University of Essex
Department of Economics

Economic Experiments on Tax Compliance, Harassment Bribes, and Diverse Committees

In partial fulfilment of the requirements for the
Doctorate of Philosophy in

ECONOMICS

Submitted by:
Zia Ul Hassan khan

Supervisor(s):
Professor Friederike Mengel
Dr. Ahrash Dianat

Colchester, April 2022

Date of Submission for Examination: 27/01/2022

Date of Resubmission for Examination (Revised): 25/04/2022

Declaration

I, Zia Ul Hassan, hereby declare that the work in this thesis titled "*Economic Experiments on Tax Compliance, Harassment Bribes, and Diverse Committees*" is a result of my work and collaborations with other researchers.

- Chapter 1, *An Experimental Study on the Role of Social Norms About Tax Compliance*, is sole authorship.
- Chapter 2, *Incentives for whistle-blowing: An Experiment*, is sole authorship.
- Chapter 3 , *Diverse Committees*, is co-authored with Prof. Friederike Mengel and Dr. Niall Hughes.

The thesis is based on experiments at the ESSEXLab, an experimental social sciences lab at the University of Essex, United Kingdom.
No part of this thesis has been submitted elsewhere for any other degree or qualification, and it is all my work unless referenced, to the contrary, in the text.

Dedication

My parents and all other parents who sacrifice their bodies, minds, and souls in providing and caring for their children. The ink of scholars is a product of their sweat and blood. I love you *Anna* and *Papa*.

Abstract

This Doctoral thesis includes three chapters. In Chapter 1, *An Experimental Study on the Role of Social Norms About Tax Compliance*, we study which social norms shape taxpayers' avoidance behaviour. We look at the direct effect on tax avoidance and evasion and how the two interact with social norms and rule-following concerns. We also experimentally manage empirical and normative expectations to look at their role in determining tax avoidance. In Chapter 2, *Incentives for whistle-blowing: An Experiment*, we investigate the effectiveness of financial rewards in encouraging optimal whistle-blowing. To do so, we conduct a laboratory experiment that features a two-player bribery game between a public official and a citizen. First, the public official decides whether (and how much) to ask for a bribe. Then, the citizen jointly decides whether to pay the bribe and whether to report the public official for corruption. We use a between-subject experimental design that varies the accuracy of the judicial system (defined as the probability of type-1 and type-2 errors) and the evidence threshold required for proving guilt (defined as the cost of reporting). In Chapter 3, *Diverse Committees*, we investigate the potential of diversity to influence committee decision-making. We present a model in which committee members receive private information on a state of the world, deliberate and vote. Committee members belong to one of two groups which may differ along two dimensions: (a) their preferences and (b) their information structures. We test the model's key predictions in a laboratory experiment.

Acknowledgements

All praises and thanks to **ALLAH** Almighty for everything!
I want to thank my supervisors, *Friederike* and *Ahrash*, for providing guidance and encouragement throughout the doctoral degree. I convey sincere gratitude to the ESSEXLab, specifically to *Dr Patrick Lown* and late *Dr Patrick Nolan*, for the fantastic support given throughout the PhD program. I am also humbled by the support provided by the Economic and Social Research Council and the British Academy. I am grateful to my colleagues, *Eike, Elisabetta, and Ludovica,* for their continued guidance and support for my research. I also want to remember all the teachers I have had in life, inside or outside the classroom.
I want to mention my late uncle, *Dr Raza*, who always saw the good in me and inspired me to be a good human being. This journey was never possible without my parents' belief and relentless sacrifices, especially my mother's. Finally, I am blessed to have shared this journey with three amazing women - *Zareen, Zainab, and Ayesha*. They have been the beacon of hope for me during moments of severe stress.

Table of Contents

List of Figures

List of Tables

Chapter 1

An Experimental Study on the Role of Social Norms About Tax Compliance

Abstract

This paper aims to study which social norms shape taxpayers' avoidance behaviour. We look at the direct effect on tax avoidance and how it interacts with social norms and rule-following concerns. We find a weak-to-modest correlation between rule-following and avoidance. Even after controlling for treatment, risk preference, demographics, and session fixed effects, these effects persist. We find no influence on tax avoidance of how justified the redistribution through taxation seems to be. We experimentally manage empirical and normative expectations and see a strong effect of expectations about others' behaviour, indicating a strong conditional cooperative element in tax avoidance behaviour. When we provide subjects with conflicting empirical and normative expectations, the former seems to dominate the latter.

Keywords: Tax-compliance, Social Information, Social Norms.
JEL: C91, D9, H26

1.1 Introduction

Tax compliance remains a topic of significant import for researchers as well as policymakers. The most prominent reason is the cost of non-compliance for governments in keeping societies functional. Tax non-compliance comes in two different forms. Evasion is a deliberate concealing of income from the tax authorities and therefore is outright illegal. On the other hand, tax avoidance involves flexing the tax rules to gain a tax advantage that the Parliament never intended. It is often based on phoney transactions that serve little or no other purpose than reducing the tax burden. Most of these transactions follow the law in letter but not in its spirit.

