Studying Oppression in the Lab

Noah Bacine[1], Haley Harwell, Ryan Rholes[1]

Oxford University[1]

December 7, 2023

What is Oppression?

Oppression is malicious or unjust treatment or exercise of power, often under the guise of governmental authority or cultural opprobrium.

Economic Oppression is the social act of placing severe economic restrictions on individuals, groups or institutions. Economic oppression may take several forms, including the practice of bonded labour (in some parts of India); serfdom; forced labour; low wages; denial of equal opportunity;

What is Oppression?

Oppression is malicious or unjust treatment or exercise of power, often under the guise of governmental authority or cultural opprobrium.

Economic Oppression is the social act of placing severe economic restrictions on individuals, groups or institutions. Economic oppression may take several forms, including the practice of bonded labour (in some parts of India); serfdom; forced labour; low wages; denial of equal opportunity;

We seek to introduce an *ethical* framework to study oppressive behavior in the lab

Why Should we Care About Oppression?

Differences between social groups has been the basis of some of humanity's darkest chapters:

- The Holocaust
- Apartheid
- The Caste system

There seems to be a contradiction between the recognition of the harm/danger of oppressive behaviour and individuals' willingness to engage in it.

- Although the costs of oppression can be measured in lives, most of the evidence regarding its underlying mechanism(s) are qualitative

"Antisocial behaviour is ubiquitious in the real world ... Yet behavioral economists have devoted almost all their attention to prosociality." – Abbink & Herrmann (2011)

- Why hasn't there been more work on one of the most detrimental forms of antisocial behaviour?

Some Historical and Potentially Unethical Work

Robber's Cave Experiment (1954)-22 boys split into two groups, encouraged to form group bonds and later compete with the other group (in some case at the other groups expense)

- **Result**: Individuals will deride and mistreat the outgroup when there is competition for resources

Milgram's Shock Experiment (1963)- Authorities encourage participants to shock other 'participants' for getting questions wrong

- **Result**: Individuals will do distasteful thing under the guidance of 'authority'

Jane Elliott's Classroom (1968)- Students divided into groups based on eye colour, told they were superior and given special privileges

- **Result**: Student's internalized their status, taunting and mistreating students in the other group

Stanford Prison Experiment (1974)- Subjects were encouraged to act as prisoners as guards in a 'prison'

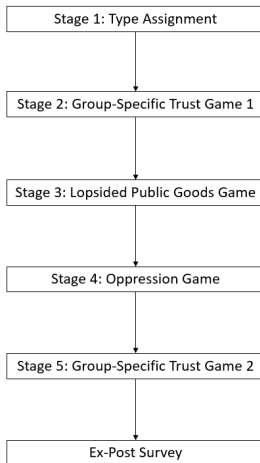- **Result**: Individuals will abuse others who are deindividualized when given authority

What we do differently

Enter some intrepid researchers:

- On the one hand, the prior work on this topic is arguably the basis for modern research ethics
- On the other hand, it seems problematic to rely on the production of observational data to study oppression

Our project seeks to build on a century of behavioural work to create an ethical fraemwork for studying oppression in a controlled setting.

Full Experimental Design

Stage 1: Type Assignment

Stage 2: Group-Specific Trust Game 1

Stage 3: Lopsided Public Goods Game

Stage 4: Oppression Game

Stage 5: Group-Specific Trust Game 2

Ex-Post Survey

Stage 1: Type Assignment

- In the first stage participants are assigned to one of two types:
  - 40% are advantaged players (Type 1's or Stars)
  - 60% are disadvantaged players (Type 2's or No stars)
- Participants unaware of the 'meaningful' difference across types until Stage 3
- Treatments vary the Type Assignment Stage (Adapted from Ball, Eckel, Grossman & Zame, 2001)
  - Baseline: Types are neutral (1/2) and randomly assigned
  - Reinforced: Types have implied status (Star/No Star) through language in instructions, celebratory ceremony and treatment by researchers but are randomly assigned
  - Earned: Types are neutral (1/2) and determined 'High' scores on a quiz [Ongoing not presented today]

Stage 2/5: Group Specific Trust Game

- Participants complete a trust game using the strategy method
  - Informed of the identity of their partner prior to making the decision
  - Partner identity is fixed across TGs
  - Endowed with 20 ECU in both roles
- First make a decision as a sender:
  - Can send amounts in discrete increments of 4
  - Any amount sent tripled by experimenters
- Then make a decision as a receiver:
  - Make a decision for every possible amount they could be sent

Why Repeat the Trust Game?

While we are primarily interested in trust/reciprocity after experiencing an oppressive environment, previous work suggests that our treatments may impact trust through a venue other than oppression:

- Lount & Petit (2012) find that high-status individuals trust more than low-status counterparts
- Wilson & Eckel (2006) find that beuatiful people are trusted more and reciprocated less
- Ball & Eckel (1996) find that low-status individuals are offered less in an ultimatum game

We run the TG (without providing any feedback) prior to the oppressive part of the experiment to establish baseline levels of trust and reciprocity

Stage 3: The Lopsided Public Goods Game

Groups of 5 play a modified public goods game for 5 rounds:

