in her payments, respectively. Analogously, for each "Strongly Agree" or "Agree" rating, player A's earnings increase by 5 and 3 ECUs, respectively. If players C or D rate player A with "Neither Agree nor Disagree," there are no changes in player A's earnings. The ratings become common knowledge to all players before starting the next round. Note that the distance to the social exclusion is different for C and D bystanders. In particular, the C bystander is closest to the potential victim and is the one that receives the ball when player A excludes her from the game. Oppositely, D bystander's game dynamic does not change much with or without social exclusion. These differences can have implications on how much they are affected by victimization.

Part 3 follows the same rules and structure as Part 2. However, there are changes in the roles previously assigned to participants. In particular, former A players become B players, while former B players become A players; we randomly resort C and D players into D or C players. With this manipulation, we can study how former victims of the exclusion behave when they can ostracize a group member and how former offenders react to this exclusion. We use the same super-groups of eight participants defined earlier and rematch them into two new groups once the roles are reassigned. This means that the group configuration changes between parts 2 and 3. In addition, Part 3 does not include self-assessment techniques to elicit changes in emotions. The computer randomly selects Part 2 or Part 3 to determine the final earnings from the activity. The earnings during the individual and group stages from the part selected by the computer are added to the earnings from Part 1 to determine final payoffs.

We performed 13 virtual sessions with students from Universidad del Rosario. Each session consisted of 24[4] participants connected to a Zoom session. We invited participants to an activity that would take up to two hours, including an identity check-up. The experiments were programmed in oTree (Chen et al., 2016), and we recruited subjects using the Online Recruitment System for Economic Experiments (ORSEE) from Rosario Experimental and Behavioral Economics Lab (REBEL). Once in the session, we admitted participants from a waiting room in groups of four and modified their screen names with a code to maintain anonymity. We randomly selected five participants from each session for identity check-ups using Zoom's private rooms. Once the experiment started, We read aloud the General Instructions and the instructions for the first individual and group stages. The rest of the instructions were available on the experimental screens. Two experimenters were available for questions through the Zoom chat. We included inactivity alerts in all rounds from the group stages. We lost three participants due to inactivity and dropped all the other members of their respective supra-groups.

We held a first group of ten sessions between October 14 and November 19, 2021. We collected data in three more sessions on May 2nd and 3rd, 2022. This second group of

[4]Session 12 was carried out with 16 participants.

sessions was necessary to replace the observations we had to drop due to inactivity (see the Results section for details on the final sample). The participants spent 89 minutes in the activity and received an average payment of COP 60.000 (about $15). Once all participants concluded their participation, they received feedback on the Part selected by the computer for payments and their earnings. At the end of the feedback screen, we included a link for a form where we collected information on the bank account where subjects wanted to receive their earnings from the activity.

3 Hypotheses

Our hypotheses follow some of the main findings of the literature on bullying. We study as our primary outcomes a measure that follows from the performance on a RET and self-reported measures of emotions (valence and arousal). We first center on the effects of social exclusion on direct victims and bystanders. We understand exclusion in its intensive margin, meaning the number of ostracism events and some defined categories for the intensity of the exclusion. Following the findings, we expect that:

Hypothesis 1. Social exclusion events will negatively affect the performance in the RET and the self-reported emotions of both victims and bystanders. They will not affect the performance and emotions of offenders.

Although the literature on the effects of bullying on bystanders indicates the existence of pro-social compensating behavior, the closeness between victims and bystanders and the individual prosociality levels of bystanders mediate these results. Then,

Hypothesis 2. The effects of social exclusion in performance and emotions are greater for direct victims of the exclusion than for bystanders.

Our experimental design includes the possibility of punishment through ratings. These ratings can potentially affect the earnings of player A in at most 10 ECUs per round if both C and D players decide on the maximum punishment for player A. A payoff maximizing offender should be indifferent between tossing the ball to player B or player C when she believes the bystanders will harshly punish her excluding decision. In fact, she can increase her expected earnings by excluding player C. However, the presence of ratings can, by itself, discourage exclusion (Fehr and Gächter, 2000):

Hypothesis 3. When both bystanders punish the offender, this reduces exclusion in upcoming rounds.

In addition, the degree of punishment should affect the RET performance and the reported emotions of the directly implicated parties (offender and victims). However, since there are no costs of the punishment to the bystanders, we should not observe

changes in the outcomes of interest.

Hypothesis 4. Punishment negatively affects the RET performance and the reported emotions of offenders and positively affects those of victims. These results remain stable for the bystanders.

Once there is a role reversal in Part 3, we expect that:

Hypothesis 5. Victims who were excluded more frequently in Part 2 and who have the offender's role in Part 3 will exclude their victims more often.

And,

Hypothesis 6. The offenders who are in the victim role in Part 3 will have minor variations in the RET as long as they were part of the top excluders. However, if they excluded below the median and face repeated exclusion, they will have worse results in the RET.

In addition, we want to explore what happens with the RET performance of victims in Part 3, depending on their former behavior as bullies in Part 2.

4 Results

We study the decisions of 278 subjects[5] in five parts. The computer initially assigns subjects to roles A (N=69), B (N=70), C (N=69), and D (N=70) and reassigns them to new roles from part 4 onward, as detailed earlier.[6] We begin our analysis with Table 1, which presents the summary statistics by the subjects' initial roles. We include our main sociodemographic variables, relevant academic information, an index factoring the Big Five test dimensions, the mean ratings for the elicited social norms,[7] the baseline measures from the emotions test (valence and arousal), and the baseline and trial RET performance. Overall, there is balance in these variables across initial roles, as shown by the p-values from the mean comparison tests. There is also no evidence of a difference in the baseline RET or the emotions test by role pairs according to the Mann-Whitney test, which is confirmed with the Kruskal-Wallis test ($\chi^2 = 3.084$ and $p-value = 0.358$

[5] 304 subjects participated in twelve sessions of 24 participants and one of 16. We eliminate 26 subjects: five dropped participants due to inattention, three research assistants that replaced participants who left the session before starting in order to be able to begin on time, and the other members of their super-groups. A dropped participant is replaced by the computer, and decisions are determined at random.

[6] A (N=70), B (N=69), C (N=69), and D (N=70).

[7] To convert the Likert scale responses into numerical scores, we follow Krupka and Weber (2013) and define that: "very socially inappropriate" = –1, "somewhat socially inappropriate" = –1/3, "somewhat socially appropriate" = 1/3, "very socially appropriate" = 1.

for the baseline performance in the RET, and $\chi^2 = 1.709$ and $p - value = 0.635$ for the baseline reported emotions).

Table 1: Summary statistics and mean comparison tests, by role

	Mean				p-value					
	A (N=69)	B (N=70)	C (N=69)	D (N=70)	A vs B	A vs C	A vs D	B vs C	B vs D	C vs D
Socio-Demographic Characteristics:										
Age	21.00	21.27	21.01	21.60	0.476	0.963	0.132	0.493	0.463	0.136
Female	0.55	0.61	0.67	0.63	0.450	0.164	0.354	0.523	0.863	0.641
Number of siblings	1.54	1.30	1.38	1.56	0.173	0.413	0.913	0.673	0.148	0.364
Number of household members	3.83	3.60	3.43	3.36	0.288	0.118	0.033	0.494	0.246	0.753
Relative income at 16	3.32	3.49	3.54	3.41	0.232	0.102	0.451	0.705	0.576	0.309
Above-average income at 16	0.43	0.43	0.46	0.34	0.942	0.734	0.269	0.679	0.300	0.147
Strata	3.61	3.74	3.75	3.59	0.431	0.414	0.892	0.954	0.380	0.366
Self-financed expenditure	31.01	32.57	34.93	34.57	0.779	0.472	0.533	0.660	0.722	0.948
University information:										
Semester	6.91	6.59	6.59	7.07	0.322	0.329	0.615	0.979	0.114	0.115
Economics	0.20	0.10	0.12	0.16	0.091	0.164	0.486	0.764	0.315	0.482
Economics or finance	0.30	0.24	0.28	0.37	0.419	0.710	0.406	0.664	0.099	0.228
Big Five:										
Personality index	0.03	0.04	-0.05	-0.03	0.871	0.152	0.332	0.095	0.238	0.645
Social norms										
No exclusion	0.86	0.84	0.83	0.81	0.765	0.658	0.486	0.869	0.694	0.835
One round of exclusion	0.50	0.45	0.36	0.31	0.485	0.054	0.015	0.274	0.107	0.555
Two rounds of exclusion	0.22	0.20	0.16	0.10	0.809	0.432	0.132	0.563	0.178	0.418
Three rounds of exclusion	-0.29	-0.23	-0.26	-0.32	0.458	0.685	0.609	0.710	0.224	0.358
Four rounds of exclusion	-0.67	-0.60	-0.64	-0.62	0.340	0.706	0.478	0.594	0.804	0.766
Five rounds of exclusion	-0.90	-0.96	-0.90	-0.90	0.240	1.000	0.891	0.221	0.213	0.888
Baseline measures										
Trial Real Effort Task result	7.20	7.59	7.70	6.91	0.254	0.097	0.414	0.747	0.087	0.030
Trial Real Effort Task mistakes	0.29	0.14	0.28	0.13	0.060	0.875	0.037	0.110	0.826	0.074
Baseline Real Effort Task result	8.29	8.70	8.54	8.36	0.121	0.381	0.787	0.595	0.221	0.544
Baseline Real Effort Task mistakes	0.38	0.29	0.23	0.40	0.474	0.251	0.871	0.628	0.379	0.192
Reported emotion	1.99	1.79	1.94	1.71	0.469	0.874	0.320	0.571	0.795	0.404
Reported intensity of emotion	0.32	0.53	0.32	0.30	0.503	1.000	0.948	0.552	0.453	0.954

