



**DEPARTMENT OF ECONOMICS
DISCUSSION PAPER SERIES**

**WHEN TO FAVOUR YOUR OWN GROUP?
THE THREATS OF COSTLY PUNISHMENTS AND IN-GROUP
FAVOURITISM**

Donna Harris, Benedikt Herrmann and Andreas Kontoleon

Number 628
November 2012

Manor Road Building, Oxford OX1 3UQ

When to Favour Your Own group?

The Threats of Costly Punishments and In-group Favouritism*

Donna Harris[†]

Benedikt Herrmann[‡]

Andreas Kontoleon[§]

November 9, 2012

Abstract

Using a laboratory experiment with minimal groups, we examined the extent to which the threats of costly punishments affect in-group favouritism behaviour. We studied three types of punishment separately: in-group, out-group, and third-party punishments. In line with previous studies, the majority of the allocators favoured their own group by allocating more money to each of the in-group members at the expense of the out-group in the baseline without punishment. In the in-group punishment treatment, we observed a slight increase in in-group favouritism behaviour. On the contrary, when only the out-group could punish the allocators, there was a significant drop in in-group favouritism behaviour as well as an increase in the equal division option. Finally, when faced with an independent third-party punisher the allocators continued to favour their own group. The threat of third-party punishment appeared to have no effect on their decisions. Our paper contributes to the literature on in-group favouritism and the nature of social norms by showing that the decision whether to favour one's own group is affected by the threats of in-group and out-group punishments and whether it leads to an increase or decrease in this behaviour depends on *who* has the punishment power. Parochial or in-group biased norm was enforced by the in-group members, whilst 'egalitarian sharing norm' (across groups) was enforced by the out-group members. We conclude firstly that people apply different 'self-serving' social norms depending on their own group identity. Secondly, unlike selfish or opportunistic behaviours, independent third-parties, who only observed this behaviour but were not directly affected by it, were not willing to punish this behaviour.

JEL: D70, D73, C92.

Keywords: In-group Favouritism, Group Behaviour, Social Identity, Social Norm, In-group Punishment, Out-group Punishment, Third-party Punishment, Favour Game.

1 Introduction

In-group favouritism is a preferential treatment given to a person or a group at the expense of outsiders (Becker, 1957). By definition, in-group favouritism generates negative externalities to those outside the group. But despite these negative externalities, previous studies have shown that people have a natural tendency to favour their own group, regardless of how the groups are formed (Bernhard *et al.*, 2006b,a; Charness *et al.*, 2007; Goette *et al.*, 2006; Guth *et al.*, 2008; Chen and Li, 2009; Tajfel *et al.*, 1971; Tajfel and Turner, 1986). Whilst in-group favouritism is often perceived as unjust by those who are not favoured (the out-group), we still know very little about whether there is a social norm which governs this behaviour and if the norm exists, what kind of norm it is. Furthermore, unlike selfish

*We are grateful for the discussions with Simon Gächter, Daniele Nosenzo, Jan Potters, Aldo Rustichini, Tim Salmon, and for the kind assistance from the members of the Economic Faculty at Chulalongkorn University, particularly Tanapong Popiti and the five Thai research assistants for their excellent work. We are also very grateful for the financial assistance from the Economic and Social Research Council (ESRC) Grant Number: RG58935, the British Academy, the Leverhulme Trust, and the Suzy Paine Trust of the Faculty of Economics, University of Cambridge.

[†]Corresponding author: Department of Economics, University of Oxford, Manor Road, Oxford, OX1 3UQ / Somerville College, Woodstock Road, Oxford, OX2 6HD. Email: donhatai.harris@economics.ox.ac.uk

[‡]Task Force Public Health, Institute for Health and Consumer Protection, Joint Research Centre, European Commission. Email: Benedikt.Herrmann@ec.europa.eu

[§]Environmental Economy and Policy Research Group, University of Cambridge, 19 Silver Street, Cambridge, UK, CB3 9EP. Email: ak219@cam.ac.uk.

and opportunistic behaviours which have been shown to universally considered as a clear violation of cooperation and egalitarian sharing norms (Fehr and Gaechter, 2000; Bernhard *et al.*, 2006b,a), we do not know whether in-group favouritism is considered as a norm violation or not. The main objective of this paper is to shed a light on these questions.

Social norms are behavioural standards which are based on socially shared beliefs on how individuals ought to behave in a given situation (Bernhard *et al.*, 2006b,a). The demand for a social norm arises when actions cause negative side effects or externalities for other people (Fehr and Fischbacher, 2004b) and thus, social norms can be considered as a mean to regulate individual actions and to limit these negative externalities. From the very early phases of human evolution, social norms have shaped and regulated human’s social and economic interactions such as food sharing, cooperative hunting, and participation in warfare (Bernhard *et al.*, 2006b,a). Even in modern societies where formal rules of law and enforcement mechanisms are well-established, social norms still play a crucial role in maintaining social orders. This is because a large part of human interactions still takes place within informal social environments such as family, workplace and residential neighbourhoods, which are not regulated by explicit contracts (Fehr and Gaechter, 2000). Social norms are typically enforced by an individual’s willingness to impose an informal sanction on those who deviate from the behavioural standards, even at the punisher’s own cost (Fehr and Fischbacher, 2004b; Hoff *et al.*, 2007). Social sanctions are not costless since there are pecuniary and non-pecuniary costs (or both) associated with imposing a sanction on others. In addition, there is also a risk of retaliation and emotional tension from the punished individual (Fehr and Fischbacher, 2004a). Consequently, the willingness to punish norm violations tends to stem from strong negative emotions (Elster, 1989), which can outweigh these costs. In this paper, we examine (i) whether a social norm exists within the context of in-group favouritism? (ii) whether such a norm is expected to be enforced? and (iii) how does the threat of norm enforcement change people’s decision whether to favour their own group?

We design a new one-shot sequential allocation experiment with minimal groups and costly punishments to address these questions. At the beginning of the experiment, we use a randomised group assignment method by labelling some subjects randomly as ‘group A’ and others as ‘group B’. We use this method rather than natural group identity because, as pointed out in Chen and Li (Chen and Li, 2009), the same person may identify with more than one natural group identity (e.g. nationality, ethnicity, occupation, religion, gender, etc). Moreover, the degree to which they identify with natural group identities may also vary¹ and there may also be different social norms associated with different natural identities. Furthermore, we want to establish a clean baseline in order to see if we can replicate previous studies which have shown that group categorisation alone can trigger in-group favouritism behaviour. Once we have established this baseline, we can introduce costly punishments such that the decision whether to punish the allocator is only related to the allocator’s behaviour and not with any other social dimensions of groups. In the first stage of the experiment, an allocator decides how to divide a fixed sum of money between the two groups. In the second stage, the decision is revealed to the punishers (either the in-group members, the out-group members or an independent third-party in three separate treatments) who decide whether to punish the allocator by deducting her payoff at a cost.

Our design departs from the previous minimal group allocation experiments in a number of respects.