Tackling tax avoidance is one of the most critical challenges for governments worldwide. We can gauge the economic significance of tax avoidance from the reported losses of £13.7 billion to the UK economy and €150 billion a year to members of the European Union (Murphy, 2012).[1] In the UK, taxes avoided on personal income, national insurance contributions, capital gains, and corporations contribute most to the tax gap. One of the challenges is how to target avoidance without pushing the taxpayer to other channels of the shadow economy. In this respect, the success of a fair and efficient tax system in achieving its revenue targets depends on enforcement and the acceptance of tax laws (Slemrod, 2016). A substantial amount of research shows that enforcement alone cannot account for high rates of tax compliance (Alm et al., 1992; Slemrod, 1985). Acceptance plays a crucial role through its two complementary aspects, i.e. the perception of existing norms about taxes and the manifestation of this perception in the actual decision of filing tax returns. Various experiments and surveys have found that the more equitable and fair the perception of the tax system, the higher the rate of compliance (Becker et al., 1987; Bordignon, 1993; Andreoni et al., 1998; Kirchler, 2007).

In this paper, we first show that there exists a (perceived) norm against tax avoidance.[2] We elicit the norm using the task by Krupka and Weber (2013). We also look at the agreement among subjects regarding the norm. Although the content of the norm is clear, there is a considerable lack of consensus among subjects about norm adherence. We use an abstract rule-following task (Kimbrough and Vostroknutov, 2018) to measure an individual's sensitivity to follow norms in general. We find that rule sensitivity measured by this task is correlated (albeit weakly) with tax avoidance. While the rule-following task is primed by the inclusion of a rule in the instructions, the use of loophole is introduced as an innocuous opportunity. Therefore, it is possible that the weak correlation is due to this different priming of each choice. This also explains the weak correlation among the two measures as due to the lack of consensus violations are expected to be more common[3]. Though it is not the objective of our paper, we see a much higher and negative correlation between rule-following sensitivity and evasion.[4] Consistent with existing literature (Blaufus et al., 2016; Malik et al., 2018), these results show that social and moral considerations related to evasion and avoidance need not be the same.

We also ran additional treatments to check for the influence of social expectations in making a particular social norm salient and for their consequent impact on tax avoidance. We use Bicheri's (2015) typology to define a social norm as a context-dependent rule such that people follow it once they have two distinct expectations. First, they believe that most others in their reference network conform to it. This is termed the empirical expectation and overlaps nicely with the descriptive norm in Cialdini's (1990) typology. Second, they believe that most others in their reference network believe they ought to conform to this rule. This is called normative expectation and resembles injunctive norm in Cialdini's typology. We manage subjects empirical and normative expectations using data from our earlier treatments. In treatments where we manage subjects' empirical expectations, we find that if expectations are selfish, tax avoidance increases, whereas if they are fair, tax avoidance decreases. Compared with the control group, only subjects in the selfish choice treatment avoid more taxes. In the second pair of treatments, we manage subjects' normative expectations. We find that tax avoidance is significantly different (higher) in the selfish belief treatment than in the control group. The results highlight that subjects respond strongly to information in line with their narrow self-interest than the one in line with the norm perceived by them. In the final pair of treatments, we provide subjects conflicting empirical and normative expectations to see which of the two better predicts behaviour. The results show that average behaviour in the conflicting treatments is comparable to that in the treatment where we only provide empirical expectation than in treatment where we provide normative expectation. Thus in situations where the two expectations conflict, subjects are much more likely to

[1] The data for the United Kingdom has been taken from the HMRC's Measuring Tax Gaps publications covering 2011-12 through 2015-16.

[2] Throughout the paper, we will use the terms 'loophole', 'avoidance opportunity', and 'tax avoidance' interchangeably, but they convey the same meaning.

[3] It will be worthwhile to see if the introduction of sanctions can lead to such a consensus as sanctions often create focal points.

[4] For details, see Section A.1 in Appendix

hedge on the empirical expectations to resolve that conflict. Subjects beliefs about what others actually do (empirical expectation) shape tax avoidance more than their belief about what others think they should do (normative expectation).

1.2 Literature Review

1.2.1 Tax Compliance

Early theoretical approaches to model tax compliance behaviour borrowed heavily from the economics of crime approach pioneered by Gary Becker (1968). In these models (Allingham Sandmo, 1972; Yitzhaki, 1974), taxpayers act as if they face a gamble such that their optimal choice is an outcome of the expected utility maximization problem. Sources of risk in such models are the probability of detection and the ensuing punishment. However, such an approach is not feasible to study avoidance behaviour as avoidance does not involve an illegal act of hiding one's income. Though the primary motive of the taxpayer (reducing taxes) and fiscal consequences of their actions (increase in tax gap) are similar, the process differs considerably owing to legal as well as moral considerations of the taxpayer (Etzioni, 2010).
Several extensions of the basic model incorporate non-pecuniary determinants of tax compliance. In these models, monetary and non-monetary (psycho-social) preferences enter additively as below.