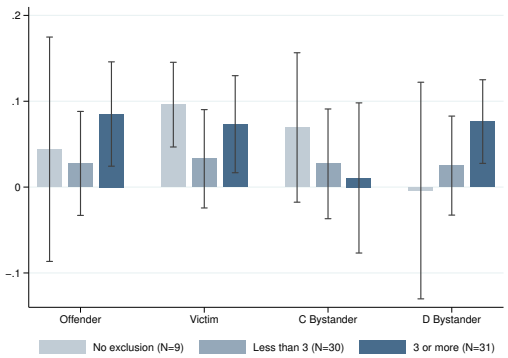
Notes: This table shows the mean by the initial role assigned (A, B, C, and D) for the main final survey variables, the Big Five dimensions, and the baseline measures. It also includes the p-value for the mean differences by role pair. We present results for age, female, number of siblings and household members, relative income at 16, above-average income at 16, and percentage of self-financed expenditures. We also add socioeconomic status following the Colombian classification. We include the current semester and if the participant is an Economics or Finance major. We build an index factoring the five dimensions of the Big Five personality test from a sample of questions. We include the average responses to the elicited norms after converting responses into numerical scores (we follow Krupka and Weber [2013] and define "very socially inappropriate" = –1, "somewhat socially inappropriate" = –1/3, "somewhat socially appropriate" = 1/3, "very socially appropriate" = 1). In addition, we present the baseline results from the encryption task and the number of mistakes, as the reported emotions and intensity of emotions at the beginning of the experiment.

4.1 Social Exclusion

We begin studying the effects of social exclusion events on the performance and emotions of all participants. During the first group stage from Part 2, we create monetary incentives for potential offenders (A players) to exclude B players (potential victims) from the ball-tossing game. From 345 rounds, corresponding to 5 rounds for each of the 69 potential offenders, we find exclusion in 49% of the opportunities. Once we include the possibility of punishment, exclusion reduces to 29.3%. Role reversal does not significantly change this dynamic as exclusion with no punishment is 46.3% and 32% with punishment. Figure 3 shows the average changes in the RET performance between Part 1 (RET_1) and the first group stage from Part 2 (RET_2), normalized by the

performance during the baseline RET (RET_1). With this normalization, we attempt to capture the individual ability in the task. We create three exclusion categories to account for the intensity of the exclusion: no exclusion, exclusion in less than three rounds, and exclusion in three or more rounds. This categorization allows us to find the effects of below-the-median and above-the-median exclusion. In addition, it is consistent with the social norm elicitation results that show that a positive social valuation of exclusion corresponds to, at most, two rounds of exclusion. From three rounds of exclusion onward, there is an average agreement of social inappropriateness. We present results for the potential offenders and victims and both bystanders.

Figure 3: Average changes in normalized RET performance by exclusion category, by role.



Notes: This figure presents the average changes in the normalized RET performance of participants between Part 1 and the first group stage from Part 2. We differentiate results by role for each defined category for exclusion intensity (no exclusion, less than two, and three or more rounds). The number of observations per category is available inside the parentheses. As there is only a possibility for exclusion, these are potential offenders or victims. Confidence intervals at 95% using t-tests.

As can be seen, the normalized changes in performance are not significantly different from zero in most cases. Notably, non-excluded subjects in the potential victim role appear to outperform themselves compared to the baseline (Wilcoxon signed ranks test results are $z = 2.512$ and $p < 0.05$). The results are also significantly better for potential offenders ($z = 3.170$ and $p < 0.01$), potential victims ($z = 2.400$ and $p < 0.05$), and D bystanders ($z = 2.949$ and $p < 0.01$) whenever three or more exclusion events occur during the first group stage from Part 2. However, we do not find significant differences for the social exclusion categories within roles using the Kuskal-Wallis tests. Table 2 presents the OLS results for the normalized changes in performance by role. Our independent variable is the categorical variable previously defined for the exclusion, with no exclusion as the omitted category. Even columns include controls for changes in emotions, sociodemographic variables (age, gender, and Colombian classification for socioeconomic status) and an index of personality traits from the Big Five Inventory. We also include the number of times the computer had to toss the ball during the first group stage from Part 2 to measure inattention.

Table 2: OLS results for within-subject changes in normalized RET performance by exclusion, by role

	$\frac{RET_2 - RET_1}{RET_1}$							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Potential Offender		*Potential Victim*		*C Bystander*		*D Bystander*	
Less than 3 exclusions by part	-0.017	-0.021	-0.063*	-0.061*	-0.042	-0.065	0.029	0.030
	(0.062)	(0.069)	(0.035)	(0.035)	(0.048)	(0.056)	(0.060)	(0.062)
3 or more exclusions by part	0.041	0.029	-0.023	-0.022	-0.059	-0.055	0.080	0.082
	(0.062)	(0.070)	(0.035)	(0.035)	(0.056)	(0.064)	(0.058)	(0.060)
Constant	0.044	0.143	0.096***	-0.098	0.069*	-0.310	-0.004	0.064
	(0.055)	(0.265)	(0.021)	(0.200)	(0.036)	(0.249)	(0.053)	(0.173)
Observations	69	69	70	70	68	68	70	70
Controls	No	Yes	No	Yes	No	Yes	No	Yes
Less than 3 = 3 or more								
F-test	1.862	1.194	1.037	1.019	0.095	0.044	1.909	1.589
Prob > F	0.177	0.279	0.312	0.317	0.759	0.835	0.172	0.212

Notes: OLS regressions by role. Dependent variable is the change in the normalized RET performance in the first group stage from Part 2 compared to the baseline, normalized by the result during the baseline task to account for ability. Reference category is no exclusion. Controls: sociodemographic (age, female, and socioeconomic status), an index for personality traits from the Big Five Inventory, and a measure of inattention. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

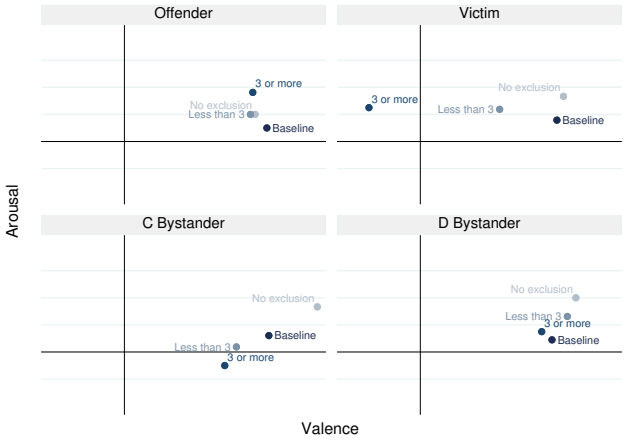
These results align with what we expected from the potential offender. In particular, exclusion events do not appear to affect the performance of the offenders in the RET. We also find that the RET result for potential victims significantly increases whenever they are not excluded, but this result is not robust to controlling for age. In addition, excluded participants in the victim position observe a lesser improvement in performance than non-excluded B players. C and D bystanders appear to have contrasting performances. When we advance in the exclusion categories, the performance of D bystanders increases, while it decreases for C bystanders. These results may hint that the relationship between the two bystanders and the victim is different. However, we cannot confirm statistically that the two bystanders behave differently when facing social exclusion or that, as hypothesized, their performance is negatively affected by victimization.

Result 1: Exclusion events do not affect the performance of potential offenders, while they do affect potential victims. When there are no exclusion events, potential victims outperform themselves compared to their baseline performance measure. In contrast, excluded victims see a reduction in their performance between parts compared to the non-excluded potential victims. The two bystanders appear to have opposing effects in performance when observing social exclusion, but the effects are not statistically significant.