- Players start each round with 20 tokens and can:
  - Put tokens in private account providing a return of 1 ECU
  - Put tokens in the public account providing 2 ECUS split as follows:
    - **2 advantaged (Type 1/Star) players get 0.73 ECUs**
    - **3 disadvantaged (Type 2/No Star) players get 0.18 ECUs**

$$U_i = 20 - c_i + \Sigma_{i=0}^{i=5} c_i * r_i$$

  - where $r_i = 0.73$ for Type 1/Star and 0.18 for Type 2/No Star
- The game is only a public good for Type 1/Star's
  - Returns selected such that Type 2/No Star will be worse off when everyone contributes max than if no one contributes anything

- Participants receive feedback on the contribution decisions of others, their types, and earnings each round

Stage 4: The Oppression Game

Costly punishment is introduced before another 10 rounds occur:

- Only advantaged (Type 1/Star) players can punish
  - Advantaged (Type 1/Star) players pay 1 ECU for each point they assign
- Only disadvantaged (Type 2/No Star) players can be punished
  - Disadvantaged (Type 2/No Star) players lose 10% of their earnings for each point they receive
- Participants receive feedback on the contribution decisions of others & types prior to punishment
- Receive additional feedback about punishment decisions (in aggregate) and earnings afterwards

We define **oppression** as the decision to apply punishment in this context

Types of Punishment

- **Herrmann, Thöni, & Gachter (2008)** distinguish between two types of punishment
  - <u>prosocial punishment</u>-punishment of a participant who gave less than the punisher
  - <u>antisocial punishment</u>-punishment of a participant who gave at least as much or more than the punisher

Types of Punishment

- **Herrmann, Thöni, & Gachter (2008)** distinguish between two types of punishment
  - <u>prosocial punishment</u>-punishment of a participant who gave less than the punisher
  - <u>antisocial punishment</u>-punishment of a participant who gave at least as much or more than the punisher

We study a scenario in which returns are unequal & there is no possibility of conditional cooperation

Punishment in the Presence of Inequality

- **Kingsley (2016)** finds punishment increases contributions in homogenous environments but has no effect in heterogenous environments

- **Reuben & Reidl (2013)**-Introduction of punishment is least effective in increasing contributions among low type players when returns are heterogeneous (but equally effective in the presence of heterogeneous endowments)

Punishment in the Presence of Inequality

- **Kingsley (2016)** finds punishment increases contributions in homogenous environments but has no effect in heterogenous environments
- **Reuben & Reidl (2013)**-Introduction of punishment is least effective in increasing contributions among low type players when returns are heterogeneous (but equally effective in the presence of heterogeneous returns)

To the best of our knowledge, we are the first to look at such an extreme return difference (such that the good is only a pgg for part of the group)

We study a scenario in which the opportunity to punish is unilateral

Ex-post Survey

Participants complete a short survey after repeating the trust game a second time:

- Demographics
- In-group favoritism
- Norms about in-game behaviour
- Social Dominance Orientation
- Feelings & Experience during the experiment
- GSS Fair
- Self-Esteem

Types of Punishment

- **Herrmann, Thöni, & Gachter (2008)** distinguish between two types of punishment
  - <u>prosocial punishment</u>-punishment of a participant who gave less than the punisher
  - <u>antisocial punishment</u>-punishment of a participant who gave at least as much or more than the punisher

Types of Punishment

- **Herrmann, Thöni, & Gachter (2008)** distinguish between two types of punishment
  - <u>prosocial punishment</u>-punishment of a participant who gave less than the punisher
  - <u>antisocial punishment</u>-punishment of a participant who gave at least as much or more than the punisher

We study a scenario in which returns are unequal & there is no possibility of conditional cooperation

Punishment in the Presence of Inequality

- **Kingsley (2016)** finds punishment increases contributions in homogenous environments but has no effect in heterogenous environments
- **Reuben & Reidl (2013)**-Introduction of punishment is least effective in increasing contributions among low type players when returns are heterogeneous (but equally effective in the presence of heterogeneous endowments)

Punishment in the Presence of Inequality

- **Kingsley (2016)** finds punishment increases contributions in homogenous environments but has no effect in heterogenous environments
- **Reuben & Reidl (2013)**-Introduction of punishment is least effective in increasing contributions among low type players when returns are heterogeneous (but equally effective in the presence of heterogenous returns)

We study a scenario in which the opportunity to punish is unilateral

Sample and Data collection

- We ran 3 pilot sessions at Oxford's lab in April 2022
- data collection for first two treatments completed:
  - 13 sessions of Baseline:
    - 150 participants in 30 groups of 5
    - 60 'oppressors' and 90 'oppressees'
  - 8 sessions of Reinforced:
    - 125 participants in 12 groups of 5
    - 50 'oppressors' and 75 'oppressees'
  - 4 sessions of Earned:
    - 65 participants
    - 26 'oppressors' and 39 'oppressees'
- AEA pre-registration: AEARCTR-0009160