$$U(x_i) = E[u(x_i)] + V(x_i, \dots)$$

While the first term corresponds to the expected utility, as in the basic model, the second term incorporates the taxpayer's preference over several non-monetary consequences of non-compliance. Non-pecuniary preferences are included as reputational or psychic costs (Gordon, 1989). They also enter as the social custom utility of honest tax payment and conformity payoff of following the norm (Myles and Naylor, 1996), as stigmatization costs (Kim, 2003), and as an internalized social norm for tax compliance (Traxler, 2010). All the contributions above focus on evasion as the non-compliance behaviour, and none talks about avoidance or the associated morale costs.
Empirical data on taxpayers' compliance is limited and lacks reliability in many ways. Taxpayer Compliance Measurement Program (TCMP) of Internal Revenue Service (IRS), based on running intensive audits, fails to distinguish between honest mistakes and intentional non-compliance and does not identify underpaid taxes for sources of income not subject to information reporting (Andreoni et al., 1998). In the absence of controls for even various observable and theoretically established determinants of compliance, multiple researchers have taken the course of running field experiments (Slemrod et al., 2001; Fellner et al., 2013; Castro and Scartascini, 2015).
Questionnaire or survey-based studies of tax compliance enable a richer analysis of the subject. They include data on socioeconomic and attitudinal variables - variables that are absent in actual tax data. McGee and others (2008) survey a sample of US and Hong Kong university students and find from their responses that tax evasion is not always ethical. Some survey studies even show that tax ethics play a vital role in taxpayers' compliance decisions and that the strength of such a role depends on the strength of their commitment to tax ethics (Wenzel, 2005). Wenzel finds that individuals with stronger personal norms of tax honesty categorize themselves as more compliant while we find no such effect vis-à-vis the social norms. However, once we control for the variable that measures how strongly taxpayers identify themselves with the social group, social norms have a statistically significant negative effect on non-compliance for more strongly identified individuals. Nonetheless, survey studies suffer from various methodological issues: taxpayers may not remember their decisions, may give socially acceptable responses, may refuse to respond to such questions, and even may not truthfully reveal their decisions (Elffers et al., 1987).

1.2.2 Social Norms

One of the non-pecuniary determinants of tax compliance is the [perceived] social norm. Cialdini and Trost (1998) defined social norms as collectively understood rules of behaviour and conduct. Elster (1989) distinguishes social norms from moral norms as the former is not a product of instrumental reasoning. Their existence and subsequent sustenance rest substantially on a belief that people share them in the reference network. They are either unconditional or, if conditional, then not future-oriented. Bicchieri (2006), on the other hand, defines social norm as a context-dependent behavioural rule such that people prefer to follow it if they have two certain expectations. The empirical expectation is a first-order belief that most others in their reference network conform to it. The normative expectation is a second-order belief that most people in their reference network believe they ought to follow it. A social norm is hence a conditional preference.

Distinct from moral norms, it is independent of such expectations and thus requires an unconditional commitment for conformity based on emotional reactions such as disgust, contempt, etc.
Binmore (2005) posits that life is an indefinitely repeated game and argues that a meaningful analysis of social norms requires repeated interaction among subjects. Opposing this approach, Bicchieri (2005; 2006) models social norms as one-shot mixed-motive interactions where norms pull in one direction and self-interest pulls in the other. According to this view, if a norm exists and is believed to exist by individuals in a given network, it transforms each individual's utility function. The original mixed-motive game becomes a coordination game where sufficient people prefer to conform to the norm. Our study follows this latter approach.
Cialdini et al. (1990) studied the role of descriptive and injunctive norms on the littering behaviour of subjects. Descriptive norm specifies what people [actually] do in a particular setting and motivates action by telling people what is generally observed in that situation. On the other hand, injunctive norms specify what kind of behaviour people approve of and thus motivate action through perceived (or actual) sanctions. This distinction overlaps considerably with Bicchieri's typology. Descriptive norms resemble empirical expectations, and injunctive norms resonate closely with normative expectations. Cialdini (1991) also stresses that a norm needs to be activated or made salient to motivate and direct action. Thus, even if a norm is in place in two distinct situations, it will better predict behaviour in the setting where it is (more) focal. Normative content is not enough on its own. It is, thus, important which social group is focused on in such comparisons. People conform to observed norm violation and norm compliance if social information is related to individuals with known social proximity. If social proximity is unknown, subjects only conform to norm violations (Bicchieri et al., 2020). There is also evidence that we can improve norm adherence by advertising a high level of tax compliance among morale leaders or members of the moral reference group (Hammar et al., 2009; Gachter Renner, 2014). Given the increase in behavioural interventions involving such social information, it is essential to design these nudges very carefully. They are likely to fail or backfire if factors such as the context, reference dependence, and information credibility have not been incorporated reasonably at the design stage (Bicchieri and Dimant, 2019).
We can see the relative impact of empirical and normative expectations (or descriptive and injunctive norms) on behaviour by managing the subjects' expectations toward fairness or selfishness. Bicchieri and Xiao (2009) study this in the context of a dictator game. They work dictator's expectations either in the direction of selfishness or fairness. In situations where the two conflict, empirical expectations predict a dictator's choice better than normative expectations. In a similar vein, the work of Croson and Shang (2006 and2008) and Frey and Meier (2004) has contributed a lot to the importance of social comparison on giving behaviour. Social comparison can lead to either an upward effect or a downward effect. In the context of our experiment, social information acts as a shock to the initial beliefs of the subjects. An upward (downward) surprise arises when subjects' initial belief of tax avoidance is lower (higher) than what the majority in the reference group chose or approved.

1.3 Experimental Design

Our research design aims to understand the relationship between social norms and avoidance. We now describe the four stages of the experiment (section 3.1) and then our treatments (Section 3.2).[5]

1.3.1 Experimental Stages

The experiment consists of four stages (as outlined in Figure 1.1). Next, we describe each of them.