We observe results consistent with our hypotheses in terms of the reported emotions. Figure 4 includes the two dimensions of the reported emotions: valence and arousal. We graph the average of the re-escalated measures from both dimensions that range from -4 to +4. This figure contains the baseline reported emotions and the results after the first group stage from Part 2. Overall, all subjects report positive high arousal emotions

in the baseline. After the exclusion possibility, a Kruskal-Wallis test of equality of populations indicates that we can only reject the null hypothesis for the victims and only in terms of the valence dimension of emotions ($p < 0.01$). As can be seen, the valence dimension of emotions is lower for victims of exclusion, and this reduction is proportional to the number of times the victim is excluded. In particular, the changes in valence for those excluded three or more times concerning baseline are statistically different from the changes for those not excluded (Mann-Whitney test results are $z = 3.897$ and $p < 0.01$) and from the changes in emotions for those excluded once or twice ($z = 3.083$ and $p < 0.01$). The results are also different for participants excluded at most twice from those not excluded ($z = 2.251$ and $p < 0.05$). From the Figure is also apparent that C bystanders who observe exclusion in three or more rounds report lower valence and arousal[8] than at baseline. However, the same is not observable for D bystanders. The arousal levels appear to increase with the levels of exclusion for A participants.[9]

Figure 4: Two dimensions of the reported emotions by intensity of the exclusion, by role



Notes: These figures present the average of the two reported dimensions in the emotion test: valence and arousal. We graph both dimensions by role and by group stage for each of the defined categories for exclusion intensity (no exclusion, less than two, and three or more rounds). The baseline corresponds to Part 1 of the experiment. We also include the results after the first group stage in Part 2 (Exclusion). As there is only a possibility for exclusion, these are potential offenders or victims.

Table 3 presents the OLS results for the two dimensions of the reported emotions by role. The top panel of the Table uses the within-subject difference in the reported valence as the dependent variable; the bottom panel of the table reports the results for

[8]Results for the Wilcoxon signed-rank test gives $z = -2.242$ and $p < 0.05$ for the cases when there is exclusion once or twice and $z = -2.684$ and $p < 0.01$ for the cases where there are three or more exclusions.

[9]A within-subject comparison gives significant results for A participants who exclude three or more times (Wilcoxon signed-rank test gives $z = 4.179$ and $p < 0.01$).

the within-subject difference in the reported arousal. We use the exclusion categories previously defined, with no exclusion as the omitted category. Even columns include controls for sociodemographic variables (age, sex, and strata), personality traits, and a measure for inattention.

Table 3: OLS results for within-subject changes in reported emotions, by role

	$Valence_2 - Valence_1$							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Potential Offender		*Potential Victim*		*C Bystander*		*D Bystander*	
Less than 3 exclusions by part	0.693**	0.642**	-1.178***	-1.182***	-0.069	-0.625	-0.767*	-0.777*
	(0.338)	(0.316)	(0.407)	(0.398)	(0.441)	(0.380)	(0.452)	(0.433)
3 or more exclusions by part	0.233	0.287	-3.057***	-3.098***	-0.613	-0.900*	-0.602	-0.647
	(0.413)	(0.375)	(0.523)	(0.533)	(0.530)	(0.457)	(0.523)	(0.497)
Constant	-0.556*	1.517	0.444	3.502	-0.000	-2.636	0.667	1.552
	(0.283)	(1.597)	(0.283)	(2.520)	(0.394)	(1.607)	(0.425)	(1.450)
Observations	69	69	70	70	69	69	70	70
Controls	No	Yes	No	Yes	No	Yes	No	Yes
Less than 3 = 3 or more								
F-test	1.708	0.951	12.682	13.097	1.792	0.606	0.231	0.147
Prob > F	0.196	0.333	0.001	0.001	0.185	0.439	0.633	0.702
	$Arousal_2 - Arousal_1$							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Potential Offender		*Potential Victim*		*C Bystander*		*D Bystander*	
Less than 3 exclusions by part	-0.307	-0.394	-0.144	-0.136	-0.617	-0.791	0.278	0.312
	(0.481)	(0.545)	(0.369)	(0.412)	(0.530)	(0.562)	(0.368)	(0.427)
3 or more exclusions by part	0.136	0.105	0.308	0.197	-0.993	-0.997	0.036	0.070
	(0.494)	(0.514)	(0.461)	(0.527)	(0.653)	(0.667)	(0.340)	(0.380)
Constant	0.444	3.459	0.111	0.016	0.444	-0.303	0.222	1.186
	(0.397)	(3.589)	(0.251)	(1.572)	(0.511)	(1.698)	(0.142)	(1.331)
Observations	69	69	70	70	69	69	70	70
Controls	No	Yes	No	Yes	No	Yes	No	Yes
Less than 3 = 3 or more								
F-test	1.222	1.737	0.920	0.504	0.764	0.281	0.277	0.297
Prob > F	0.273	0.192	0.341	0.480	0.385	0.598	0.600	0.588

Notes: OLS regressions. Dependent variables: changes in the reported valence and arousal in the first group stage from Part 2 compared to the baseline. Reference category is no exclusion. Controls: sociodemographic (age, female, and socioeconomic status), an index for personality traits from the Big Five Inventory, and a measure of inattention. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

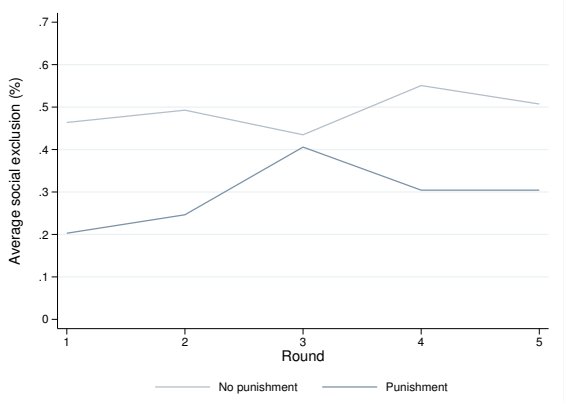
We first observe that subjects in the offender's role that do not exclude B players report a reduction in the valence dimension of emotions. This effect disappears once we control for sociodemographic variables and changes direction when we include controls for age. However, we find consistent evidence that offenders who decide to exclude report higher valence than their baseline measures. This result is statistically significant and higher in magnitude for those who exclude only once or twice during the first group stage from Part 2. On the other hand, and as expected, victims report feeling worse in the presence of exclusion. The negative changes in valence are deepened as we move up in the exclusion categories. Last, we see that both C and D bystanders are negatively affected in terms of valence when there is social exclusion. There are no significant changes in the arousal dimension of emotions for any of the players.

Result 2: Exclusion increases the reported valence for participants in the role of offender who decide to exclude, and decreases the reported valence of excluded players and of both bystanders.

4.2 Punishment

We now move to analyze the punishment behavior of our subjects. On average, our offenders exclude in 2.45 of the five rounds whenever there is a possibility of exclusion without punishment. Once we include the possibility of punishment, the average exclusion reduces to 1.46 rounds. Using a Kuskal-Wallis test we reject the null hypothesis of equality of population between the exclusion behavior of potential offenders under the punishment and the no punishment conditions ($p < 0.01$).[10] Figure 5 shows the average social exclusion behavior of A players by round. As can be seen, ostracism is mostly stable in the no-punishment condition, varying between 43.5% and 55.1%. In the punishment condition, social exclusion drops to 20% in round 1, compared to 46% in the no-punishment condition. This may indicate that the threat of punishment disciplines potential offenders. The exclusion then rises to around 25% in round 2 and over 40% in round 3. It then stabilizes at 30%. This variability is most likely responding to the punishment decisions of the bystanders.

Figure 5: Average social exclusion with and without punishment possibility, by round



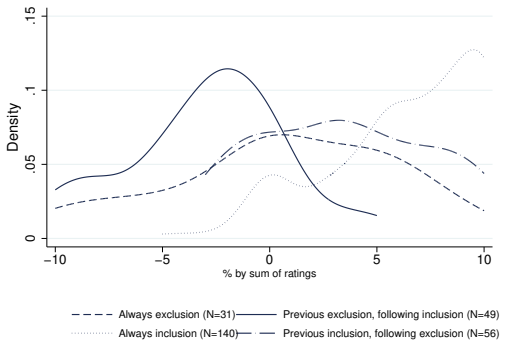
Notes: This figure presents the average social exclusion behavior of potential offenders in Parts 2 and 3 (no punishment and punishment conditions, respectively), by round.

In an initial exploration of bystander behavior when there is a possibility of punishment, we find that when there is social exclusion in a round, the sum of the opinions of both bystanders is negative and significant, and this is robust to the usual controls and role

[10] There is a significant difference in the exclusion behavior with and without punishment ($z = 4.128$ and $p < 0.01$) using a Wilcoxon signed-rank test.

reversal (see Table A.1 in the Appendix). In addition, social exclusion decisions appear to be influenced by punishment. Figure 6 includes the density for the social exclusion decisions of potential offenders, depending on the sum of ratings from bystanders in the immediately previous round. The continuous line illustrates that changes in social exclusion behavior happen mostly when the sum of ratings after an exclusion event is negative. In contrast, generalized approval or indifference (nor approval nor disapproval) of exclusion events favors a continuation of social exclusion (dashed line). When bystanders do not forcefully reward the inclusion in a particular round, offenders tend to exclude in the following round (dash-dots line); instead, when bystanders reward inclusion, it is perpetuated in the immediately following round (dotted line).