Trust Game 1
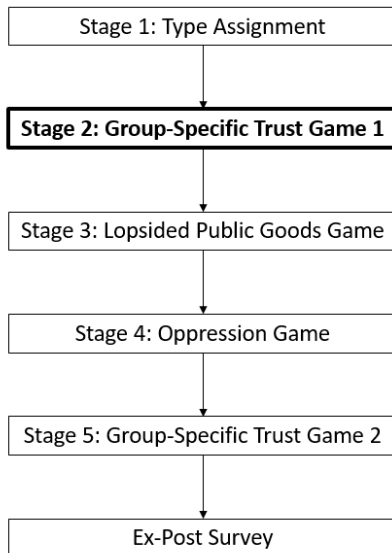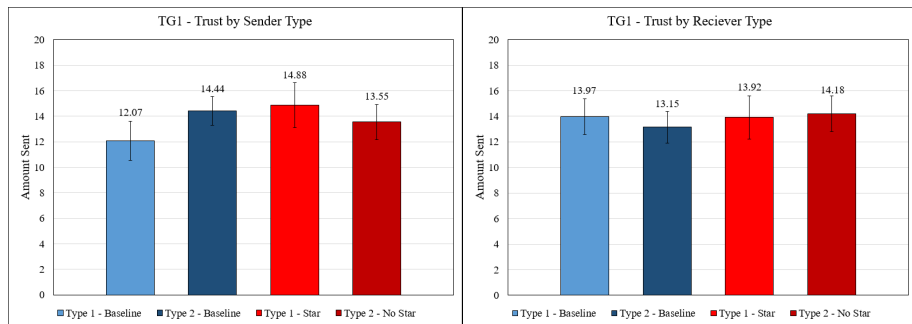
The only noticeable difference is Type 1's in the baseline treatment who exhibit significantly less trust than Type 2's in the baseline and Type 1's in the status treatment

TG1 - Trust by Shared Identity
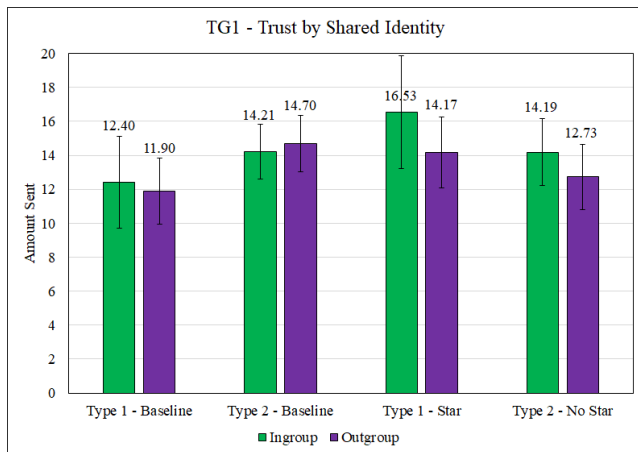
Suggestive evidence that the 'Status' treatment is doing something but none of the differences are significant
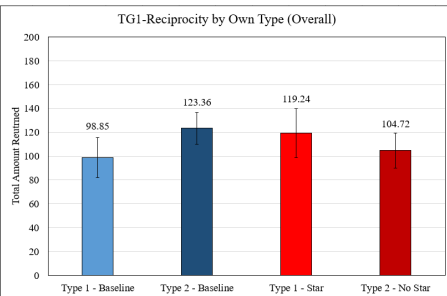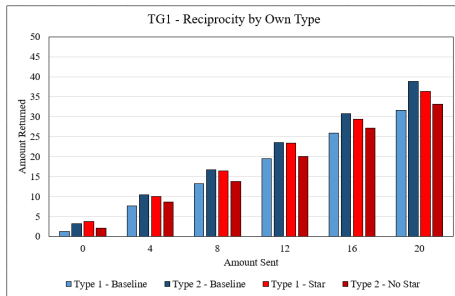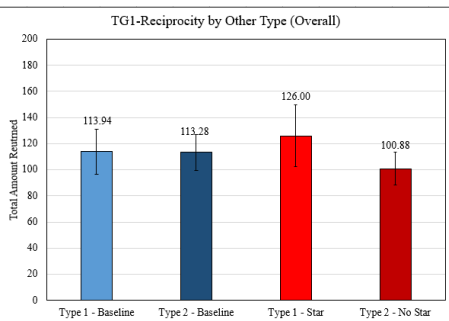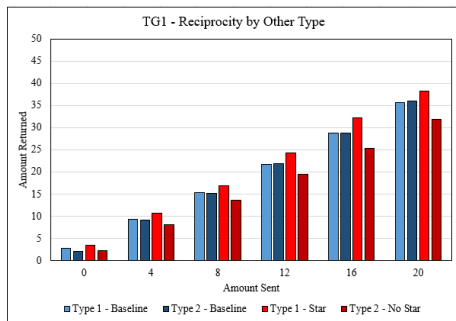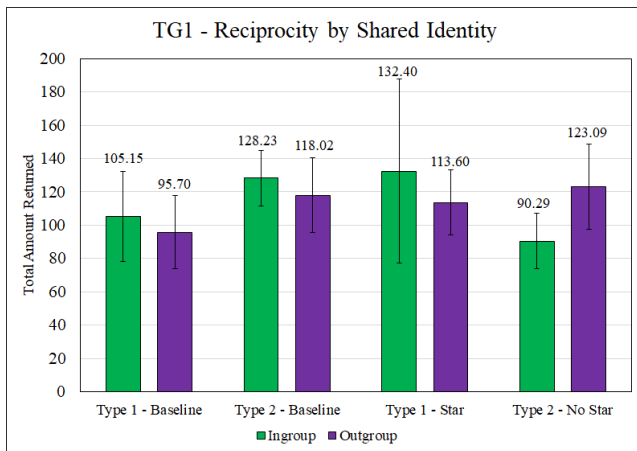
No stars are significantly more reciprocal to their star counterparts (p¡0.05)
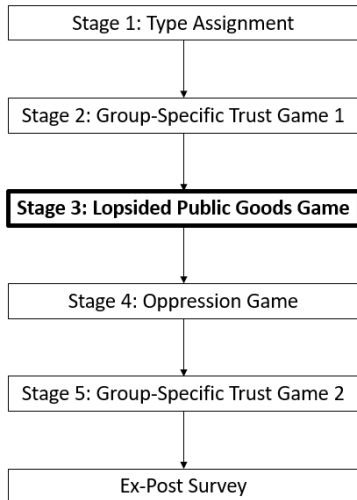
LPGG
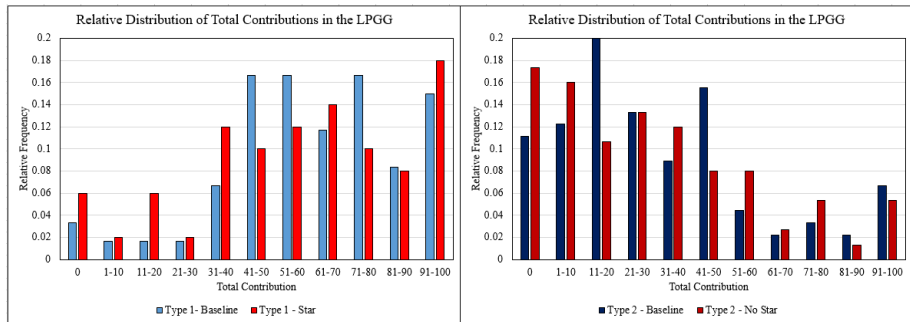
Stage 3: The Lopsided Public Goods Game