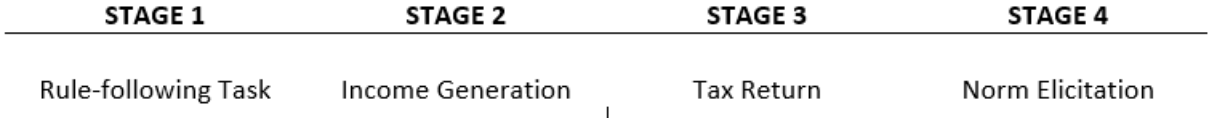


Figure 1.1: Stages of the Experiment

[5]This paper is a combination of results from two studies. In the first study, we construct two treatments based on the source of inequality among the subjects. The second study comprises six treatments based on the nature of the information received by the participants.

Stage 1: Rule Following

We use the norm-following task due to Kimbrough and Vostroknutov (2018) to measure individual propensity to follow the rules. The task requires subjects to allocate 100 balls into two buckets, yellow or blue. They drop balls one by one into any of the two buckets (see Appendix A.5.1). For each ball they put in the blue bucket, participants get 3 pence whereas, for each ball they put in the yellow bucket, participants get 6 pence. However, the instructions explicitly state that "the rule is to put the ball in the blue bucket". Subjects' earnings from this task have a lower bound at £3 (if they follow the rule entirely) and an upper bound at £6 (if they violate the rule completely). We use the loss incurred to follow the rule to measure rule-following proclivity (Kimbrough and Vostroknutov, 2018). The task involves an abstract context of elicitation. It is a variant of their previous 'traffic light' task (Kimbrough and Vostroknutov, 2016), which suffered from cultural variations in traffic conventions and a lack of variation in the norm-proclivity estimate.

Stage 2: Income Generation

In the next stage, subjects can earn income by completing an encryption task (see Appendix A.5.2). The task is due to Malik et al. (2018), a modified version of Erkal and others (2011). Each subject has 5 minutes to complete as many encryptions and earn as much as possible. The computer screen presents them an encryption table assigning a unique number to each of the 26 letters of the English alphabet. On the same screen, they use the table to encrypt words of English alphabets. The task requires them to find the corresponding number from the table for each letter of the word and type it in the box beneath that letter. Once they have correctly encrypted a word, the screen presents the next word. At the end of the 5 minutes, they move to the payoff screen, displaying their earnings. Each correct encryption earns subjects 50 Lab Sterlings equivalent to £0.25. This task provides us with an estimate of true taxable income for each subject.

Stage 3: Tax Return

In this stage, subjects declare their income for tax collection purposes. They observe that the income from the bucket task is tax-free and will be paid out to them in full at the end of the experiment. Instructions also inform them of the exogenous parameters like the tax rate (50 percent), the audit probability (10 percent), and the fine (100 percent of evaded amount). To increase the salience of the incentives, they see an example tax declaration along with post-tax income with and without audits taking place. They further find that the tax system involves redistribution. Once everyone submits their tax return, all collected taxes will be lumped together and redistributed equally among them. Finally, the instructions inform them that they have an opportunity to exempt some of their encryptions from taxes. This opportunity has arisen due to a loophole in the tax system which they can work in their favour. The use of the loophole ranges from 0 encryptions to 5 encryptions, where 0 means no tax avoidance and 5 means maximal tax avoidance. After they have understood the instructions, they can file their tax return (see Appendix A.5.3). They make two different choices on the tax return: declare their income and use the loophole. The first input gives us an estimate of tax evasion (actual income minus reported income), whereas the second input gives us an estimate of tax avoidance. To reflect real-life tax filing incentives, we allowed subjects to make a decision on evasion even though its not the objective of our study.

Stage4: Norm Elicitation Task

In this stage, we use a framework introduced by Krupka and Weber (2013) to directly elicit norms for the two treatments (see Appendix A.5.4). In this task, we present subjects with three scenarios in which an anonymous individual 'X' has to choose among several available actions. Subjects provide their social appropriateness ratings corresponding to each possible action on a four-level Likert scale: very socially appropriate, somewhat socially appropriate, somewhat socially inappropriate, and very socially inappropriate. We incentivize their responses using a coordination game. At the end of the task, a random draw selects one of the three scenarios and one of the possible actions. Subjects earned £2 if their response was the same as the one chosen by most other subjects in their session. Otherwise, they made £0.

Questionnaire

After the main experiment, participants answer demographic questions.The experiment involved eight treatments (two for the first study and 6 for the second), 20 sessions, and 329 subjects. All sessions took place at ESSExLAB, the University of Essex's social science research laboratory. We used the online recruitment system – Hroot, to invite participants (Bock et al., 2014). Sixty-five percent of our sample is composed of

	Treatment	Label	Study	Participants
1.	Unequal Opportunity	UEO	1	59
2.	Equal Opportunity	EO	1	50
3.	Fair Choice	FC	2	34
4.	Selfish Choice	SC	2	35
5.	Fair Belief	FB	2	47
6.	Selfish Belief	SB	2	42
7.	Fair Belief and Selfish Choice	FB + SC	2	31
8.	Selfish Belief and Fair Choice	SB + FC	2	31

Table 1.1: Summary information for all treatments.

female subjects, and the mean age of participants is 22.6 (minimum is 18 and maximum is 67). We used z-tree (Fischbacher, 2007) to run the experiment. As our experiment involved human subjects, we obtained ethical approval from the ethics committee at the University of Essex.