Figure 6: Potential offender's behavior depending on previous ratings



Notes: This figure includes the density for the social exclusion decisions of offenders in t and t+1, in terms of the sum of ratings from both bystanders in t.

Next, we estimate a logit regression for the social exclusion at t+1 depending on the feedback from bystanders at t. Table 4 presents the results for the estimations. Columns (1) and (2) regress on the sum of opinions of bystanders, and columns (3) and (4) use the number of bystanders who punish the offender (0, 1, or 2), with no bystanders punishing the offenders as the omitted category. In addition, we create a variable that combines the ratings from bystanders after the exclusion decisions of potential offenders. We first identify if a social exclusion event is generally approved or disapproved by both bystanders using the sum of their ratings in a given round. If the sum of ratings is positive, we say that bystanders approve of potential offenders' actions; if it is negative, we say that bystanders disapprove of potential offenders' decisions. If the sum of ratings is zero, we say that, on average, bystanders are indifferent to potential offenders' social exclusion behavior. This classification gives us six possible categories: social exclusion is punished (15.36%), social exclusion is rewarded (5.80%), indifference to social exclusion (8.12%), inclusion (no social exclusion) is punished (4.06%), inclusion is rewarded (57.10%), and indifference to inclusion (9.57%). Columns (5) and (6) of Table 4 regress social exclusion at t+1 on these categories, with indifference to inclusion as the omitted category. We include controls for sociodemographic characteristics, personality traits,

and a measure for inattention on the even columns. Standard errors are clustered at the super-group level.

Table 4: Logit results for exclusion behavior after feedback

	Social exclusion in t+1					
	(1)	(2)	(3)	(4)	(5)	(6)
Social exclusion in t	-0.289	-0.411	2.143***	2.141***		
	(0.419)	(0.419)	(0.516)	(0.496)		
Sum of opinions of bystanders in t	-0.194***	-0.209***				
	(0.053)	(0.054)				
Social exclusion in t × Sum of opinions	0.376***	0.390***				
	(0.107)	(0.111)				
One bystander punishes offender in t			2.114***	2.182***		
			(0.497)	(0.518)		
Two bystanders punishes offender in t			-1.727*	-1.772*		
			(1.026)	(1.029)		
Social exclusion in t × One bystander punishes			-4.072***	-4.179***		
			(0.805)	(0.843)		
Exclusion is punished					-0.799	-1.041*
					(0.563)	(0.624)
Indifferent to exclusion					0.330	0.239
					(0.551)	(0.658)
Exclusion is rewarded					1.709**	1.484**
					(0.731)	(0.658)
Inclusion is punished					2.035**	1.843*
					(1.012)	(1.012)
Inclusion is rewarded					-0.654	-0.836*
					(0.438)	(0.470)
Constant	0.053	1.784	-1.227***	-0.231	-0.531	1.334
	(0.343)	(1.782)	(0.189)	(1.631)	(0.402)	(1.726)
Observations	276	276	276	276	276	276
Controls	No	Yes	No	Yes	No	Yes
Clusters	35	35	35	35	35	35

Notes: Logit regressions. Dependent variable: social exclusion in t+1. Reference category is no exclusion of B players in columns (1) and (2), no exclusion and no punishment from bystanders in columns (3) and (4), and indifference to social exclusion in columns (5) and (6). Controls: sociodemographic (age, female, socioeconomic status), personality traits from the Big Five Inventory, and a measure of inattention. Standard errors clustered at the super-group level. *** p<0.01, ** p<0.05, * p<0.1.

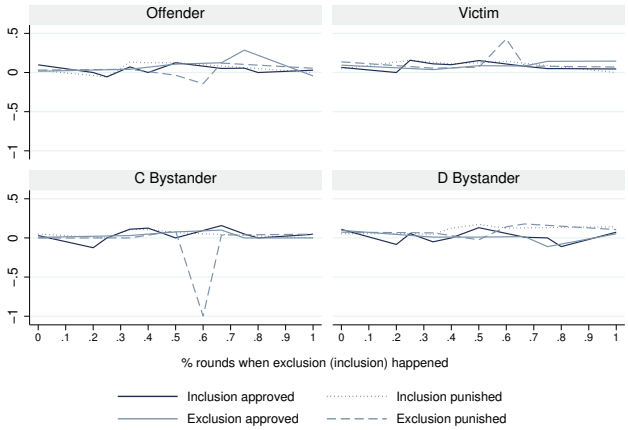
Here we see that if there is exclusion at t and bystanders reward the offender, the probability of exclusion in the following round increases. Once we analyze the results for the number of times bystanders punish a potential offender in period t, we find that if there is no exclusion at t, and if one of the bystanders punishes the offender compared to none, this increases the probability of exclusion in t+1. In the presence of social exclusion at t, and with a single punishment decision, the probability of exclusion reduces in the following round. The results once we include the opinion of bystanders to potential offenders' actions indicate that if bystanders punish ostracism, social exclusion reduces in the following round, but that when bystanders are either indifferent or reward social exclusion, or when they punish inclusion, the exclusion increases the following round.

Result 6: Ratings from bystanders affect potential offenders' decisions. In particular,

generalized disapproval of exclusion favors changes towards inclusion in the game, while generalized approval of inclusion maintains that behavior.

Since our performance and emotions outcomes are available for the part and not the round, we compute the number of times social exclusion (inclusion) is approved (disapproved) relative to the number of times social exclusion (inclusion) occurs during that group stage. Figure 7 shows the average changes in the normalized RET performance of subjects after the second group stage of Part 2, compared to the baseline results. On the x-axis, we have the percentage of rounds when social exclusion (inclusion) happened. We can observe that when ostracism is approved in more than 70% of the rounds where it happens, the RET performance of offenders appears to increase. In contrast, when it is punished in around 60% of rounds, the performance of offenders decreases. Oppositely, victims who observe punishment in around 60% of the exclusion events outperform themselves regarding baseline measures. When bystanders punish either decision of potential offenders, D players appear to have an increase in their performance. The same is not true for C bystanders.

Figure 7: Average changes in normalized RET by the reactions to social exclusion decisions, by role.



Notes: These figures present the average changes in the normalized RET performance of participants (y-axis). On the x-axis we include the percentage of rounds where either exclusion or inclusion happened. We present the by role and by the combination of potential offender actions and bystanders' reactions. As there is only a possibility for exclusion, these are potential offenders or victims.

A regression analysis further explores these findings by considering the interaction of social exclusion and punishment. Table 5 presents the OLS results for the normalized changes in RET performance between Part 1 and the second group stage of Part 2 for the potential offenders and the potential victims (analogous results for the bystanders are available in Table A.2 in the Appendix). In columns (1), (2), (5), and (6), we present the results using the number of times potential offenders get punished during the

group stage as the independent variable and the categories of social exclusion previously defined. The rest of the columns use categories for punishment that use the median as reference (no punishment, punished once or twice, and punished more than twice, with $p_{50} = 2$). The even columns include the usual controls and control for the changes in the reported emotions between the second group stage from Part 2 and the baseline. The reference categories are no exclusion and no punishment.

Table 5: OLS results for within-subject changes in RET performance by punishment, for potential offenders and victims

	$\frac{RET_3 - RET_1}{RET_1}$							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Potential Offender				*Potential Victim*			
Number of punishment events by part	0.017	-0.032			0.040***	0.086**		
	(0.031)	(0.041)			(0.013)	(0.038)		
Less than 3 exclusions by part	-0.048	-0.013	-0.307***	-0.210***	0.089	0.093*	0.147***	0.121
	(0.083)	(0.076)	(0.101)	(0.074)	(0.053)	(0.049)	(0.051)	(0.087)
3 or more exclusions by part	0.143**	0.151**	-0.403***	-0.188	-0.016	-0.076	0.158*	0.170
	(0.061)	(0.073)	(0.141)	(0.132)	(0.091)	(0.117)	(0.084)	(0.111)
Punishment events × Less than 3 exclusions	0.017	0.037			-0.054**	-0.101**		
	(0.048)	(0.048)			(0.021)	(0.045)		
Punishment events × 3 or more exclusions	-0.052	0.003			-0.022	-0.048		
	(0.034)	(0.047)			(0.030)	(0.045)		
Punished once or twice			0.005	-0.027			0.065***	0.153***
			(0.062)	(0.066)			(0.022)	(0.046)
Punished more than twice			0.375***	0.211*			-0.106*	-0.072
			(0.123)	(0.117)			(0.059)	(0.094)
Punished once or twice × Less than 3 exclusions			0.330**	0.253**			-0.152**	-0.203*
			(0.122)	(0.107)			(0.060)	(0.102)
Punished once or twice × 3 or more exclusions			0.495***	0.302**			-0.187**	-0.323**
			(0.151)	(0.137)			(0.092)	(0.141)
Constant	0.055	0.759*	0.057	0.727*	0.041*	0.026	0.041*	0.053
	(0.037)	(0.375)	(0.038)	(0.369)	(0.020)	(0.171)	(0.021)	(0.201)
Observations	69	69	69	69	70	70	70	70
Clusters	35	35	35	35	35	35	35	35
Controls	No	Yes	No	Yes	No	Yes	No	Yes

Notes: OLS regressions. Dependent variable is the change in the RET performance on the second group stage from Part 2 compared to the baseline, normalized by the result during the baseline task to account for ability. Reference categories in columns (3), (4), (7), and (8) are no exclusion and no punishment. Controls: changes in emotions (between the second group stage of Part 2 and the baseline), sociodemographic variables (age, female, number of siblings, number of household members, socioeconomic status), personality traits from the Big Five Inventory, economics or finance major, and a measure of inattention. Standard errors clustered at the super-group level. *** p<0.01, ** p<0.05, * p<0.1.