Firstly, the allocation decision is concerned with two groups of three members rather than only one

¹Ben-Ner *et al.* (2009), for example, show that there is heterogeneity in the degree to which people identify with different natural groups. The authors investigate the existence and relative strength of favouritism for in-group versus out-group along multiple identity categories (body type, political views, nationality, religion, and more) in four alternative contexts: (1) giving money in a dictator game, (2) sharing an office, (3) commuting, and (4) work. They find that people who belong to the in-group are treated more favourably than those who belong to the out-group. However, family and kinship are the most powerful source of differentiation, followed by political views, religion, sports-team loyalty, and music preferences, with gender being basically insignificant.

member in each group. This is because in practice, many economic decisions involve allocating resources between groups, such as politicians in charge of a budget allocation often have to decide how to divide resources between different interest groups (Naegelen and Mougeot, 1998; Zantman, 2002). In addition, a number of studies have also shown that increasing the number of people in a social interaction can alter preferences and decisions (Isaac *et al.*, 1994; Gaube, 2000; Goeree *et al.*, 2002; Bandiera *et al.*, 2006; Stahl and Haruvy, 2006; Andreoni, 2007)². Therefore, we want to see whether we are able to replicate previous findings when there are more than one person in each group. Secondly, our experiment is a zero-sum game in which the amount of money to be allocated is fixed and thus, the allocator's decision only affects income distribution between the two groups without changing the level of economic efficiency (the size of the pie remains constant). Thirdly, we control for self-interest preference by design because the allocator is not allowed to take any share of her group's payoff. Instead, the allocator gets a fixed payoff from the experimenter. Finally, we minimise potential confound from the allocator's (disadvantageous) inequity aversion preference (Fehr and Schmidt, 1999; Bolton and Ockenfels, 2000) by fixing the amount that the allocator receives to be equal to the maximum amount that could be allocated to each member of the two groups.

Our game is also different from the third-party punishment experiment carried out by Bernhard *et al.* (Bernhard *et al.*, 2006a,b) which examines parochial nature of altruistic norm enforcement. The first and most obvious difference is that in Bernhard *et al.*, the authors used natural group identities in Papua New Guinea: the Wolimbka and the Ngenika instead of minimal groups. They also argued that in-group favouritism observed in previous minimal group experiments does not reveal parochial tendencies in human altruism because the subjects only distributed resources between two other subjects. Subjects thus bore no cost, regardless of how they allocated the available resources between the other two individuals (Bernhard *et al.*, 2006a). In addition, they also pointed out that costless in-group favouritism in these experiments is due to the expectation that ingroup members will receive some reciprocation from other ingroup members. Thus, what looked like a preference for ingroup favouritism was in fact based on the expectation of ingroup reciprocity. We believe that our design overcomes these problems because firstly we run an anonymous one-shot experiment in which the allocator's payoff is exogenously fixed by the experimenter and thus, in-group reciprocity is not possible. Secondly, we introduce costly punishments which means that there is a potential cost to the allocator's decision if he is punished. We implement the standard punishment mechanism in which the punisher can pay to deduct the allocator's payoff. Therefore, in-group favouritism is not costless in our game.

Secondly, in Bernhard *et al.* (Bernhard *et al.*, 2006a) the allocator was asked to divide the money between herself and another subject who either belonged to the same group or belong to a different group. This is very different from our game in which the allocator's decision is only concerned with other people's payoffs and not their own. This is because our main interest is in in-group favouritism behaviour rather than sharing behaviour (between self and others). In our design, the allocator does not explicitly act in a selfish or opportunistic manner if she favours her own group since she does not stand to gain anything from her decision. Hence, there is no violation of egalitarian sharing norm in the same sense as in Bernhard *et al.*'s paper. Finally, the design of the punishments in Bernhard *et al.* is also different from ours. In their study, the third party can either come from the same group as both the dictator and the recipient (ABC treatment), the same group as only the dictator (AC treatment),

²For example, increasing the number of people within a group of recipients in a dictator game has been shown to decrease altruistic preference (Andreoni, 2007). Other studies have found that a larger number of people involved in a voluntary contribution to public goods lead to more selfish preference as observed by an increase in free-riding behaviour (Goeree *et al.*, 2002). On the contrary, Stahl and Haruvy (2006) find that a larger number of recipients increases the dictator's kind action (by allocating a larger sum of money) because it increases the 'warm glow' from this action which has a positive effect on the dictator's utility function.

the same group as only the recipient(BC), or from a completely different group (AB treatment). In our design, we have three types of punishments: in-group, out-group, and third-party. Our in-group punishment may appear to be similar to the ABC design; however, the payoffs of in-group punishers in our experiment are directly affected by the allocator’s decision and thus, they cannot be considered as ‘third-party’ punishers. On the contrary, the ABC treatment in Bernhard et al, the payoff of player C who is the third-party punisher is completely independent of the dictator’s decision. In our out-group punishment, the out-group punishers’ payoffs in our favour game are also dependent on the allocator’s decision and thus, they cannot be considered as third-party punishers. We introduce a completely independent person who belongs to neither groups but observes the allocator’s decision in our third-party punishment. This is similar to the AB treatment in Bernhard et al. But we are interested in different type of behaviour which is preferential treatment given to the in-group rather than sharing behaviour as in their study. Given a number of the differences outlined above together with other features of our design (for example, the size of the group and the fixed choice set), we believe that our experiment can offer new insights on the effects of costly punishment and norm enforcements on in-group favouritism behaviour.

We find that in the baseline without punishment, the majority of the allocators exhibit in-group favouritism by allocating more money to each member of their own group at the expense of the out-group. Our results replicate previous minimal group finding that people have a natural tendency to favour their own group even when the notion of a group has no social meaning. In addition, we also show that this tendency persists even when each group is consisted of more than just one member and in a one-shot setting. Secondly, we find that the allocator’s decision whether to favour her own group is influenced by the threat of punishment, but only in some circumstances and the direction of the effect depends on the *identity* of the punishers. Our results also indicate that people appear to follow different social norms when deciding whether to favour their own group or not. It seems that when people are part of a group (the in-group), they expect to be favoured. However, when they are not part of the in-group and directly lose out from in-group favouritism decision, they expected egalitarian distribution. More importantly, when they do not stand to gain or lose anything from in-group favouritism, they do not perceive this behaviour as a violation of a social norm. We find these results very interesting. We interpret our results in the following way: that in-group favouritism behaviour is different from purely selfish or opportunistic behaviour and is not universally perceived as a violation of a social norm. From our results, it seems that it is considered as unacceptable only by those who stand to lose out (the out-group), but not by those who would benefit (the in-group) nor those who are not involved (the third-party). We believe that our work paves way for further work on the nature of social norms which influence in-group favouritism behaviour.

The rest of the paper is structured as follows. Section 2 describes the details of the experimental design and section 3 reports the main behavioural results. Section 4 analyses internal validity of our results and section 5 concludes.

2 Experimental Design

2.1 The Baseline

In the baseline allocation game, seven subjects are randomly matched and divided into two groups: four subjects are assigned to ‘group A’ and three to ‘group B’ and each subject is given an initial endowment of 3,000 tokens regardless of their group (which serves as a protection against bankruptcy). We only

allow members of group A (the in-group) to make the allocation decisions and then randomly select one person from group A to be the allocator. We take allocation decisions from all members of group A in order to maximise the number of observations per session. The allocator then receives a fixed payoff from the experimenter. This explains why we need to have asymmetrical group sizes initially i.e. four in group A and three in group B, such that after the allocator is selected, we can have an equal number of the recipients in each group (three in-group versus three out-group). Our random group assignment method adheres to the minimal group paradigm criteria (Tajfel and Turner, 1986)³, which include assigning subjects to non-overlapping groups; no social interaction takes place between subjects (no face-to-face communication or online chat); group membership is anonymous; and the decision task requires no link between the allocator’s self-interest and her choice. If our design is able to replicate previous findings, we expect to observe in-group favouritism behaviour in the baseline (subjects choosing options a, b, or c).