1.3.2 Treatments

Equal Opportunity vs. Unequal Opportunity

Table 1.1 provides summary information for all eight treatments (two for the first study and 6 for the second). We base the first two treatments on the source of income inequality between the groups. The equal-opportunity (EO) treatment is identical to the unequal-opportunity (UEO) treatment except that in the latter, some participants earn systematically less (half as much) than the other group of participants. We achieve this by manipulating the length of words participants must encrypt in stage 2. Any differences in income in the equal-opportunity treatment are solely due to individual effort (the task itself is not demanding in skills and abilities). Therefore, redistribution seems more justified in unequal-opportunity treatment than equal-opportunity treatment. Moreover, moral costs of non-compliance are more salient in unequal-opportunity treatment than in equal-opportunity treatment.

Conjecture 1a: If subjects take the fairness/justification of redistribution into account in their tax decision, then we expect more compliance in the unequal-opportunity than in the equal-opportunity treatment.

The channel of this behavioural difference is the difference in the [perceived] norm. We expect subjects to rate avoidance more inappropriately in the unequal-opportunity treatment. As they expect a stronger norm against tax avoidance in the unequal opportunity treatment, norm dependent utility maximizers will choose closer to the no-avoidance option.

Conjecture 1b: Alternatively, if subjects do not take the fairness/justification of redistribution into account in their tax decision, then we do not expect significantly different compliance in the unequal opportunity than in the equal-opportunity treatment.

Expectation Management

To look at the role of social expectations on tax avoidance, we use data from the sessions on the first two treatments (equal vs unequal opportunity) to manage subjects' expectations. We provide each subject with information on what the majority chose in a previous session (empirical expectation) and (or) what the majority believed one should choose (normative expectation). This information is provided after they have learnt about the tax system and before they file their tax return. We also elicit their initial belief about what subjects actually did in that session and (or) their belief about what subjects felt others needed to do.[6] We then provide them with the actual value. Table 1.2 describes the content of the information for each treatment. We call no use of the tax avoidance opportunity fair and maximal use of that opportunity as selfish. By providing subjects with a fair choice (belief), we want to create a fair empirical (normative) expectation in the subjects. In the selfish (fair) choice treatment, irrespective of their initial belief, each subject is informed that most people in the previous session chose to make a full (no) use of the loophole. Likewise, in the selfish (fair) belief treatment, each subject is informed that most people in the previous session believed one should make full (no) use of the loophole. The control or benchmark group does not receive any message.

[6]Specifically, we ask them what level of the loophole was chosen by the majority (from 0 to 5) and (or) which of those levels the majority believed one ought to choose.

Treatment	Message
Fair Choice	Most people in a previous session chose to exempt 0 Encryptions, i.e. majority of the subjects did not make any use of the tax loophole.
Selfish Choice	Most people in a previous session chose to exempt 5 Encryptions, i.e. majority of the subjects made full use of the loophole.
Fair Belief	Most people in a previous session said using 0 Encryptions is the most socially approved action. In other words, majority of the subjects said that one should make no use of the loophole.
Selfish Belief	Most people in a previous session said using 5 Encryptions is the most socially approved action. In other words, majority of the subjects said that one should make full use of the tax loophole.
Fair Belief & Selfish Choice	Most people in session ALPHA said using 0 Encryptions is the most socially approved action. In other words, majority of the subjects said that one should make no use of the loophole. Most people in session BETA chose to exempt 5 Encryptions, i.e. majority of the subjects made full use of the tax loophole.
Selfish Belief & Fair Choice	Most people in the session ALPHA said using 5 Encryptions is the most socially approved action. In other words, majority of the subjects said that one should make full use of the tax loophole. Most people in session BETA chose to exempt 0 Encryptions, i.e. majority of the subjects did not make any use of the tax loophole.

Table 1.2: Messages containing information on social expectations for treatments in the second study.

Conjecture 2: We expect that there will be less avoidance in the fair choice (belief) treatment as compared with that in the selfish choice (belief) treatment.

Conjecture 3: Furthermore, the impact of negative social information (selfish choice or selfish belief) will be stronger than that of positive social information (fair choice or fair belief) compared to the benchmark treatment.

Under negative social information, norm and self-interest pull subjects in the same direction. It is in the narrow self-interest of each individual to undertake maximal tax avoidance irrespective of which treatment they fall under. The norms, however, are distinct under each treatment. In selfish treatment, the implied norm is to undertake full tax avoidance, the same as self-interest prescribes. Under fair treatment, the indicated norm is not to undertake tax avoidance, whereas self-interest is to avoid taxes maximally. While the norm pulls the subject in one direction, self-interest pulls them in the opposite.