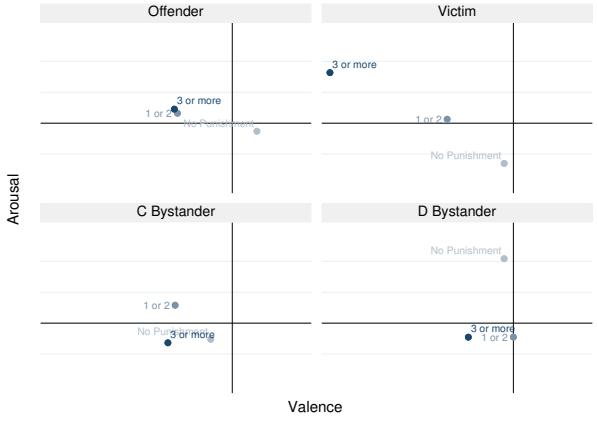
When the number of punishment events increases, the RET performance of potential victims increases with respect to the baseline. At the same time, it decreases for potential offenders, although this last result is not statistically significant. When we include the categories for punishment, we see that when offenders decide to exclude and are lightly punished (once or twice), they outperform themselves in terms of the RET ($p = 0.054$). The effect is larger for those that decide to exclude in three or more rounds ($p < 0.05$). The opposite happens to victims of exclusion that observe few punishment decisions of bystanders, but the effect is not significant. There are no important changes to be reported of bystanders.

Result 7: The number of times a potential offender is punished positively affects the RET performance of potential victims.

Result 8: When punishment to social exclusion is moderate (once or twice per group stage), the RET performance of offenders increases.

Figure 8 presents the changes in the two dimensions of the reported emotions from Part 1 to the second group stage in Part 2, using the punishment categories defined above. It appears that punishment negatively affects the emotions of all players. In particular, the valence of punished potential offenders and of potential victims who observe punishment reduces. The changes in the valence reported by potential victims is larger when punishment happens three or more times. C bystanders report lower valence of emotions compared to baseline measures, even in cases when there is no punishment. D bystanders, however, only report feeling relatively worse than during baseline when the punishment happens three or more times during the third group stage. In terms of arousal, the victims who observe above median punishment report intenser emotions, as do D bystanders in cases of no punishment.

Figure 8: Changes in the two dimensions of the reported emotions by punishment, by role



Notes: These figures present the changes between Part 1 (baseline) and the second group stage from Part 2 (punishment) for the two reported dimensions in the emotion test. We graph both dimensions by role for each of the defined categories for exclusion intensity (no exclusion, one or two, and three or more rounds). As there is only a possibility for exclusion, these are potential offenders or victims.

Now we move on to analyze the results in the reported emotions after the punishment possibility. Table 6 presents the results for within-subjects changes in the reported emotions for potential offenders and potential victims, interacting categories for punishment and exclusion. Columns (1) to (4) present the results for the changes in valence; columns (5) to (8) present the results for the changes in arousal. We include the usual

controls in the even columns. Here we can observe that heavy punishment, compared to no punishment, decreases the valence dimension of emotions for potential offenders. The changes are larger and significant for those punished above the median. The results are not significant for the interaction of bystanders' punishment decisions with offenders' exclusion behavior. There are no general changes in the reported arousal levels of potential offenders after punishment. Potential victims reported valence is also negatively affected by above the median punishment, compared to no punishment to potential offenders. However, the reported arousal increases when potential offenders are lightly punished compared to the no-punishment situation.

Result 9: Above the median punishment negatively affects valence of potential offenders and potential victims compared to the no punishment situation.

Table 6: OLS results for within-subject changes in reported emotions by punishment, for potential offenders and victims

	$Valence_3 - Valence_1$				$Arousal_3 - Arousal_1$			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Potential Offender								
Number of punishment events by part	-0.375	-0.392			0.000	-0.287		
	(0.304)	(0.328)			(0.625)	(0.847)		
Less than 3 exclusions by part	-0.532	-0.564	0.684	1.411*	0.286	-0.028	-0.947	-1.584**
	(0.662)	(0.806)	(0.784)	(0.791)	(0.631)	(0.647)	(0.843)	(0.649)
3 or more exclusions by part	-0.125	0.588	2.384**	3.213**	1.306*	1.538*	1.086	0.585
	(0.748)	(0.924)	(1.030)	(1.224)	(0.771)	(0.875)	(1.409)	(1.184)
Punishment events × Less than 3 exclusions	-0.003	0.128			-0.193	0.151		
	(0.409)	(0.414)			(0.671)	(0.846)		
Punishment events × 3 or more exclusions	0.308	0.211			-0.106	0.128		
	(0.336)	(0.380)			(0.646)	(0.887)		
Punished once or twice			-0.816*	-0.765			0.553	0.262
			(0.448)	(0.472)			(1.192)	(1.521)
Punished more than twice			-2.700***	-3.207***			0.300	0.807
			(0.942)	(1.010)			(1.131)	(0.944)
Punished once or twice × Less than 3 exclusions			-1.093	-1.687*			0.675	1.397
			(1.005)	(0.845)			(1.391)	(1.562)
Punished once or twice × 3 or more exclusions			-2.084*	-2.641*			-1.386	-0.567
			(1.223)	(1.351)			(1.866)	(1.882)
Constant	0.292	0.706	0.316	1.521	-0.000	1.170	-0.053	0.436
	(0.221)	(3.087)	(0.224)	(2.974)	(0.379)	(3.566)	(0.383)	(3.303)
Observations	69	69	69	69	69	69	69	69
Clusters	35	35	35	35	35	35	35	35
Controls	No	Yes	No	Yes	No	Yes	No	Yes
Potential Victim								
Number of punishment events by part	0.500**	0.006			1.406	1.542		
	(0.191)	(0.542)			(1.090)	(1.028)		
Less than 3 exclusions by part	-0.704	-0.618	-1.658	-1.598**	1.672**	1.550*	2.605**	3.167**
	(0.665)	(0.567)	(1.167)	(0.768)	(0.697)	(0.866)	(1.232)	(1.321)
3 or more exclusions by part	-2.168**	-2.091	-0.658	-0.785	0.093	-0.115	3.505**	3.793**
	(1.006)	(1.295)	(1.466)	(1.119)	(1.081)	(1.356)	(1.505)	(1.496)
Punishment events × Less than 3 exclusions	-1.017***	-0.509			-1.582	-1.696		
	(0.330)	(0.583)			(1.109)	(1.051)		
Punishment events × 3 or more exclusions	-0.537**	-0.095			-1.014	-1.125		
	(0.246)	(0.568)			(1.126)	(1.067)		
Punished once or twice			0.842***	-0.058			3.105*	3.095**
			(0.306)	(0.851)			(1.586)	(1.383)
Punished more than twice			-1.700	-1.584*			-1.200	-1.651
			(1.465)	(0.913)			(1.336)	(1.329)
Punished once or twice × Less than 3 exclusions			-0.251	0.543			-4.423**	-5.262***
			(1.233)	(1.004)			(1.982)	(1.812)
Punished once or twice × 3 or more exclusions			-2.342	-1.687			-6.505***	-6.992***
			(1.529)	(1.286)			(2.372)	(2.206)
Constant	0.167	1.557	0.158	2.492	-1.010*	-2.995	-1.105**	-3.207
	(0.300)	(2.186)	(0.306)	(2.204)	(0.497)	(2.639)	(0.499)	(2.741)
Observations	70	70	70	70	70	70	70	70
Clusters	35	35	35	35	35	35	35	35
Controls	No	Yes	No	Yes	No	Yes	No	Yes

Notes: OLS regressions. Dependent variables are the changes in both dimensions of the reported emotions after the second group stage from Part 2 compared to the baseline. Reference categories in columns (3), (4), (7), and (8) are no exclusion and no punishment. Controls: sociodemographic (age, female, number of siblings, number of household members, socioeconomic status), personality traits from the Big Five Inventory, economics or finance major, and a measure of inattention. Standard errors clustered at the super-group level. *** p<0.01, ** p<0.05, * p<0.1.