Once all the members of group A have made their allocation decisions, the experimenter rolls a die to select the decision to be implemented and the person whose decision is picked is assigned the role of the allocator. This random selection mechanism of the allocator reduces the probability that each subject can affect the other in-group members’ payoffs⁴. Finally, the allocation decision is for *each* member of the group rather than for the group as a whole. This is in order to maintain equal distribution of the payoff amongst the group members within each group.

Another key feature of the design is that we use a fixed and symmetrical *choice set* to control for the decision space and to keep the stake constant across all treatments. As shown in Figure 1, our choice set consists of three different magnitudes of in-group favouritism ($a = \text{high}$, $b = \text{medium}$, $c = \text{low}$); an egalitarian distribution option (d); three different magnitudes of out-group favouritism in a reverse order to the in-group favouritism options (e, f, g); and an option which allocates zero to both groups (h)⁵. All options, except for option h , add up to the same total fixed amount of 3,000 tokens. The exchange rate is 100 tokens = 7 pence⁶ for two main reasons: firstly, to focus on the impact of in-group favouritism on income distribution between the two groups; and secondly, to make it clear to the subjects that the choices (except option h) have no effect on economic efficiency. For example, if the allocator chooses option a , the largest amount (4,500 tokens) is given to each of the in-group member at a directly cost of 1,500 tokens to each of the out-group member. As a result, this option leads to an extremely unequal income distribution between the two groups. For option d , the income gap between the in-group and the out-group is at the minimum (zero), which is the best outcome as far as income distribution is concerned. Options e, f, g , also create income inequalities between the in-group and the out-group members, but in the reverse order of options a, b, c . Finally, similar to option d , the difference between the payoffs between the two groups for option h is also zero, but in this case there is a loss in economic efficiency of 3,000 Tokens.

³The aim of this method is to test whether group categorisation alone is sufficient to trigger in-group favouritism and out-group discrimination behaviour (Tajfel *et al.*, 1971; Billig and Tajfel, 1973; Brewer, 1979). For more detailed review of previous experiments, we refer to Chen and Li (2009).

⁴In order to test whether it is still possible that in-group favouritism observed in our design is driven by *generalised reciprocity* among the in-group members (Yamagishi and Kiyonari, 2000), we carry out a separate treatment which has the same features as in the baseline game, except for group A only consists of two members. Therefore, the probability of being chosen to be the allocator is higher (0.5 instead of 0.25 in our original baseline game). If generalised reciprocity is the main driving factor for in-group favouritism, we should observe a higher magnitude of in-group favouritism in this treatment compared to the original game. However, we find a much lower magnitude of in-group favouritism in this additional treatment which indicates that in-group favouritism behaviour observed in our original design is not driven by generalised reciprocity.

⁵We use lower case letters for the options in the choice set to prevent confusion with the group assignments A and B.

⁶The highest amount that each subject can get from the choice set is 4,500 tokens which is approximately 3.15 Pounds and the lowest amount is -1,500 which is -1.05 Pounds. However, none of the subjects leaves with a negative payoff since they all receive an initial endowment of 3,000 tokens.

Figure 1: The Choice Set

Decision Tasks								
Please complete the following two tasks:								
Task 1: Please make only decision.								
Task 2: Please rate how much you LIKE or DISLIKE for EACH of the choices.								
To complete Task 1: Type an X (it does not matter if upper or lower case) in the box corresponding to the choice that you want to choose. To complete Task 2: Type a number from 1 (DISLIKE VERY MUCH) 2 (DISLIKE) 3 (LIKE) to 4 (LIKE VERY MUCH) or 0 (INDIFFERENT) for EACH choice.								
ALLOCATION CHOICE	A	B	C	D	E	F	G	H
Each member of YOUR GROUP	4500	3000	2000	1500	1000	0	-1500	0
Each member of THE OTHER GROUP	-1500	0	1000	1500	2000	3000	4500	0
TASK 1: YOUR DECISION (choose only ONE) PLEASE REMEMBER THAT YOU ONLY MAKE THIS DECISION ONCE.								
TASK 2: RATE. Please indicate how much you like EACH choice by rating from 1 (DISLIKE VERY MUCH) 2 (DISLIKE) 3 (LIKE) to 4 (LIKE VERY MUCH) or 0 (INDIFFERENT).								
After you have made your decision, please confirm your entries by clicking the OK button. PLEASE REMEMBER THAT YOU WILL ONLY GET TO MAKE THIS DECISION ONCE (there will be no more round).								
								OK

2.2 The Punishment Treatments

We test the existence of social norms by looking at whether there is any change in the allocators' behaviours when they are faced with the threats of costly punishments. In all punishment treatments, all subjects were informed that the experiment consists of two stages: (i) the '*decision stage*' and (ii) the '*punishment stage*'. In stage two, after the allocator is randomly selected, the decision is revealed to the punishers⁷. It is important to note that only one type of punishment is implemented at a time.

In the *in-group punishment treatment (T2)*, the punishers are the members of group A (in-group members) who are not selected to be the allocator. At a cost of 100 tokens, they can each deduct the allocator's payoff by 500 tokens⁸. The allocator's payoff can be written as: $\pi_{allocator} = 7,500 - \sum_{n=1}^3 P_{in-group}$, where 7,500 tokens is made up of 3,000 tokens initial endowment and 4,500 tokens fixed payment from the experimenter for being selected to be the allocator. $P_{in-group}$ is the punishment imposed by each of the other in-group members. The payoff of the in-group punishers can be written as: $\pi_{in-group} = 3,000 + T - C$, where T is the amount of tokens distributed to them by the allocator and C is the cost of punishment. In this treatment, the out-group members are just passive players and thus, their payoffs are determined solely by the choice of the allocator.

In the *out-group punishment treatment (T3)*, the allocator's decision is revealed to the out-group members (group B) who can punish the allocator using the same cost to punishment ratio as in T2, whilst in the *third-party punishment treatment (T4)* it is revealed to an independent third-party punisher who is given an endowment of 4,500 Tokens in the second stage of the experiment which she can spend on punishment. The third-party's endowment is equal to the allocator's fixed payoff. This is to ensure that

⁷Only the decision but not the personal identity of the allocator is revealed to the other participants.

⁸Neutral frame is used in all instructions. No loaded words such as 'punish', 'sanction' are used, instead the word 'deduct' is used.

the third-party does not decide to punish the allocator because of (disadvantageous) inequity aversion (Fehr and Schmidt, 1999). The cost of punishment for the third-party is as follows: for a cost of 100 tokens, the third-party punisher can deduct 300 tokens from the allocator’s payoff, which is similar to the cost to punishment ratio used in previous studies (Fehr and Fischbacher, 2004b; Bernhard *et al.*, 2006a,b). The slightly higher cost for the third-party punisher compared to the in-group and out-group punishment treatments also reflects the fact that in practice, due to opportunity cost and the risk of retaliation, it is more costly for an independent person to get involved in a conflict between two other parties (or groups). Moreover, the higher cost will ensure that the third-party punisher has to feel sufficiently strong about the norm violation in order to punish the allocator.

Whether or not the allocators’ decisions would be affected by the threats of punishments depends on their perceived likelihood. If the subjects only care about own money, a threat of punishment is not credible in a one-shot game since the punishment is costly and there is no future benefit for the punishers. Therefore, we should not observe any change in the allocator’s behaviour in any of the punishment treatments compared to the baseline. If we do observe changes in the allocators’ behaviours, it would indicate that a social norm exists and is expected to be enforced. That is, the allocators expect that they would be punished if they did not change their behaviour. The direction of the change in the allocators’ behaviours will depend on the type of norm that is expected to be enforced.