1.3.3 Norm-driven Preferences and the Utility Function

This section presents our model with a norm-dependent utility specification in the spirit of Krupka and Weber (2013) and Thomson and Vostroknutov (2017). The utility function has two components: financial payoff and non-financial payoff. Suppose that the subject i makes use of the avoidance opportunity, let's say, $x \in \{0, 1, 2, 3, 4, 5\}$. The norm-dependent utility of this subject, therefore, is,

$$U_i(x) = \beta_i(x) + \gamma_i g^r(x)$$

$\beta_i(x)$ is the personal or consumption utility of the subject; $\gamma_i > 0$ is the norm-proclivity estimate of subject i; $r \in \{1, 2, 3\}$ represents the treatments (scenarios); and $g^r : \{0, 1, \ldots, 5\} \to [-1, 1]$ is the shared social appropriateness of action x. Therefore, the second term captures the costs (benefits) associated with violating (following) the elicited norm. The rule-following task (aka the bucket task) estimates γ_i, whereas the norm elicitation task estimates g^r under the three scenarios. The social norm is captured by the function $g^r(x)$ and expressed as the subject's beliefs about the most appropriate action given the context. Moreover, the norm is also assumed to be common knowledge among the target population. Individuals who do not care about norms ($\gamma_i = 0$) will be more likely to select the payoff maximizing action. An individual who cares about norms ($\gamma_i > 0$) will choose in the direction of the social norm. With the change in context, the social norm can change significantly. Any differences in tax avoidance across the treatment will be due to the differences in perceived social norms. In contrast, variation within the same treatment (context) will be due to differences in the norm-proclivities of subjects. Thus, in a given context, rule-followers (high γ_i) will choose closer to the norm, whereas rule breakers (low γ_i) will choose away from the norm and closer to the payoff maximizing action.

Conjecture 4: Norm-dependent utility implies that agents who are more sensitive to following the rule in the bucket task will avoid less in the tax task than those who are less inclined to follow the rule.

In the bucket task, subjects observe a non-binding and unenforced injunctive norm to put the balls in the blue bucket. However, putting the balls in the yellow bucket pays twice as much. The task creates incentives such that self-interest pulls subjects in one direction while a preference for norms pulls subjects in the opposite direction. Keeping in mind the incentive scheme, we consider subjects who put more balls in the blue bucket as those who care about norms in general. If there is a norm of compliance in the tax task, subjects who show higher sensitivity towards rule-following in the bucket task will be more compliant with their taxes.

Conjecture 5: The role of norm proclivity will be mediated by the strength of the norm, calculated as the level of agreement on what the norm prescribes or proscribes. When the agreement is weak, violations will be more common despite the emergence of clear norm content.

	EQUAL OPPORTUNITY	**UNEQUAL OPPORTUNITY**	
DEMOGRAPHICS		*Earn-High*	*Earn-Low*
Mean Age	22.14	24.06	21.63
Share female	0.66	0.59	0.80
Share student	1.00	0.93	1.00
DECISIONS		*Earn-High*	*Earn-Low*
Rule-following	0.52	0.62	0.63
Sheets Avoided	2.76	2.83	1.80
Income Evaded	452.70	493.10	215.83
% Income Evaded	0.21	0.23	0.22

Table 1.3: Descriptive Statistics for Study 1.

1.4 Results

In this section, we first look at the results from the rule-following task. Next, we look at the results from the norm-elicitation stage and compare the elicited norm across the treatments. Finally, we examine the association between subjects' rule-following sensitivity and tax compliance.

1.4.1 Rule Following behaviour

The histogram in Figure 1.2 shows the distribution of rule-following or the number of balls put in the blue bucket. In contrast to self-interest, many subjects (more than 20 percent) put all the balls in the blue bucket, whereas fewer subjects (almost 5 percent) put all the balls in the yellow bucket. This task has been employed to measure norm-sensitivity across many contexts and cultures (Gurdal et al., 2020; Thomsson and Vostroknutov, 2017; Kimbrough and Vostroknutov, 2018).[7]

1.4.2 Elicited Norm

Figure 1.3 uses data from the norm elicitation task to describe the norm under each scenario. We describe a (perceived) social norm as the action with the highest appropriateness rating. We can see that not avoiding taxes is the most appropriate action under both treatments. The appropriateness ratings decrease monotonically for each additional use of the loophole such that maximal tax avoidance is considered the least appropriate action. The norm perceived by the subjects is the same across the two treatments. The emergence of a similar norm rejects conjecture 1a and supports conjecture 1b.
Bicchieri (2005) discusses two aspects of a social norm: qualitative and quantitative. The qualitative element deals with the content of the norm and describes which actions in the set of available actions are considered appropriate or inappropriate. However, the content of the norm alone is not sufficient to make an individual comply with the norm. Compliance also depends on norm strength which depends on the agreement about the norm among the relevant network. Figure 1.3 shows that subjects in our experiment agree on the content of the norm. The norm that emerges is against the use of tax avoidance. However, is there a consensus among them on that?
In table 1.4 we present the social appropriateness ratings by treatment. The top row corresponds to each possible action individual 'A' in the hypothetical situation could take. The next four rows provide full distribution of responses, and the next row provides the mean of the appropriateness ratings. The mean can

[7]The mass on full rule-following is not very different in our study from other studies. Though, the probability mass on full-violation and midway violation (similar to the evidence on partial lying) is slightly different.