Groups of 5 play a modified public goods game for 5 rounds:

- Players start each round with 20 tokens and can:
  - Put tokens in private account providing a return of 1 ECU
  - Put tokens in the public account providing 2 ECUS split as follows:
    - **2 advantaged (Type 1/Stars) players get 0.73 ECUs**
    - **3 disadvantaged (Type 2/No Stars) players get 0.18 ECUs**
    - where $r_i = 0.73$ for Type 1 and 0.18 for Type 2
  - The game is only a public good for Type 1's
    - Returns selected such that Type 2 will be worse off when everyone contributes max than if no one contributes anything
- Participants receive feedback on the contribution decisions of others, their types, and earnings each round
- Similar to the the first TG, the LPGG serves to both familiarize participants with the environment and to serve as a baseline for comparison when oppression is introduced.
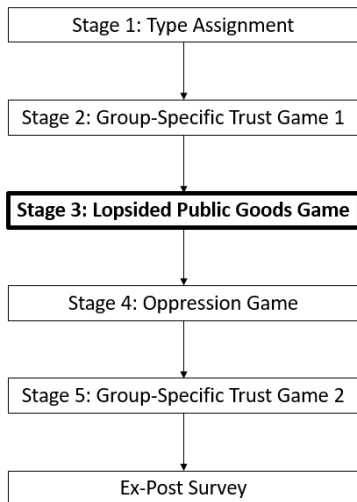
Lopsided Public Goods Game Contributions
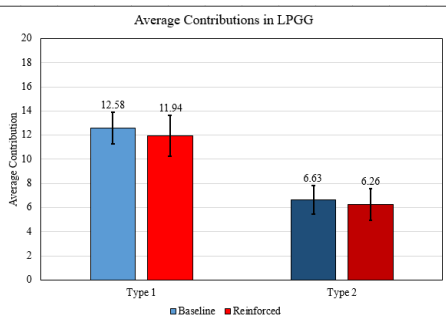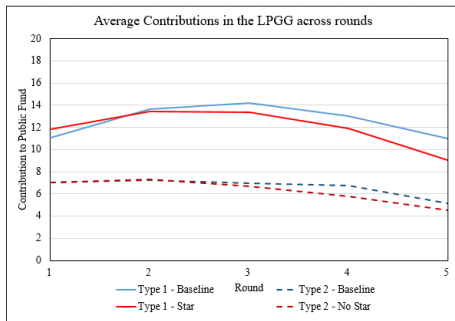
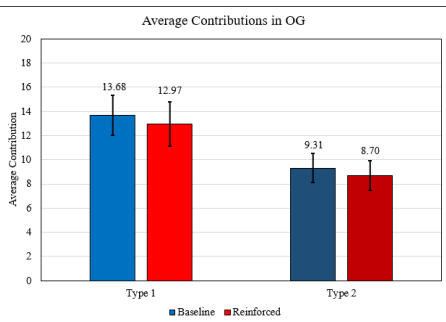# Distribution of Lopsided Public Goods Game Contributions

Oppression Game

Stage 1: Type Assignment

Stage 2: Group-Specific Trust Game 1

Stage 3: Lopsided Public Goods Game

Stage 4: Oppression Game

Stage 5: Group-Specific Trust Game 2
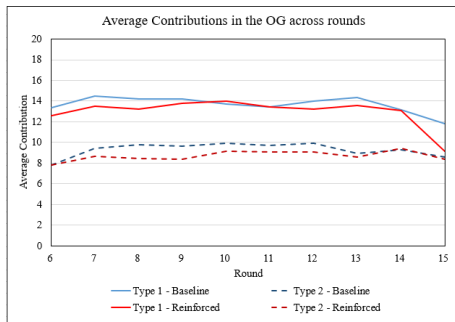
Ex-Post Survey