2.3 The Subject Pools

The experiment was carried out in April 2009 at the University of Cambridge and University of Nottingham in the UK⁹. The experiment was administered by z-Tree software (Fischbacher, 2007) and a reward medium of experimental monetary unit called ‘token’ was used. The tokens were converted into cash and paid to each subject in private immediately after the experiment ended together with an on-time show-up fee of 3 Pounds. Finally, independence of the observations was ensured by using the between subject design approach. In all sessions, after the subjects were seated, each was given a written instruction which explained what they would be asked to do in the experiment. A summary of the instruction in the baseline treatment for group A (the in-group) is as follows:

Members of Group A

If you are assigned to be in GROUP A, you will be asked to complete the following TWO tasks.

TASK 1: In task 1, on the screen you will see a set of eight different allocation options (an example of the allocation options is shown below). Each option allocates Tokens between your fellow GROUP A’s members (excluding you) and the members of GROUP B. To indicate your decision, mark an ‘x’ (it does not matter whether it’s lower or upper case) in the box under the option that you would like to choose. Please choose only ONE option and please remember that you will only make this decision ONCE.

TASK 2: In task 2, after you have made the decision, you are asked to RATE EACH option using the following scale: 1 (Dislike very much), 2 (Dislike), 3 (Like), 4 (Like very much), or 0 (Indifferent).

The rating in Task 2 provided a crude robustness check that subjects chose the option that they actually preferred (like) and not because they think an option was a ‘correct’ one to choose (experimenter’s demand) even if they did not prefer that option. They were also clearly informed of the randomised

⁹The reason for running the experiments in two locations was purely due to the logistics and the availability of experimental labs since two authors were based in Cambridge and one was based in Nottingham.

method by which the allocator was selected:

You and your fellow GROUP A members will each make the decision, but only ONE of the four decisions will be randomly selected. Once all the GROUP A members have completed both tasks and confirmed their decisions and ratings by clicking the OK button, the administrator will throw a dice which will determine whose decision will be selected:

If the dice shows number 1, the decision of GROUP A's member A1 will be selected.

If the dice shows number 2, the decision of GROUP A's member A2 will be selected.

If the dice shows number 3, the decision of GROUP A's member A3 will be selected.

If the dice shows number 4, the decision of GROUP A's member A4 will be selected.

If the dice shows numbers 5 or 6, no decision will be selected and the dice will be thrown again until it shows the numbers between 1 and 4.

The instruction also included explanations of what members of group B (the out-group) were asked to do. Whilst the members of group A were making the allocation decisions, we elicited beliefs of the members of group B by asking them what they thought the allocator would choose. We also asked them to make a 'hypothetical' allocation decision if there were to be assigned as an allocator:

Members of Group B

If you are assigned to be in GROUP B, you will be asked to complete TWO tasks:

TASK 1: You will be asked to select an option which you think the decision-maker in GROUP A is most likely to choose. To indicate your decision, mark an 'x' (it does not matter whether it's lower or upper case) in the box under the option that you think the decision-maker in GROUP A is most likely to choose. Please choose only ONE option and you can only make this decision ONCE.

TASK 2: Suppose that you were in the position to decide how to allocate the Tokens between the members of your group (GROUP B) and the other group (GROUP A). Which option would you choose? To indicate your decision, mark an 'x' (it does not matter whether it's lower or upper case) in the box under the option you would like to choose. Please choose only ONE option.

The tasks for both groups are common information. Members of group A know what group B are asked to do and vice versa. They are also told explicitly how their payoffs are calculated. For the punishment treatments, the first stage was identical to the baseline treatment and the following instruction was given for the second stage:

(In-group punishment treatment)

In the second stage, the chosen decision will be shown to the other three members of Group A whose decisions are not chosen (but the identity of the Group A's member whose decision is chosen or the 'decision-maker' will be kept confidential). Each of the other three members of Group A will be asked to decide whether to:

A) Deduct the payoff of the Group A member whose decision is chosen (the decision-maker) at their own cost;
OR

B) Do nothing (Leave the box BLANK).

If the other three members of Group A choose option (A), each will have to individually decide which of the following deduction strategies they would like to impose on the decision-maker:

1) Deduct 500 Tokens of the decision-maker's payoff, at a cost of 100 Tokens.

2) Deduct 1,000 Tokens of the decision-maker's payoff, at a cost of 200 Tokens.

3) Deduct 1,500 Tokens of the decision-maker’s payoff, at a cost of 300 Tokens.

The subjects were given approximately 10-15 minutes to read the instruction and then a summary of the instruction was read out loud to the group. Each subject made the decision in private on the computer screen behind a partition. Once everyone in the session had made their decisions, they were asked to complete a post-experimental questionnaire on their computer screen. Each session lasted approximately 40 minutes and each subject was paid privately in cash at their seat in exchange for a receipt confirming the amount paid and their participation in the experiment. The average payment was 8 Pounds in addition to the show-up fee.

3 The Results

A total of 359 subjects took part in the experiment, most of which were undergraduate students (the average age is 20) randomly selected from different faculties recruited via a self-recruiting system, ORSEE (Greiner, 2004). As a manipulation check, we used a psychometric test called ‘*Inclusion of Other in the Self Scale* (IOS scale)’ (Aron *et al.*, 1992; Cialdini *et al.*, 1997) to examine whether the randomised group assignment method that we used was sufficient to create group identity. The IOS scale asked the subjects to identify how ‘close’ they felt towards the in-group and the out-group members (1 = not overlap/not close, 7 = very overlap/very close)(Aron *et al.*, 1992; Cialdini *et al.*, 1997). Subjects indicated that those who shared the same letter (A or B) were *closer* to them than those who did not. The difference is statistically significant (Wilcoxon signed-rank test, $z = 3.102, p - value = 0.0019$). Therefore, our method of simply labelling subjects as group A and group B appeared to create different group affiliations¹⁰. Table 1 presents summary statistics¹¹.

Note: Standard deviations in parentheses

RESULT 1: The majority of the allocators favoured their own group in the baseline treatment without punishment.

Sixty-three subjects took part in the baseline treatment. Eighty-one percent chose in-group favouring options (A, B or C) and about half of which chose option *a* which deducts 1,500 tokens from each of the out-group members. Only 19% chose the equal distribution option and none chose to favour the out-group. Our results are in line with previous findings that people have a tendency to favour their own group at a cost to the out-group, even when the notion of a group has no real social meaning. Our results also provide new evidence that in-group favouritism occurs when the allocation decisions are concerned with more than just one member in each group.

Out-group’s Beliefs (Baseline)

Sixty-three percent of the out-group members believed that the allocators would (on average) choose

¹⁰However, we do not think that our results are due to experimenter’s demand effect. This is because we also ran an additional treatment with two members in group A and one member in group B using the same labelling method, but we did not observe the same magnitude of in-group favouritism behaviour. In fact, in that treatment the majority of the allocators chose to split the money equally between the two groups.

¹¹We used all the allocation decisions from all members of group A (not just those who were selected to be the allocator) in the analysis.