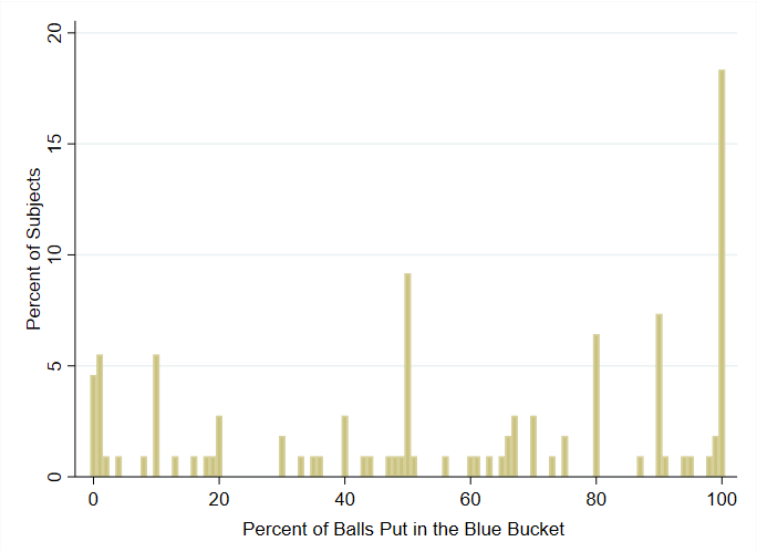


Figure 1.2: Results from the Rule-following Task

range from complete agreement on socially appropriate (1) to complete agreement on socially inappropriate (-1). The distribution of ratings shows a lack of agreement on perceived appropriateness or inappropriateness of actions. The average rating for no tax avoidance is 0.604 under the UEO treatment and 0.56 under the EO treatment, far from the consensus rating of +1. Similarly, the average rating for maximum tax avoidance

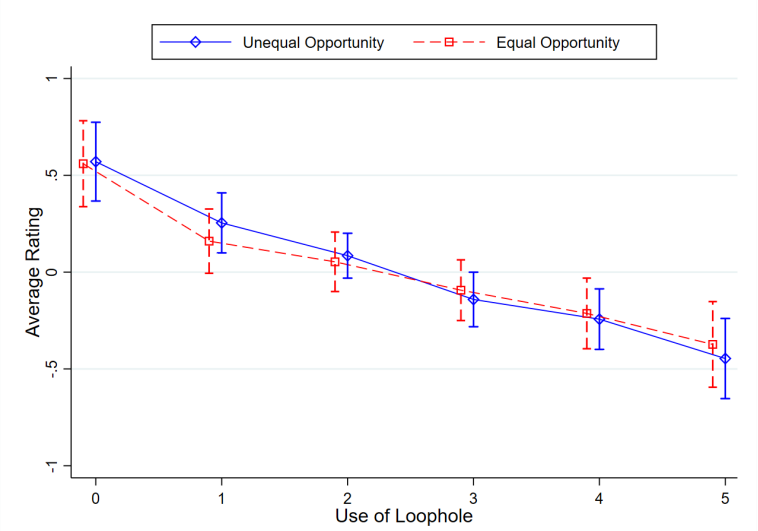


Figure 1.3: Norm perceived by subjects in the Equal Opportunity and the Unequal Opportunity treatments. Figure shows the appropriateness ratings for each use of tax loophole (0,1,2,3,4,5).

is -0.412 under the UEO treatment and -0.373 under EO treatment which is even further from the consensus rating of -1. Though subjects may perceive a norm against avoidance, lack of solid agreement can likely lead to common occurrences of norm violation in each treatment (more so in the EO treatment). In line with the predictions from conjectures 4 and 5 about norm strength and norm sensitivity, disagreement will dilute the norm effect so that we will see a clear norm effect only among strongly rule-following and strongly rule-breaking individuals.

1.4.3 Tax Avoidance

Changes in the context of a decision environment can lead to changes in the social norm. So, differences in behaviour across contexts can be attributed to perceived norms. However, differences in behaviour within the same context can be explained by differences in individual willingness to conform to the norm.
In Figure 1.4, we plot tax avoidance across the two treatments. First, tax avoidance is lower under the

	Norm Strength by Treatment					
Unequal Opportunity	**0**	**1**	**2**	**3**	**4**	**5**
Very Socially Inappropriate	13.56	8.47	3.39	13.56	28.81	54.24
Somewhat Socially Inappropriate	6.78	16.95	28.81	44.07	42.37	22.03
Somewhat Socially Appropriate	5.08	38.98	50.85	32.2	20.34	5.08
Very Socially Appropriate	74.58	35.59	16.95	10.17	8.47	18.64
Mean ratings	**0.604**	**0.345**	**0.209**	**-0.07**	**-0.276**	**-0.412**

Equal Opportunity	**0**	**1**	**2**	**3**	**4**	**5**
Very Socially Inappropriate	15.25	8.47	8.47	11.86	23.73	44.07
Somewhat Socially Inappropriate	1.69	20.34	27.12	37.29	30.51	16.95
Somewhat Socially Appropriate	6.78	40.68	40.68	28.81	22.03	8.47
Very Socially Appropriate	61.02	15.25	8.47	6.78	8.47	15.25
Mean ratings	**0.56**	**0.16**	**0.053**	**-0.093**	**-0.213**	**-0.373**

Table 1.4: Level of agreement among subjects about the norm under the equal opportunity to earn and unequal opportunity to earn treatments. For each level of 'loophole use' (from 0 sheets avoided to 5 sheets avoided), we provide the percentage of subjects who gave it a certain rating. We also provide the average of ratings provided by subjects (ranging from -1 to +1)