Stage 4: The Oppression Game

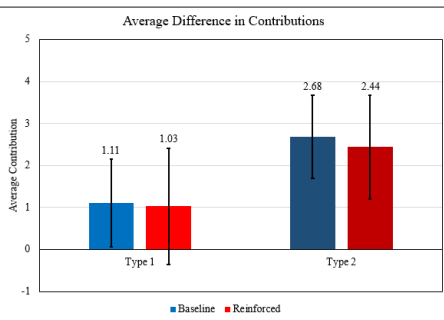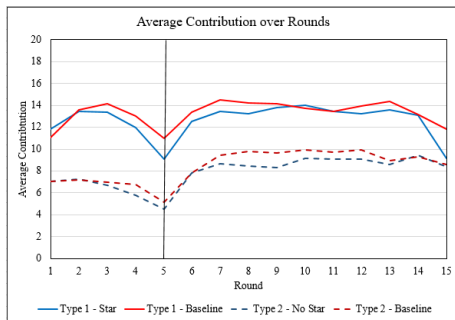Costly punishment is introduced before another 10 rounds occur:

- Only advantaged (Type 1/Star) players can punish
    - Advantaged (Type 1/Star) players pay 1 ECU for each point they assign
- Only disadvantaged (Type 2/No Star) players can be punished
    - Disadvantaged (Type 2/No Star) players lose 10% of their earnings for each point they receive
- Participants receive feedback on the contribution decisions of others & types prior to punishment
- Receive additional feedback about punishment decisions (in aggregate) and earnings afterwards

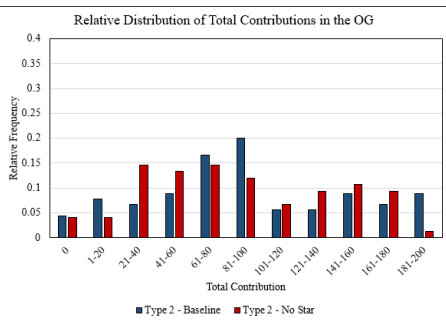We define **oppression** as the decision to apply punishment in this context
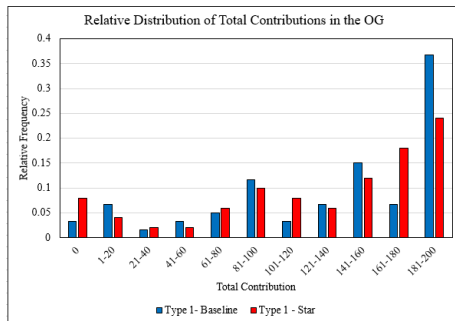
Comparison of LPGG and OG 2

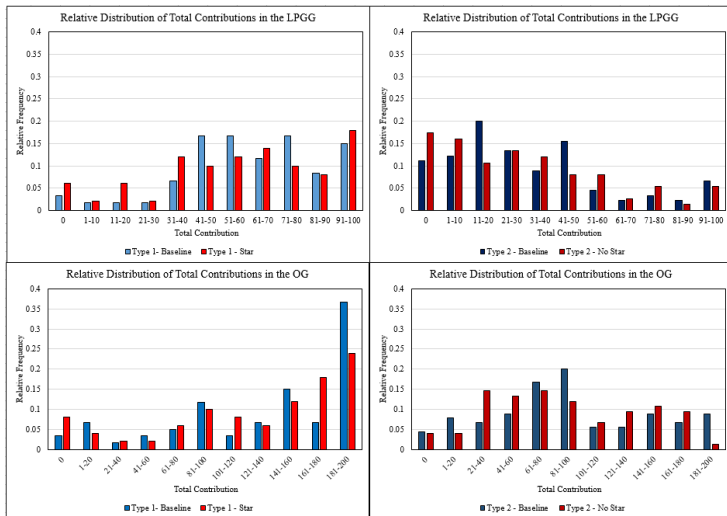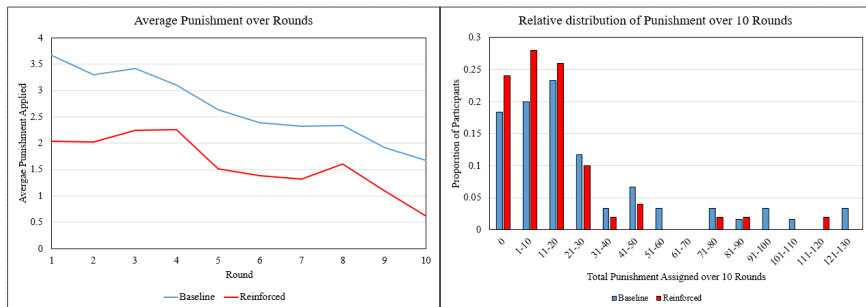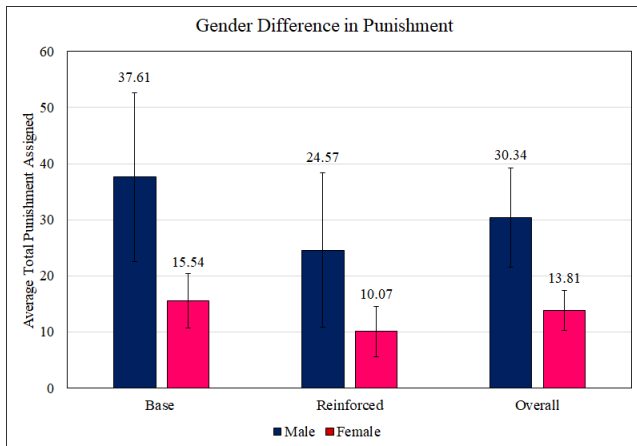# Punishment Behaviour

Average punishment over 10 rounds:

- Baseline: 26.73
- Reinforced: 16.1

The difference is significant at the 5% level.