Table 1: Summary Statistics: All Treatments

Variables	Baseline	In-group Punish	Out-group Punish	Third-Party Punish
Total number of obs.	63	84	84	128
Male	35%	55%	65%	52%
Average Age	21	20	20	20
Undergraduate	87%	98%	100%	100%
Study Economics	14%	14%	15%	16%
IOS Scale (in-group)	2.40 [1.49]	2.32 [1.80]	2.58 [1.62]	2.40 [1.71]
IOS Scale (out-group)	1.94 [1.52]	1.65 [1.38]	2.17 [1.60]	2.00 1.60
%In-group favouritism (A, B, C)	81%	85%	52%	84%
Equal split allocation (D)	19%	13%	38%	16%
Out-group favouritism (E, F, G)	0%	0%	10%	0%
No Allocation (H)	2%	0%	0%	0%
Out-group Beliefs of in-group favouritism	85%	97%	89%	88%
Out-group Hypothetical Choice to favour own group	56%	78%	58%	73%
Decided NOT to punish (% of total)		92%	61%	94%
Punished in-group favouritism (A,B,C)		5.6%	11%	6%
Punished equal split (D)		2.4%	14%	7%

to favour their own group which coincided with the actual behaviour of the allocators in the baseline treatment. In addition, we also asked each of the out-group members to make a hypothetical allocation decision if they were to be assigned the role of the allocator. We find that 56% of the out-group members would also favour their own group if they were given the opportunity (even though this was only a hypothetical choice)¹².

RESULT 2: The threat of in-group punishment slightly increased in-group favouritism behaviour

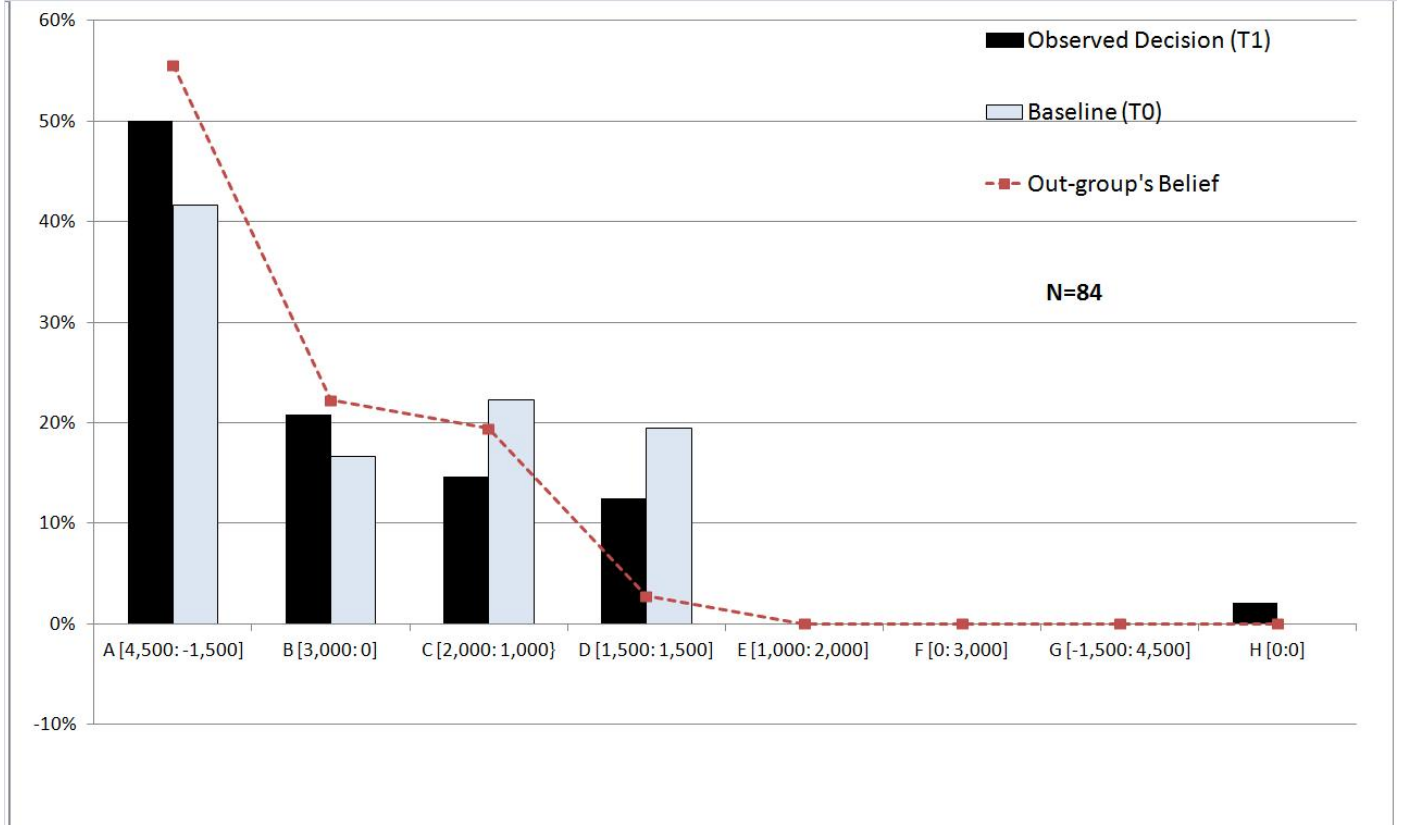
The threat of punishment from the other in-group members increased the proportion of in-group favouring options from 81% in the baseline to 85%. There was also a small drop in the equal distribution option (option *d*) from 19% to 13%. However, these changes were not statistically significant (although this could be due to the ‘*ceiling effect*’ since the proportion of in-group favouritism options was already above 80% in the baseline).

¹²It is important to note that the hypothetical decision was not influenced by the allocator’s decision in any way since the out-group members were making the hypothetical decisions at the same time as the in-group members were making the allocation decisions. Therefore, they did not know the chosen decision whilst making the hypothetical choices.

Out-group's Beliefs in In-group Punishment Treatment

Almost all of the out-group members (97%) expected that the allocators would favour their own group. The proportion was significantly higher than in the baseline. When the out-group members were asked to make a hypothetical allocation choice in this treatment, the majority (78%) also chose to favour their own group.

Figure 2: Allocation Decisions and Out-groups' Beliefs: *In-group Punishment* Treatment vs Baseline



Actual In-Group Punishment Behaviours

In the second stage of the game, punishment opportunity was given to the other in-group members who were not selected to be the allocator. Almost all of the in-group punishers (92%) chose not to punish the allocators. Only three in-group punishers decided to punish. Of which, two people punished options *b* and *c* by deducting 500 and 1,500 tokens respectively as shown in Table 1. It is interesting to note that none of the allocators who chose option *a* was punished. We used direct-response method rather than the strategy method in the punishment stage. In direct-response method, subjects make decisions whenever it is their time to do (which makes the decision ‘hot’ in a psychological sense), whilst in strategy method subjects make contingent decisions for all nodes at which they may have to play (Brandts and Charness, 2011). We used the direct response method because we wanted subjects to make

‘actual’ punishment decisions rather than considering all possible scenarios. In all three punishment treatments, the punishment decisions took place only after the allocator’s decision was revealed to the punishers, which could explain why only a few punishment decisions were observed. If the allocators expected to be punished, they would already change their behaviours in the decision stage depending on their expectations about the punishment. For example, in the in-group punishment treatment, if the allocators expected to be punished when they do not favour their own group, they would be more likely to choose options *a*, *b*, or *c*. As a result, there was no need for the in-group punishers to implement punishment in the second stage. Since we are interested in the extent to which the *threat* of punishment could change the allocators’ behaviour compared to the baseline in a one-shot game, it did not affect our results whether or not the actual punishment took place or not.