Table 7 presents the results for the same specifications from Table 6 but now considering both bystanders separately. The results for the reported emotions of the two bystander types go in the opposite direction. In particular, it seems that punishment negatively affects the valence dimension of emotions of C bystanders. Oppositely, punishment positively affects the reported valence of D bystanders.

Table 7: OLS results for within-subject changes in reported emotions by punishment, for bystanders

	$Valence_3 - Valence_1$				$Arousal_3 - Arousal_1$			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
C Bystander								
Number of punishment events by part	-0.938***	-0.607			-0.719	-0.380		
	(0.319)	(0.458)			(1.443)	(1.355)		
Less than 3 exclusions by part	-0.020	-0.202	0.421	-0.128	0.308	0.533	-0.237	-0.486
	(0.538)	(0.646)	(0.428)	(0.812)	(0.700)	(0.768)	(0.587)	(0.757)
3 or more exclusions by part	-0.798	-1.282	0.543	0.046	1.241	1.267	0.985	0.698
	(0.759)	(1.108)	(1.206)	(1.586)	(1.421)	(1.646)	(0.962)	(1.290)
Punishment events × Less than 3 exclusions	0.847**	0.489			0.552	0.118		
	(0.358)	(0.499)			(1.462)	(1.382)		
Punishment events × 3 or more exclusions	1.055***	0.780			0.486	0.142		
	(0.383)	(0.548)			(1.472)	(1.394)		
Punished once or twice			-1.579***	-1.221*			0.263	0.627
			(0.428)	(0.622)			(3.049)	(2.746)
Punished more than twice			-0.900*	-0.585			-0.500	-0.264
			(0.493)	(0.829)			(0.685)	(0.752)
Punished once or twice × Less than 3 exclusions			0.988**	0.957			0.464	0.405
			(0.471)	(0.846)			(3.101)	(2.775)
Punished once or twice × 3 or more exclusions			0.457	0.092			-0.385	-0.401
			(1.306)	(1.645)			(3.366)	(3.113)
Constant	-0.437	0.567	-0.421	0.533	-0.135	1.808	-0.263	1.164
	(0.421)	(3.192)	(0.428)	(3.268)	(0.461)	(3.580)	(0.451)	(3.272)
Observations	69	69	69	69	69	69	69	69
Clusters	35	35	35	35	35	35	35	35
Controls	No	Yes	No	Yes	No	Yes	No	Yes
D Bystander								
Number of punishment events by part	1.094***	1.039***			0.594	0.776**		
	(0.263)	(0.339)			(0.512)	(0.374)		
Less than 3 exclusions by part	-0.422	-0.389	-2.658	-2.370	-1.873**	-1.904**	-2.316**	-2.550**
	(0.650)	(0.668)	(1.897)	(1.758)	(0.769)	(0.899)	(0.894)	(1.212)
3 or more exclusions by part	-0.006	0.041	-1.058	-1.242	-2.577**	-2.017	-2.316	-2.161
	(0.786)	(0.745)	(1.971)	(1.827)	(1.145)	(1.368)	(1.431)	(1.658)
Punishment events × Less than 3 exclusions	-1.315***	-1.341***			-0.523	-0.756		
	(0.310)	(0.376)			(0.607)	(0.509)		
Punishment events × 3 or more exclusions	-1.123***	-1.210***			-0.358	-0.639		
	(0.310)	(0.399)			(0.546)	(0.464)		
Punished once or twice			1.842***	1.757***			1.184	1.350**
			(0.274)	(0.402)			(0.738)	(0.511)
Punished more than twice			1.000	0.431			0.800	0.978
			(1.800)	(1.628)			(1.170)	(1.494)
Punished once or twice × Less than 3 exclusions			0.476	0.135			-0.593	-0.690
			(1.915)	(1.817)			(1.114)	(1.281)
Punished once or twice × 3 or more exclusions			-0.942	-0.464			-0.984	-1.252
			(2.034)	(1.896)			(1.724)	(1.697)
Constant	0.177	2.273	0.158	1.077	1.344***	0.808	1.316**	0.385
	(0.269)	(2.137)	(0.274)	(2.262)	(0.481)	(2.986)	(0.485)	(3.047)
Observations	70	70	70	70	70	70	70	70
Clusters	35	35	35	35	35	35	35	35
Controls	No	Yes	No	Yes	No	Yes	No	Yes

Notes: OLS regressions. Dependent variables are the changes in both dimensions of the reported emotions after the second group stage from Part 2 compared to the baseline. Reference categories in columns (3), (4), (7), and (8) are no exclusion and no punishment. Controls: sociodemographic (age, female, number of siblings, number of household members, socioeconomic status), personality traits from the Big Five Inventory, economics or finance major, and a measure of inattention. Standard errors clustered at the super-group level. *** p<0.01, ** p<0.05, * p<0.1.

Result 10: The position C bystander occupies concerning potential offender and potential victim seems to be generating negative emotions, while the position D bystander occupies leads them to report positive emotions.

4.3 Role Reversal

Now we turn to study the behavior of former potential victims and potential offenders once there is role reversal. We begin our analysis by exploring the behavior of potential offenders depending on how they were treated as victims in terms of exclusion. Table 8 presents the results for the OLS estimation of the number of times a former victim, now in the role of potential offender, decides to exclude participants in role B, depending on the number of times she was excluded in Part 2. Column (1) presents the results for the estimation without including controls. From Column (2) onward, we include sociodemographic variables, if the subject is an economics or finance major, and a measure of intention as controls. In Column (3), we add the changes in the reported emotions when they had the role of victims after the possibility of exclusion, compared to the baseline measures. Here we want to explore if the change of emotions we created when subjects were in the victim position explains this result. In Column (4) we include the personality index.

Table 8: Exclusion behavior of former potential victims in the role of potential offender

	Number of exclusion events in Part 3			
	(1)	(2)	(3)	(4)
Number of times excluded as potential victim	0.359***	0.413***	0.368***	0.382***
	(0.108)	(0.117)	(0.116)	(0.112)
$Valence_2 - Valence_1$			-0.078	-0.063
			(0.093)	(0.088)
$Arousal_2 - Arousal_1$			-0.110	-0.072
			(0.097)	(0.088)
Personality index				-0.987
				(0.726)
Constant	1.442***	0.971	0.958	0.817
	(0.359)	(1.472)	(1.605)	(1.414)
Observations	70	70	70	70
Clusters	35	35	35	35
Controls	No	Yes	Yes	Yes

Notes: OLS regressions. Dependent variable is number of times a potential offender excludes in Part 3. Controls: sociodemographic (age, female, number of siblings, number of household members, socioeconomic status), economics or finance major, and a measure of inattention. Standard errors clustered at the super-group level. *** p<0.01, ** p<0.05, * p<0.1.

Although it is clear in the instructions that potential victims are not the necessarily the same players that excluded the now potential offenders, we observe that the social exclusion decision of former potential victims is positively and significantly associated with the number of times they were excluded from the game in Part 2. Positive changes in valence and arousal when they were in the potential victim role and after observing the possibility for social exclusion decrease the number of times a potential offender decides to exclude, but these results are not statistically significant. Neither is our measure for personality traits.

Result 11: Former victimization increases retaliation against someone that is not necessarily at fault for former exclusion. This result does not appear to be driven by the emotions generated after being a victim of exclusion or personality traits.

Last, Table 9 presents the results for participants in the roles of potential offenders and victims in Part 3. These participants had the role of potential victims and potential offenders during parts 1 and 2, respectively. Our dependent variable is the within change in the RET performance after the first group stage from Part 3 and the baseline, normalized with the baseline performance to account for individual ability. We include as independent variables the intensity of exclusion in both parts in levels and using categories (with no exclusion as the omitted category). Odd columns include controls. We see that participants in the role of offender that decide to exclude see a reduction in their RET performance, compared to those that decide to not exclude at all. The changes in RET are higher for those that heavily exclude B players than for those that exclude once or twice. The results improve for those that decide to exclude three or more times when they are offenders, and that were excluded as victims.

Result 12: Former victims that decide to exclude once they are in the offender position observe a negative change in their performance compared to those that do not exclude. However, the performance increases for those that decide to intensely exclude that were also excluded as victims.