Gender Heterogeneity in Punishment

Table 1: Regression Table: Does Punishment Respond to Contributions?

	(1) Punishment	(2) Punishment	(3) Punishment	(4) Punishment
Contribution	-0.113****			-0.064****
	(0.014)			(0.015)
$\Delta Contribution$		-0.063****		-0.005
		(0.009)		(0.006)
$Cont. - \overline{Cont._{g,t}}$			-0.145****	-0.080****
			(0.018)	(0.019)
Constant	1.830****	0.946****	0.936****	1.460****
	(0.202)	(0.121)	(0.123)	(0.199)
N	4080	3672	4080	3672
Clusters	68	68	68	68

Random effects regressions with time FE. Robust standard errors clustered at group level in parentheses.

$^{*}\ p < .10,\ ^{**}\ p < .05,\ ^{***}\ p < .01,\ ^{****}\ p < .001$

Table 2: Regression Table: Punishment Decisions By Treatment

	(1) Punishment	(2) Punishment	(3) Punishment	(4) Punishment	(5) Punishment
Contribution	-0.113****	-0.115****	-0.112****	-0.107****	-0.150****
	(0.014)	(0.015)	(0.025)	(0.019)	(0.043)
Reinforced		-0.477**			
		(0.231)			
Earned		-0.618***			
		(0.228)			
Constant	1.830****	2.134****	2.170****	1.515****	1.774****
	(0.202)	(0.261)	(0.378)	(0.260)	(0.400)
N	4080	4080	1800	1500	780
Clusters	68	68	30	25	13
Treatment	All	All	Baseline	Reinforced	Earned

Random effects regressions. Robust standard errors clustered at the group level in parentheses.

* $p < .10$, ** $p < .05$, *** $p < .01$, **** $p < .001$

Note: We achieve qualitatively identical results if we instead use a Newey-West estimator or if we use Bootstrapped standard errors.

Table 3: How Punishment Responds to Contributions By GSS_{Fair}

	(1) Punishment	(2) Punishment	(3) Punishment	(4) Punishment
Contribution	-0.115****	-0.095****	-0.154****	-0.108****
	(0.004)	(0.009)	(0.011)	(0.006)
Female	-0.638****	-0.649***	-0.810***	-0.612****
	(0.116)	(0.228)	(0.271)	(0.156)
Constant	0.340**	0.293	0.422	0.355*
	(0.137)	(0.236)	(0.408)	(0.188)
N	4020	1050	840	2130
GSSType	All	Take Advantage	Fair	Other

Random effects regressions with time FE. Robust standard errors clustered at group level in parentheses.
* $p < .10$, ** $p < .05$, *** $p < .01$, **** $p < .001$

We asked participants how they felt during the experiment

Type 1 Qual Responses – Justified

- "It was very fun to penalize people, although I think this was because they were not very human (just types on a screen). I felt a little bit bad, but I somehow felt like the other type 1 and I were a team to maximize our earnings. Thinking back, I don't think I did the right thing in penalizing people, but I did the more entertaining and lucrative thing."

- "I was Type 1 (in Part 4) and I felt I was educating Type 2 to behave in a particular way, it got more interesting."

- "I felt somewhat powerful being a Type 1. It was frustrating when Type 2 werent putting what I thought to be enough into the public fund, and I felt as though they needed to be "punished" for what they were doing to hopefully get them to change their behaviour."

- "it annoyed me that type 2 people continued not putting all their monezy into the public account even though it is obvious they will be punished for it."

Type 1 Qual Responses – Guilt

- "I felt good being in Type 1. I felt a slight guilt when I tried to take advantage of the other participants to earn more money. If it were not for the veil of anonymity, I would probably not have acted as unfairly as I did. In a real-life situation, though, if greater amounts of money were at stake, I would be less comfortable doing what I did."

- "It's scary how behind closed doors so to speak, we can become ruthless and lose all compassion. If interacting with those people IRL i probably would've made different decisions... I usually think of myself as pretty altruistic but now I might how to revise that in light of today! Good job with the experiment though, well designed"

- "I was angry at myself as well at some points and ashamed when i realized I was being harsh and unfair toward the type 2s - I started to understand that it was normal for them not to want to put anything in the public account since that meant virtually giving it all up - and so i felt guilty for having penalized them for doing so when it's actually a perfectly rational reaction"

- "Being type 1 was odd, it didn't feel fair I had more power than others but it was by random which made me feel slightly better"

Star Qual Responses

- "I felt like i started seeing the No stars as means to an end and expected them to be compliant to the rules imposed by the penalisation more than the fellow star"
- "It actually felt good to be in charge, I knew nothing bad could happen to me given the rules that favoured Stars and I think I used up all my penalty point to incentivize Player 4 who kept donating very little to the common pot, even if, in reality, I felt quite sorry for them as we (the two Starts) kept taking advantage of them."
- "Slightly uncomfortable to be in the dominant group"
- "Sometimes I feel difficult and bad to penalize non-stars, but given my role and my plausible "aim", I guess my doing could be justified by reason."
- "didn't want to penalise the No Stars - i felt i had no right to their ECUs."
- "I felt a little guilty penalising some of the No Star participants."
- "I felt bad for the no stars and felt uncomfortable taking advantage of my power and position. However, I did try to benefit from my situation and enforce penalty points to take advantage of my position."
- "I felt uncomfortable standing at the front to receive applause as a Star, as the status of Star was not something I had done anything to earn."
- "Felt uncomfortable being a Star and having the power to penalize"

Type 2 Qual Responses

- "I was very unhappy about being powerless in part 2 and 3, and angry at the Type 1 people applying penalties even though i was already earning less than them; it felt like they did not understand that my situation was different from theirs, or they simply did not care about it"

- "We had to put some money into the public account, to appease the type 1's so they wouldn't penalise us, whilst we knew they were still making loads more than us. I felt annoyed and helpless, but got by knowing I was going to come out of this with at least something."