So far, we have observed prevalent in-group favouritism behaviour in the baseline with no punishment opportunity and that the threat of in-group punishment had a positive effect on this behaviour. It seems that in-group favouritism was *expected* amongst the in-group members such that the allocators expected to be punished if they did not favour their own group. Our results showed that the threat of punishment sufficed to sustain (and even increase) in-group favouritism behaviour, indicating that ‘Parochialism’ or a preference for favouring one’s own group (Bernhard *et al.*, 2006b) seemed to be a norm which was enforced within the in-group.

RESULT 3: The threat of out-group punishment reduced in-group favouritism behaviour

Once the out-group members were given the punishment power, the proportion the allocators who chose in-group favouring options (*a*, *b*, or *c*) significantly decreased from 81% in the baseline to 52% in this treatment. At the same time, the proportion of the equal distribution option (*d*) increased from 19% to 38% (Two-sample Wilcoxon rank-sum test [H0: Choices (baseline) = Choices (out-group punishment)]: $z = -2.405$, $Prob > |z| = 0.0162$). Our results indicate that the allocators expected to be punished by the out-group and thus, shifted their decisions toward the equal distribution option (*d*). Furthermore, the threat of out-group punishment also led to some *out-group favouritism* choices among the in-group members. Ten percent of the in-group members chose to favour the out-group and of which, 4% went as far as giving the maximum payoff of 4,500 Tokens to the out-group (option *g*) at the expense of the in-group members! (this proportion was, of course, a very small proportion of the total decisions).

Out-group’s Beliefs in Out-group Punishment Treatment

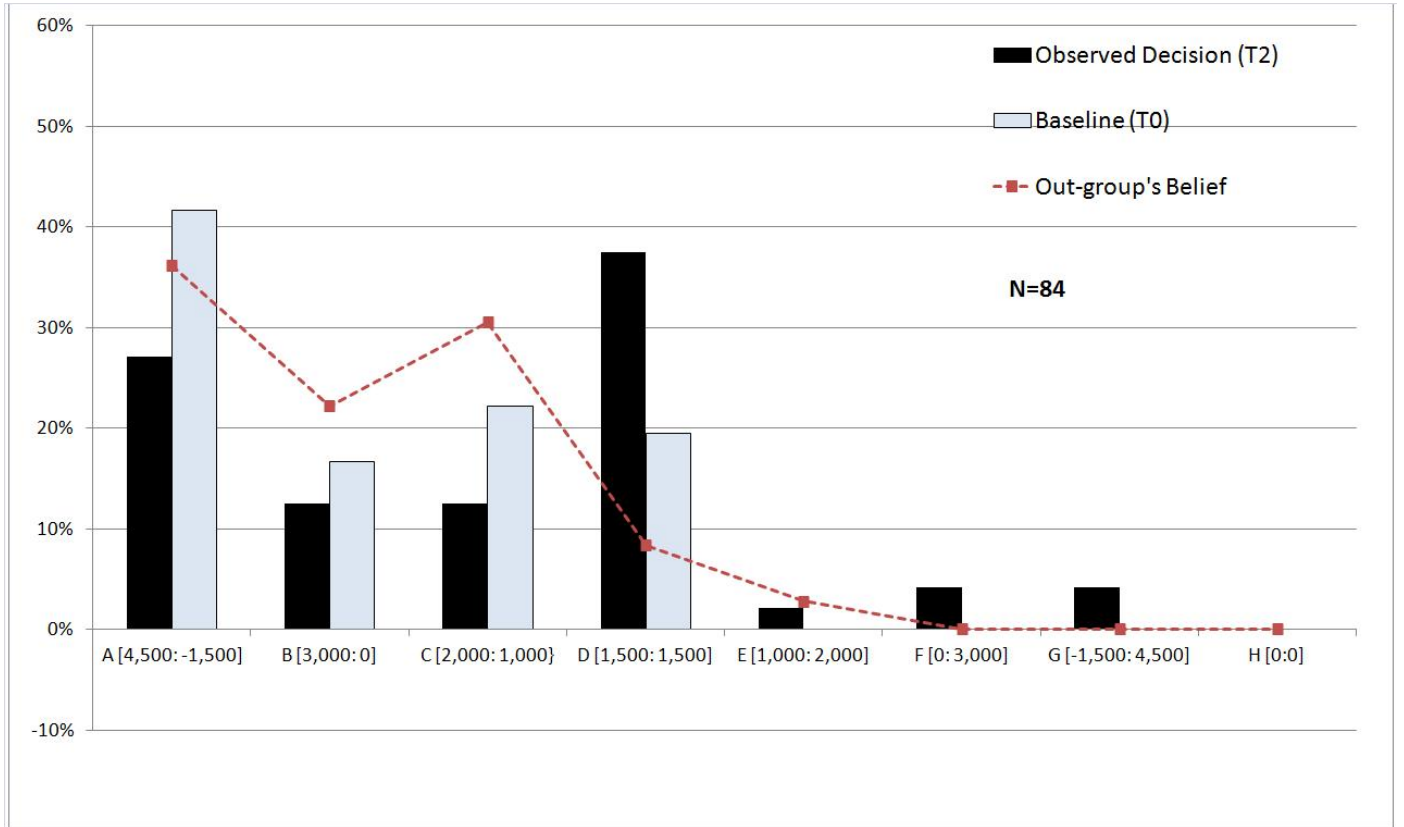
When asked what they thought the allocators would do (task 1 for members of group B described above), the majority of the out-group members (89%) thought that the allocators would still favour their own group. The high expectation of in-group favouritism by the out-group members was somewhat surprising since we observed that the threat of out-group punishment effectively reduced in-group favouring behaviour. If the out-group members considered in-group favouritism to be unacceptable, since they had the power to punish the allocators and the allocators also knew this, why would they still think that the majority of the allocators would favour their own group? It is possible that there may be a general expectation that most people will choose to favour their own group when making an allocation decision such as the one in our game (i.e. allocating a fixed amount of resource) even when there is a probability that they would be punished. Further study is needed to explore this interpretation, however. One idea is to ask the out-group members directly for their reasons for expecting the

allocators to favour their own group as well as eliciting the allocators' beliefs about what the out-group members would do. Interestingly, when asked to make a hypothetical allocation decisions, a smaller proportions of the out-group members now chose to favour their own group (58%) compared to the in-group punishment treatment (78%) although it was slightly higher than the baseline (56%). The rest of the out-group members chose to distribute the endowment equally between the two groups.

Actual Out-Group Punishment Behaviours

Contrary to the in-group punishment treatment, the majority of the out-group members (61%) chose to *punish* the allocators who chose options *a*, *b* or *c* and most of them chose the highest punishment level by deducting 1,500 tokens from the allocator's payoff. This result supports the allocators' behaviour i.e. that they expected the out-group to punish them if they favoured their own group.