Table 9: OLS results for within-subject changes in RET, role reversal

	$\frac{RET_4 - RET_1}{RET_1}$							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Potential Offender				*Potential Victim*			
Number of exclusion events in Part 3	0.008	-0.004			0.036	0.044		
	(0.018)	(0.019)			(0.024)	(0.028)		
Number of exclusion events in Part 2	0.004	0.018			0.033	0.038		
	(0.016)	(0.018)			(0.026)	(0.031)		
Exclusion events Part 3 × Exclusion events Part 2	-0.006	-0.004			-0.011	-0.014		
	(0.006)	(0.006)			(0.008)	(0.010)		
Less than 3 exclusions in Part 3			-0.011	-0.026			-0.174	-0.184
			(0.102)	(0.099)			(0.119)	(0.148)
3 or more exclusions in Part 3			-0.154***	-0.217**			0.271***	0.314**
			(0.041)	(0.082)			(0.074)	(0.137)
Less than 3 exclusions in Part 2			-0.105	-0.119			-0.101	-0.067
			(0.078)	(0.072)			(0.081)	(0.104)
3 or more exclusions in Part 2			-0.154***	-0.121**			0.037	0.104
			(0.041)	(0.059)			(0.074)	(0.135)
Less than 3 exclusions in Part 3 × Less than 3 exclusions in Part 2			0.034	0.081			0.313**	0.283
			(0.132)	(0.129)			(0.147)	(0.176)
Less than 3 exclusions in Part 3 × 3 or more exclusions in Part 2			0.108	0.168			0.225*	0.171
			(0.107)	(0.105)			(0.133)	(0.187)
3 or more exclusions in Part 3 × Less than 3 exclusions in Part 2			0.221**	0.290***			-0.162*	-0.229
			(0.088)	(0.104)			(0.095)	(0.164)
3 or more exclusions in Part 3 × 3 or more exclusions in Part 2			0.207***	0.277***			-0.264**	-0.363***
			(0.058)	(0.080)			(0.105)	(0.116)
Constant	0.102***	0.033	0.154***	0.100	-0.010	0.546	0.063	0.567
	(0.037)	(0.214)	(0.041)	(0.223)	(0.067)	(0.493)	(0.074)	(0.544)
Observations	70	70	70	70	69	69	69	69
Clusters	35	35	35	35	35	35	35	35
Controls	No	Yes	No	Yes	No	Yes	No	Yes

Notes: OLS regressions. Dependent variable is the change in the RET performance on the first group stage from Part 3 compared to the baseline, normalized by the result during the baseline task to account for ability. Reference categories in columns (3), (4), (7), and (8) are no exclusion in Part 2 and no exclusion in Part 3. Controls: sociodemographic (age, female, number of siblings, number of household members, socioeconomic status), changes in emotions after possibility of exclusion in Part 2 with respect to baseline, personality traits from the Big Five Inventory, economics or finance major, and a measure of inattention. Standard errors clustered at the super-group level. *** p<0.01, ** p<0.05, * p<0.1.

Participants in the victim role who first decided as potential offenders outperform themselves concerning baseline measures whenever they are heavily excluded (more than twice), regardless of how much they excluded when they were in the offender position.

Result 13: Former offenders that are heavily excluded do not see their performance affected and outperform themselves.

5 Discussion and conclusion

We propose a laboratory experiment to study social exclusion. This particular manifestation of peer aggression is difficult to identify and discourage. Thus, understanding its dynamics can help inform policies to reduce the prevalence of this phenomenon and the negative consequences it generates. We are interested in studying the performance and emotional effects of exclusion. In addition, we want to see the capacity of bystanders to reduce the incidence of this phenomenon and the negative consequences it has on the direct victims of the exclusion.

We find that social exclusion events do not generally affect the performance of the victims or the bystanders as we expected. The older non-excluded victims and the heavily excluded participants outperform themselves compared to baseline, but this last result is not statistically significant. Using an incentivized task to measure performance may be causing subjects excluded in three or more opportunities to increase their effort during the task to try to compensate for the losses from social exclusion. Maybe only those for whom the psychological costs are large enough are willing to incur the costly effort required to increase productivity. Lightly excluded participants have a smaller performance change than non-excluded B players. The emotions capture the effect that the exclusion generates in them. The results in terms of the reported emotions are more consistent with what we expected. Victims of the exclusion and participants in the position of an offender that decides to include B players in all rounds report feeling significantly worse than baseline. This last result for victims is consistent with the literature that uses an ostracism manipulation, even though the incentives to exclude participants are common knowledge to all. Oppositely, offenders in three or more rounds report feeling relatively better than those who did not exclude.

Although we do not find results on emotions and performance for bystanders as they do in the psychology literature, we observe that bystanders negatively rate exclusion from offenders. This behavior can be seen as empathy-related, even when punishment is not costly for them. In addition, participants in the offender position change their conduct in response to the signals sent by the bystanders. As a result, social punishment effectively reduces the incidence of exclusion. Another highlight of punishment is that victims outperform themselves compared to baseline. These findings hint at

the importance of bystanders actively intervening against the exclusion. However, even when there are no monetary costs of punishment for most participants, forcing peers to intervene appears to be causing an emotional toll on potential offenders and victims and the bystanders closer to the victims. On the other hand, the bystanders farther from the victims report feeling relatively better than baseline. Although both bystander types are statistically balanced in terms of observables, they are impacted differently by the game dynamics. This difference in results may have to do with the closer distance between victims and C bystanders and with the fact that they are the ones receiving the ball whenever offenders decide on exclusion.

Last, we explore the effects of being a former victim regarding future victimization behavior and changes in performance. We find that the possibility of excluding when they were previously ostracized increases the RET performance of former victims. This result contrasts with the lack of effect from victimization on the RET performance of the original offenders. We also find that the higher the number of times a person was excluded as a victim, the higher the number of times she decides to exclude other participants. We observe negative reciprocity even when experiment instructions mention that they were not interacting with the person who was their potential offender in a previous game stage. This dynamic may suggest how these behaviors become socially prevalent and socially accepted exclusion rules arise.

References

Beeney, Joseph E, Robert G Franklin Jr, Kenneth N Levy, and Reginald B Adams Jr, "I feel your pain: emotional closeness modulates neural responses to empathically experienced rejection," *Social Neuroscience*, 2011, *6* (4), 369–376.

Benndorf, Volker, Holger A Rau, and Christian Sölch, "Minimizing learning in repeated real-effort tasks," *Journal of Behavioral and Experimental Finance*, 2019, *22*, 239–248.

Berniell, Lucila, Dolores de la Mata, Raquel Bernal, Adriana Camacho, Felipe Barrera-Osorio, Fernando Álvarez, Pablo Brassiolo, and Juan Vargas, "RED 2016. Más habilidades para el trabajo y la vida: los aportes de la familia, la escuela, el entorno y el mundo laboral," 2016.

Bradley, Margaret M and Peter J Lang, "Measuring emotion: the self-assessment manikin and the semantic differential," *Journal of behavior therapy and experimental psychiatry*, 1994, *25* (1), 49–59.

Chen, Daniel L, Martin Schonger, and Chris Wickens, "oTree—An open-source platform for laboratory, online, and field experiments," *Journal of Behavioral and Experimental Finance*, 2016, *9*, 88–97.

Cristofori, Irene, Sylvain Harquel, Jean Isnard, François Mauguière, and Angela Sirigu, "Monetary reward suppresses anterior insula activity during social pain," *Social cognitive and affective neuroscience*, 2015, *10* (12), 1668–1676.

Eisenberger, Naomi I, Matthew D Lieberman, and Kipling D Williams, "Does rejection hurt? An fMRI study of social exclusion," *Science*, 2003, *302* (5643), 290–292.

Fehr, Ernst and Simon Gächter, "Cooperation and punishment in public goods experiments," *American Economic Review*, 2000, *90* (4), 980–994.

Gonsalkorale, Karen and Kipling D Williams, "The KKK won't let me play: Ostracism even by a despised outgroup hurts," *European Journal of Social Psychology*, 2007, *37* (6), 1176–1186.

Gutierrez, Italo A, Oswaldo Molina, and Hugo Ñopo, "Stand against bullying: An experimental school intervention," 2018.

Heckman, James J and Stefano Mosso, "The economics of human development and social mobility," *Annu. Rev. Econ.*, 2014, *6* (1), 689–733.

John, Oliver P, Eileen M Donahue, and Robert L Kentle, "Big five inventory," *Journal of Personality and Social Psychology*, 1991.

Krupka, Erin L and Roberto A Weber, "Identifying social norms using coordination games: Why does dictator game sharing vary?," *Journal of the European Economic Association*, 2013, *11* (3), 495–524.

Lereya, Suzet Tanya, William E Copeland, E Jane Costello, and Dieter Wolke, "Adult mental health consequences of peer bullying and maltreatment in childhood: two cohorts in two countries," *The Lancet Psychiatry*, 2015, *2* (6), 524–531.

Masten, Carrie L, Naomi I Eisenberger, Jennifer H Pfeifer, and Mirella Dapretto, "Witnessing peer rejection during early adolescence: Neural correlates of empathy for experiences of social exclusion," *Social neuroscience*, 2010, *5* (5-6), 496–507.