- "I felt quite negative through the experiment. I did not like my group as It felt a lot like the struggles I face as a both a disabled and a class act member. I am often disadvantaged by both of these issues while I am also advantaged in other manners. Only I am not able to always prove this since I am not given the chances other people are thrown for doing nothing but being raised through wealth, success and perfection."

- "being assigned to Type B (with lower conversion rates) makes me feel I am in an inferior group and my effort and gain are not positively related (feel a bit helpless)"

Type 2 Qual Responses 2

- "Felt powerless at times to be Type 2 participant, especially when one of the Type 1 participants were being unfair and exploitative in addition to self-serving. Self-serving goals could have been achieved in this experiment without being unfairly exploitative (which was also the point I think) but it needs the cooperation of all, at least of a similar Type, especially when richer."

- "I was uncomfortable that some people had a higher chance of earning more than others because of a randomly assigned group, however, I was more uncomfortable with the idea that some people could penalise others because of their decisions simply because they did not agree with them."

- "I felt that type 1 people took advantage of their advantageous position. Effectively forcing type 2 individuals to contribute money to the public account (and being worse off than if they kept their money in the private account) while failing to contribute to the public account themselves."

No Star Qual Responses

- "Feeling some unfairness towards being part of the No Star group, also feeling solidarity towards my No Start peers. Certainly expecting high moral standards from people in the Start Group and holding them to principle of fairness and wealth re-distribution. Feeling very disappointed when I felt they took decisions that negatively impacted no stars"

- "I was very frustrated being a No-Star. Whatever I did, was not enough to escape from penalties of Starts, which showed no compassion. Just awful"

- "At first I was very chilled and did not feel strongly at all about it. I got quite upset realising that one of the stars was behaving like an arsehole. Penalising us for not wanting to loose more money or invest our money only to make them more money. I tried to be really nice and even return money to people not giving me money and also putting into public account much more than I would usually give - but then the stars decided to be very mean. Upset me."

- "Feeling like the underdog- unfairly treated and made to do things i wouldn't necessarily have done"

- "I was sad all the time for not being a star, injustice made me anxious and a bit resentful."

- "a bit triggering as it kind of mimics real life inequalities in privilege"

Some Rebellious No Stars

- "I felt a bit frustrated at one point when I tried to implement a kind of 'reverse' penalising of the stars at the stage where 8 out of 10 stars penalised me for putting in 25 percent of my allowance, instead of 50 percent like the other non stars. With the number of penal points given to me growing, I decided to decrease the amount that I put in the public account. Unfortunately, the other non stars stuck with their 50% investment. I suppose my idea would only have worked out if all non stars would have decided the same way."