Figure 3: Allocation Decisions and Out-groups' Beliefs for the *out-group Punishment* Treatment vs the Baseline



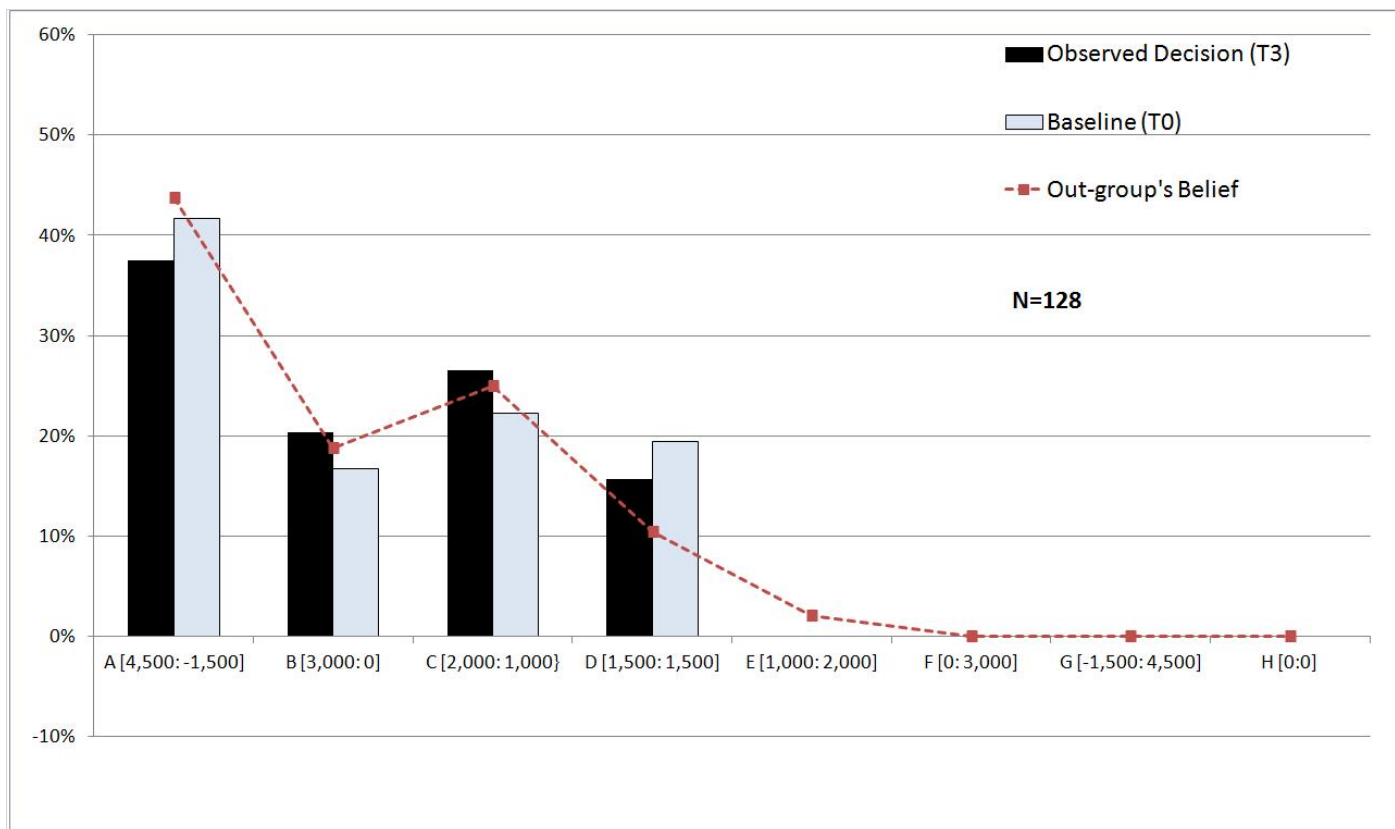
RESULT 4: The threat of punishment from an independent third-party significantly did not have any significant impact on the allocation decisions.

Since the third-party belongs to neither the in-group nor the out-group and thus, they are not directly affected by the allocator's decision, her decision to impose a costly punishment on the allocator would indicate the willingness to enforce a social norm. If the threat of third-party punishment is perceived to be credible, we expect to see a change in the allocator's behaviour - either increase or decrease in-group

favouritism behaviour, depending on what kind of norm is being enforced.

We observed a slight increase in the proportion of in-group favouring options amongst the allocators in the presence of the third-party punisher, from 81% in the baseline to 84% in this treatment. The treatment effect was not statistically significant compared to the baseline¹³. Our results show that without the threat of punishment, in-group favouritism seemed to be a common behaviour among the allocators. But when the punishment power was given to the out-group members, the allocators expected to be punished if they favoured their own group and thus, chose the equal distribution option instead, which implied that the allocators believed that this behaviour would be considered as unacceptable by the out-group members. However, since the threat of punishment from the third-party did not have any effect on this behaviour, it seems that only when the subjects were directly (and negatively) affected by in-group favouritism i.e. when they belonged to the out-group that they were willing to retaliate and punish the allocators. However, when they were not directly involved (third-party), they did not see this behaviour as a norm violation and thus, did not punish the allocators. The allocators also knew this and thus, did not change their behaviour (the threat of third-party punishment was not credible).

Figure 4: Allocation Decisions and the Out-groups' Beliefs for the *third-Party Punishment* Treatment



Out-group's Beliefs in the Third-Party Punishment Treatment

The majority of the out-group members expected that on average the allocators would choose to favour their own group (88%), perhaps indicating to some extent that the out-group members did not

¹³Although this could also be due to the 'ceiling effect' as mentioned previously.

expect the threat of third-party punishment to be credible. Finally, when they were asked to make a hypothetical allocation choice, 73% of the out-group members also chose to favour their own group, which were similar to those in the in-group punishment treatment.

Actual Third-Party Punishment Behaviours

Almost all the third-party punishers (94%) decided *not* to punish the allocator and thus, it is not possible to draw inference from this result. The post-experimental feedback revealed that most of the third-party punishers were *not* willing to incur a cost in order to punish the allocators, which was contrary to the cases observed within the context of cooperation norm in the public goods (Fehr and Gaechter, 2000; Fehr *et al.*, 2002) and the norm of egalitarianism in ultimatum game and prisoner’s dilemma game (Bernhard *et al.*, 2006a; Goette *et al.*, 2006).

4 Discussions and Concluding Remarks

Despite the now vast literature on the effect of group identity on individual decisions, little attention has been given to the extent to which the threat of costly punishment affects in-group favouritism behaviour. This question is important to the general audience of the economic discipline because in-group favouritism distorts decisions and limits access to resources only to those who are part of the in-group and thus, it is important to identify whether there is a punishment mechanism which could deter the natural bias towards one’s own group.

We first replicated the previous finding and showed that people have a natural tendency to favour their own group. In addition, we provide new evidence that this behaviour is observed when the allocation decision is concerned with two groups three individuals rather than just one person in each group. A question may be raised about whether our results are influenced by experimenter’s demand effect since we deliberately told participants that they belonged to different groups (A and B) and then asked the members of one group to divide a fixed sum of money. As a result, they might behave in a way which they thought was expected of them. However, we believe that if this was the case, we would be more likely to observe the ‘fair’ allocation option which would be more likely to be, as most would expect, the ‘correct’ thing to do. In addition, experimenter’s demand effect could not have a strong effect on our results in the punishment treatments which are the main focus of this paper. This is because it would be difficult for the participants to ‘second-guess’ what the experimenter expected them to do. Consequently, their decisions in the punishment treatments would be mainly influenced by the threat of punishment rather than experimenter’s demand. Our results indicate that the effects of the threat of costly punishment on in-group favouritism behaviour is not that straight forward. We find that without the threat of punishment the allocators favoured their own group regardless of the costs imposed on the out-group. The allocators also expected the in-group members to enforce ‘Parochial norm’ or in-group biased norm which, in turn, resulted in an increase in in-group favouritism behaviour. The allocators, however, refrained from favouring their own group when those who were directly affected by their decisions (the out-group) could punish them. Finally, unlike opportunistic and selfish behaviour which has been shown to be considered as a violation of cooperative norm and is punished by an independent third-party in a one-shot game and even when punishment is costly (Fehr and Gaechter, 2000; Fehr *et al.*, 2002; Carpenter and Matthews, 2009), in-group favouritism did not trigger the same reaction by the third-party who were not directly affected by this behaviour.