_ , Sylvia A Morelli, and Naomi I Eisenberger, "An fMRI investigation of empathy for 'social pain'and subsequent prosocial behavior," *Neuroimage*, 2011, *55* (1), 381–388.

Olweus, Dan, "Bully/victim problems in school: Facts and intervention," *European journal of psychology of education*, 1997, *12* (4), 495–510.

Polanin, Joshua R, Dorothy L Espelage, and Therese D Pigott, "A meta-analysis of school-based bullying prevention programs' effects on bystander intervention behavior," *School Psychology Review*, 2012, *41* (1), 47–65.

Riem, Madelon ME, Marian J Bakermans-Kranenburg, Renske Huffmeijer, and Marinus H van IJzendoorn, "Does intranasal oxytocin promote prosocial behavior to an excluded fellow player? A randomized-controlled trial with Cyberball," *Psychoneuroendocrinology*, 2013, *38* (8), 1418–1425.

Rivara, Frederick and Suzanne Le Menestrel, *Preventing bullying through science, policy, and practice*, National Academies Press, 2016.

Salgado, Elvira, Elvia Vargas-Trujillo, Jana Schmutzler, and Eduardo Wills-Herrera, "Uso del Inventario de los Cinco Grandes en una muestra colombiana," *Avances en Psicología Latinoamericana*, 2016, *34* (2), 365–382.

Sarzosa, Miguel, "Victimization and Skill Accumulation: The Case of School Bullying," *Journal of Human Resources*, 2021, pp. 0819–10371R2.

_ **and Sergio Urzúa**, "Bullying among adolescents: The role of skills," *Quantitative Economics*, 2021, *12* (3), 945–980.

Sebastian, Catherine, Essi Viding, Kipling D Williams, and Sarah-Jayne Blakemore, "Social brain development and the affective consequences of ostracism in adolescence," *Brain and cognition*, 2010, *72* (1), 134–145.

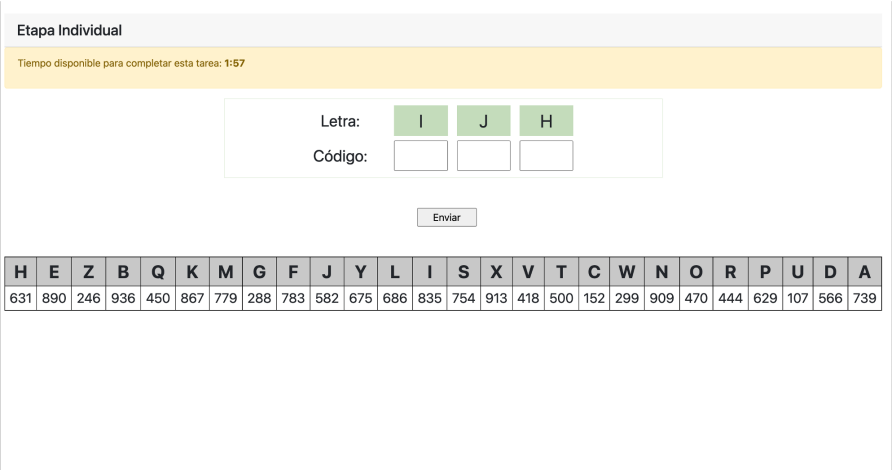
Smith, Peter K and Paul Brain, "Bullying in schools: Lessons from two decades of research," *Aggressive Behavior: Official Journal of the International Society for Research on Aggression*, 2000, *26* (1), 1–9.

Vrijhof, Claudia I, Bianca G van den Bulk, Sandy Overgaauw, Gert-Jan Lelieveld, Rutger CME Engels, and Marinus H van IJzendoorn, "The prosocial cyberball game: compensating for social exclusion and its associations with empathic concern and bullying in adolescents," *Journal of adolescence*, 2016, *52*, 27–36.

Williams, Kipling D and Blair Jarvis, "Cyberball: A program for use in research on interpersonal ostracism and acceptance," *Behavior research methods*, 2006, *38* (1), 174–180.

_ **, Christopher KT Cheung, and Wilma Choi**, "Cyberostracism: effects of being ignored over the Internet.," *Journal of personality and social psychology*, 2000, *79* (5), 748.

A Appendix



H	E	Z	B	Q	K	M	G	F	J	Y	L	I	S	X	V	T	C	W	N	O	R	P	U	D	A
631	890	246	936	450	867	779	288	783	582	675	686	835	754	913	418	500	152	299	909	470	444	629	107	566	739

Figure A.1: Individual stage RET screen

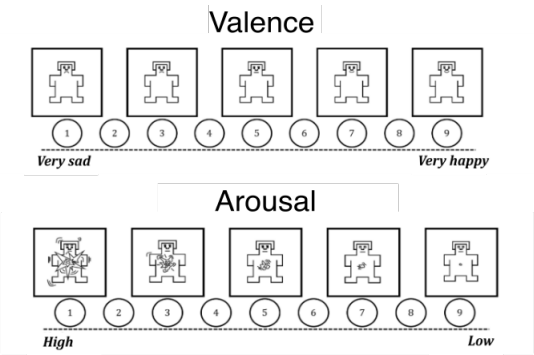


Figure A.2: Images from the Emotion Test

Table A.1: Punishment behavior of bystanders, with and without role reversal

	Sum of opinions of C and D by round		Sum of opinions of C and D by round (RR)	
	(1)	(2)	(3)	(4)
Social exclusion in the round	-7.049***	-7.086***	-5.943***	-5.963***
	(0.801)	(0.833)	(0.959)	(0.948)
Age		0.217**		0.00101
		(0.0935)		(0.180)
Female participant		-0.316		-0.0307
		(0.491)		(0.677)
Socioeconomic strata		0.327		0.427*
		(0.207)		(0.235)
Personality, using BF		-0.614		-1.552*
		(0.940)		(0.858)
Econ and finance students		-0.642		
		(0.474)		
Inattention		-0.709		2.826
		(0.940)		(2.492)
Constant	5.492***	0.105	3.858***	2.195
	(0.431)	(2.328)	(0.534)	(3.747)

Notes: OLS regressions. Dependent variable is the sum of opinions of both bystanders, by round. Standard errors clustered at the super-group level in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table A.2: OLS results for within-subject changes in RET performance by punishment, for bystanders

	$\frac{RET_3 - RET_1}{RET_1}$							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	C Bystander				*D Bystander*			
Number of punishment events by part	-0.014	-0.033			0.010	0.034		
	(0.045)	(0.059)			(0.044)	(0.033)		
Less than 3 exclusions by part	-0.049	-0.010	-0.099	-0.059	0.062	0.083	0.072*	0.118
	(0.080)	(0.095)	(0.061)	(0.100)	(0.069)	(0.068)	(0.041)	(0.098)
3 or more exclusions by part	-0.056	0.037	-0.213*	-0.139	-0.011	-0.004	0.108	0.157
	(0.073)	(0.083)	(0.105)	(0.148)	(0.080)	(0.067)	(0.074)	(0.099)
Punishment events × Less than 3 exclusions	0.012	0.022			-0.016	-0.061		
	(0.051)	(0.073)			(0.050)	(0.041)		
Punishment events × 3 or more exclusions	0.000	0.009			0.007	-0.028		
	(0.046)	(0.061)			(0.047)	(0.034)		
Punished once or twice			-0.043	-0.078			-0.026	0.009
			(0.076)	(0.103)			(0.088)	(0.074)
Punished more than twice			0.093	0.068			-0.018	-0.120
			(0.075)	(0.114)			(0.059)	(0.089)
Punished once or twice × Less than 3 exclusions			0.070	0.082			-0.007	-0.093
			(0.079)	(0.130)			(0.087)	(0.110)
Punished once or twice × 3 or more exclusions			0.198*	0.229			-0.121	-0.186
			(0.117)	(0.214)			(0.136)	(0.130)
Constant	0.097	0.311	0.099	0.213	0.029	0.461*	0.033	0.480
	(0.060)	(0.362)	(0.061)	(0.350)	(0.040)	(0.250)	(0.040)	(0.286)
Observations	68	68	68	68	70	70	70	70
Clusters	35	35	35	35	35	35	35	35
Controls	No	Yes	No	Yes	No	Yes	No	Yes

Notes: OLS regressions. Dependent variable is the change in the RET performance on the second group stage from Part 2 compared to the baseline, normalized by the result during the baseline task to account for ability. Reference categories in columns (3), (4), (7), and (8) are no exclusion and no punishment. Controls: changes in emotions (between the second group stage of Part 2 and the baseline), sociodemographic variables (age, female, number of siblings, number of household members, socioeconomic status), personality traits from the Big Five Inventory, economics or finance major, and a measure of inattention. Standard errors clustered at the super-group level. *** p<0.01, ** p<0.05, * p<0.1.