- "Increasingly resentful toward the stars for not taking a lead earlier on to help us all do better, especially when they benefitted more and had less risk, when I realised this I felt angry and tried to succeed on my own, when they punished me I felt it was more important to punish them than help myself."
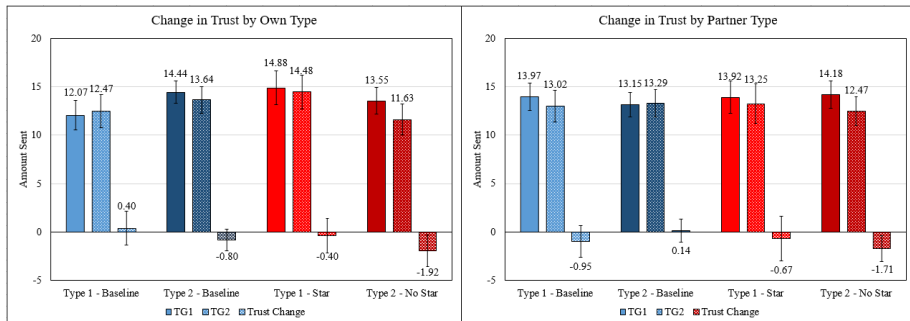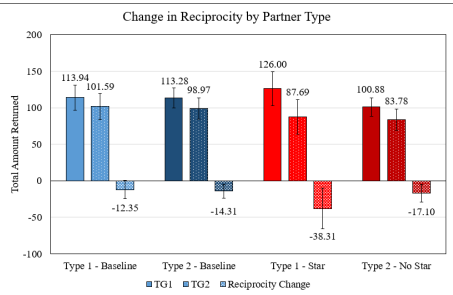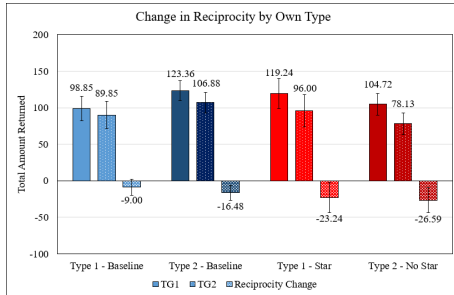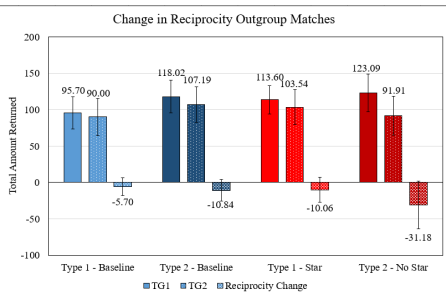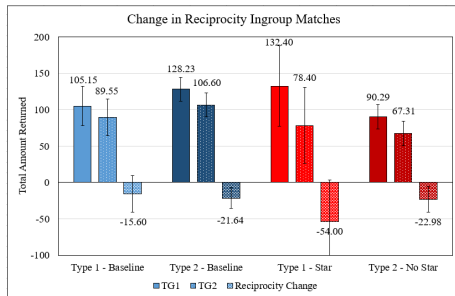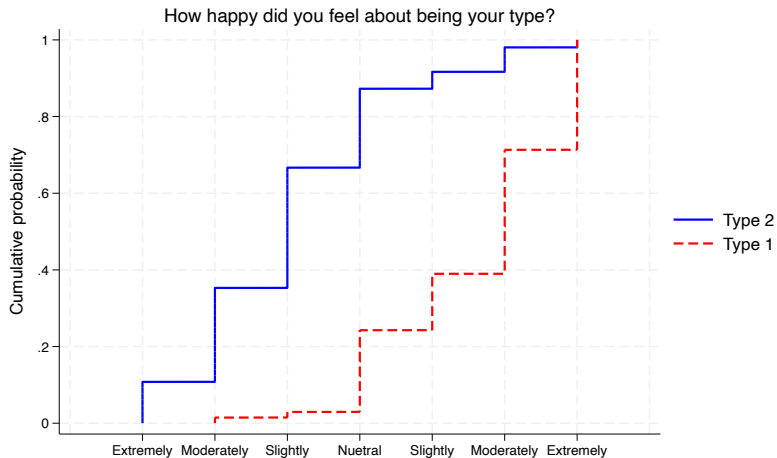
Impacts on Trust - Group identity
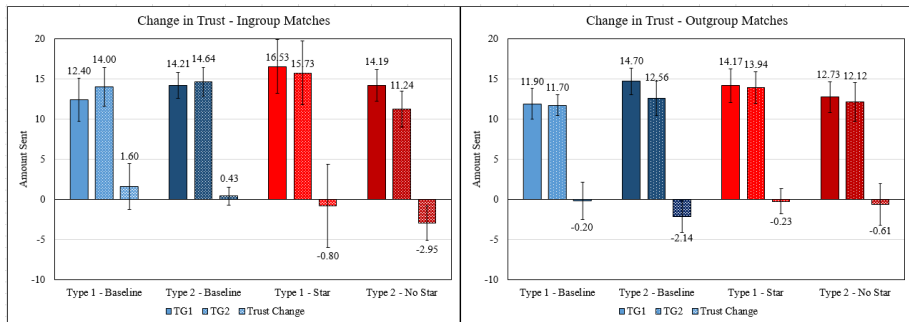
Impacts on Reciprocity - Type

Impacts on Reciprocity - Group identity
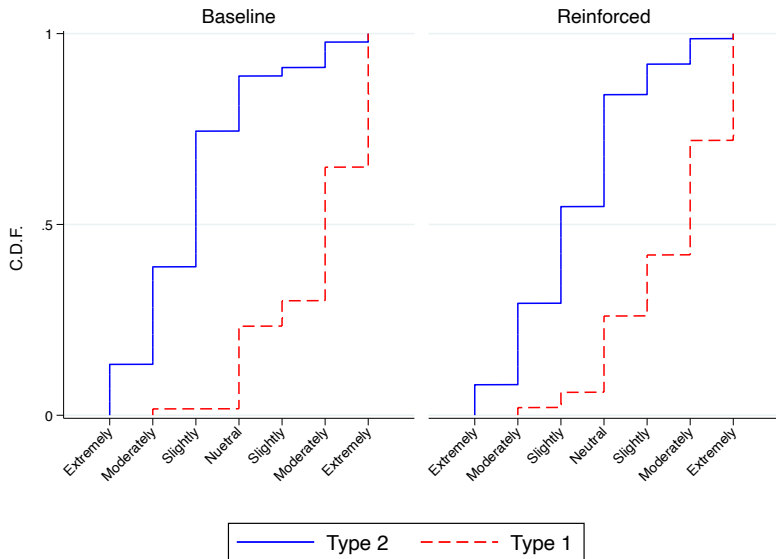
Happiness about being One's type

How happy did you feel about being your type?

Happiness about being One's type

Summary of Findings

- Despite a large literature on the importance of fairness and equality concerns in the PGG setting, we document a willingness to punish in an unfair scenario
- People will oppress others given the opportunity (will encourage payoff decreasing behavior in unfair situations)
- Willingness to oppress differs across individuals with some fully abstaining & others being as extractive as possible
- The experience of an oppressive environment impacts both trust and reciprocity
- Despite our expectations, our 'status' treatment actually made participants less willing to oppress
- **Additional analysis on individual experience during the experiment is essential to unpacking our treatment effects**

Thank you!

Contact us at:

noah.bacine@nuffield.ox.ac.uk
ryan.rholes@economics.ox.ac.uk