Like corruption, in-group favouritism is usually carried out behind closed door and thus, it is difficult to clearly identify when it is considered as acceptable or not, which is why laboratory experiment is a useful tool to study in-group favouritism behaviour. Based on our results, we conclude that people only consider this behaviour as unacceptable when they are *not* part of the in-group. But when they are part of the in-group, they expect the group members to treat them more favourably and when they are not directly involved, they do not particularly care about this behaviour. Our study has an important welfare implication, particularly in terms of income distribution. According to the Second Fundamental Theorem of Welfare Economics, every Pareto efficient allocation can be achieved as a competitive equilibrium with *appropriate transfer* of wealth (usually assumed to be carried out by a social planner or in practice a public official). But what constitutes ‘appropriate transfer’ if efficiency is not a concern? When two groups with equal level of productivity compete for scarce resource, should a social planner (or the allocator in our experimental context) allocate the resource such that the final distribution yields an equitable outcome across groups? What happens if he chooses to favour his own group, would he be punished? We believe that our paper has paved way for future research on in-group favouritism and the nature of social norms which govern this behaviour and the simple design of our game is easy to replicate in many other subject pools and the choice set can be modified to test the effect of the stake size, the loss incurred by the out-group, the cost incurred to the allocator (there is no cost for the allocator to favour her own group in our current version of the experiment), and different ways to induce group identity.

References

- Andreoni J. (2007) ‘Giving gifts to groups: How altruism depends on the number of recipients’. *Journal of Public Economics*, 91, pp. 1731–1749.
- Aron A., Aron E. and Smollan D. (1992) ‘Inclusion of Other in the Self Scale and the structure of interpersonal closeness’. *Journal of Personality and Social Psychology*, 63, pp. 596–612.
- Bandiera O., Barankay I. and Rasul I. (2006) ‘The Evolution of Cooperative Norms: Evidence from a Natural Field Experiment’. *Advances in Economic Analysis and Policy*, 6(2).
- Becker G. (1957) *The Economics of Discrimination*. The University of Chicago Press.
- Ben-Ner A., McCall B.P., Stephane M. and Wang H. (2009) ‘Identity and in-group/out-group differentiation in work and giving behaviors: Experimental evidence’. *Journal of Economic Behavior and Organization*, 72, pp. 153–170.
- Bernhard H., Fehr E. and Fischbacher U. (2006a) ‘Group Affiliation and Altruistic Norm Enforcement’. *American Economic Review*, 96(2), pp. 217–221.
- Bernhard H., Fischbacher U. and Fehr E. (2006b) ‘Parochial altruism in humans’. *Nature*, 442, pp. 912–915.
- Billig M. and Tajfel H. (1973) ‘Social categorization and similarity in intergroup behaviour’. *European Journal of Social Psychology*, 3, pp. 27–51.
- Bolton G. and Ockenfels A. (2000) ‘A Theory of Equity, Reciprocity, and Competition’. *American Economic Review*, 90(1), pp. 166–193.

- Brandts J. and Charness G. (2011) ‘The strategy versus the direct-response method: a first survey of experimental comparisons’. *Experimental Economics*, 14, p. 375398.
- Brewer M. (1979) ‘Intragroup bias in the minimal intergroup situation: a cognitive-motivational analysis’. *Psychological Bulletin*, 86, pp. 307–324.
- Carpenter J. and Matthews P. (2009) ‘What Norms Trigger Punishment’. *Experimental Economics*, 12(3), pp. 272–288.
- Charness G., Rigotti L. and Rustichini A. (2007) ‘Individual Behaviour and Group Membership’. *American Economic Review*, 97(4), pp. 1340–1352.
- Chen Y. and Li X. (2009) ‘Group Identity and Social Preferences’. *American Economic Review*, 99(1), pp. 431–457.
- Cialdini R., Brown S., Lewis B., Luce C. and Neuber S. (1997) ‘Reinterpreting the Empathy - Altruism Relationship: When one Into One Equals Oneness’. *Journal of Personality and Social Psychology*, 73(3), pp. 481–494.
- Elster J. (1989) *The Cement of Society: A Study of Social Order*. Cambridge University Press.
- Fehr E. and Fischbacher U. (2004a) ‘Social norms and human cooperation’. *TRENDS in Cognitive Sciences*, 8(4), pp. 184–189.
- Fehr E. and Fischbacher U. (2004b) ‘Third-party punishment and social norms.’ *Evolution and Human Behaviour*, 25, pp. 63–87.
- Fehr E., Fischbacher U. and Gaechter S. (2002) ‘Strong reciprocity, human cooperation, and the enforcement of social norms’. *Human Nature*, 13.
- Fehr E. and Gaechter S. (2000) ‘Fairness and Retaliation: The Economics of Reciprocity’. *Journal of Economic Perspectives*, 14(3), pp. 159–181.
- Fehr E. and Schmidt F. (1999) ‘A Theory of Fairness, Competition, and Cooperation’. *Quarterly Journal of Economics*, 114(3), pp. 817–868.
- Fischbacher U. (2007) ‘z-Tree: Zurich Toolbox for Ready-made Economic Experiments’. *Experimental Economics*, 10(2), pp. 171–178.
- Gaube T. (2000) ‘Group size and free riding when private and public goods are gross substitutes’. *Bonn Economic Discussion Papers* 13/2000.
- Goeree J., Holt C. and Laury S. (2002) ‘Private costs and public benefits: unraveling the effects of altruism and noisy behavior’. *Journal of Public Economics*, 83(2), pp. 255–276.
- Goette L., Huffman D. and Meier S. (2006) ‘The Impact of Group Membership on Cooperation and Norm Enforcement: Evidence using Random Assignment to Real Social Groups’. *American Economic Review*, 96(2), p. 212216.
- Greiner K. (2004) ‘An Online Recruitment System for Economic Experiments.’ *MPRA Paper, University Library of Munich, Germany*.
- Guth W., Levati M. and Ploner M. (2008) ‘Social Identity and Trust: An Experimental Investigation’. *Journal of Socio-Economics*, 37(4), pp. 1293–1308.

- Hoff K., Kshetramade M. and Fehr E. (2007) 'Norm Enforcement under Social Discrimination: Experimental Evidence from India'. *Centre for International and Security Studies at Maryland (CISSM) Working Paper*.
- Isaac M., Walker J. and Williams A. (1994) 'Group size and the voluntary provision of public goods: experimental evidence utilizing large groups'. *Journal of Public Economics*, 54(1), pp. 1–36.
- Naegelen F. and Mougeot M. (1998) 'Discriminatory public procurement and cost reduction incentives'. *Journal of Public Economics*, 67, pp. 349–367.
- Stahl D. and Haruvy E. (2006) 'Other-regarding preferences: Egalitarian warm glow, empathy, and group size'. *Journal of Behavior and Organization*, 61, pp. 20–41.
- Tajfel H. and Turner J. (1986) 'The social identity theory of intergroup behavior'. In S. Worchel and L. Austin, eds., 'Psychology of Intergroup Relations', Chicago: Nelson-Hall.
- Tajfel J., Billig M., Bundy R. and Flament C. (1971) 'Social Categorization in Intergroup Behaviour'. *European Journal of Social Psychology*, 1, pp. 149–178.
- Yamagishi T. and Kiyonari T. (2000) 'The Group as the Container of Generalized Reciprocity'. *Social Psychology Quarterly*, 63(2), pp. 161–197.
- Zantman W. (2002) 'Constitutional design and regional favouritism'. *Journal of Public Economics*, 4(1), pp. 71–93.