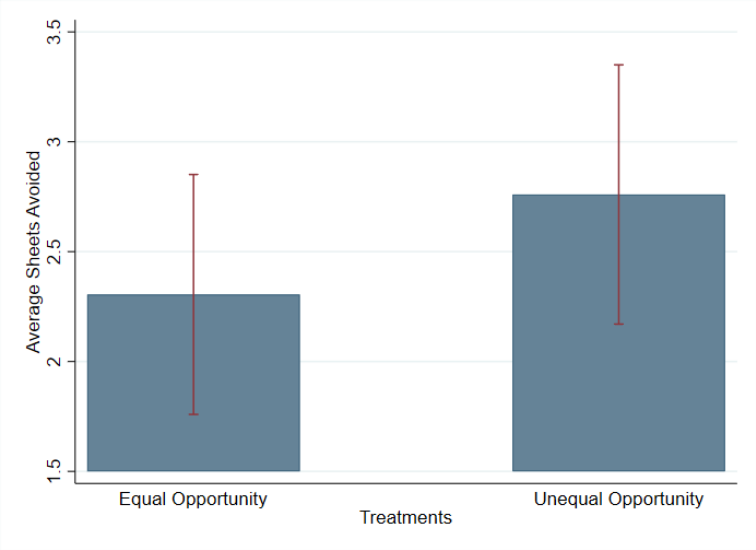


Figure 1.4: Bars indicate the average number of sheets avoided from tax payment for treatments with equal and unequal opportunity to earn.

treatment where redistribution seemed more justified due to the presence of an unequal opportunity to earn. Subjects, on average, avoided 2.76 sheets in the EO treatment and 2.3 sheets in the UEO treatment. However, according to the Wilcoxon rank-sum test, the difference is not statistically significant at any level of significance ($z = -1.083; p - value = 0.278$).

Spearman's rank correlation coefficient between the number of sheets avoided and the fraction of balls put in the blue bucket is negative ($\rho = -0.173, p = 0.07, N = 109$). Thus that tax avoidance is modestly correlated with rule following.[8] Results from our study show that rule-followers, on average, avoid less than rule breakers. Figure 1.5 below shows that the percentage of individuals who decide not to avoid taxes increases as we move up the quartiles of rule-following. The data fit well with conjecture 4 that rule-following in the bucket task indicates the extent of avoidance in the tax game. However, these differences are statistically significant only among strongly rule-following and rule-breaking individuals. It is not surprising if we read these results together with the results on norm consensus. Due to a weak consensus about the true norm, only subjects with strong sensitivity will differ in behaviour. To highlight this difference in tax avoidance, we classify rule followers and rule breakers as subjects in the 4th and the 1st quartile of rule sensitivity,

[8]We find that norm sensitivity and evasion are strongly negatively correlated ($\rho = -0.2922$; $p - value = 0.0021$ (see Appendix A.1).) which can be interpreted as an indication of strong morale costs.

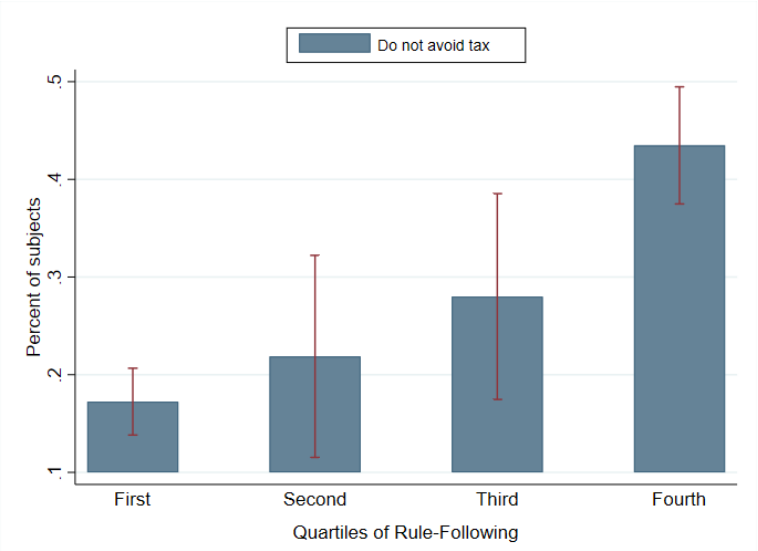


Figure 1.5: Percentage of subjects who did not avoid taxes by quartiles of rule-following. Rule following increases with the quartiles such that the 1st quartile represents the least rule-following and the 4th quartile represents the most rule-following subjects.

respectively. In Figure 1.6, we see that tax avoidance is lower among strong rule followers (2.03 sheets) than among strong rule-breakers (3.13 sheets). A Wilcoxon rank-sum (Mann-Whitney) test shows that this difference is significant at 10 percent significance level and only marginally insignificant at 5 percent level (p-value=0.0527).

We then conduct OLS regressions to analyze the relationship between rule sensitivity and tax avoidance. We report these regressions in Table 1.5. The dependent variable in the reported regressions is the number of sheets subjects avoid tax on. In the first column, we regress only on rule sensitivity and find it negatively related to avoidance. This effect is significant at a 1 percent level of significance. In column 2, we add the treatment dummy and find that while rule sensitivity is still significant, the coefficient on the treatment dummy is insignificant. Again, the results show that tax avoidance is not different across the two treatments. In the third column, we add the interaction of rule sensitivity and the treatment dummy and control for demographics like age and gender. The coefficient on rule-sensitivity is now insignificant. Although the interaction is also insignificant, we test whether rule-sensitivity + interaction = 0. The test rejects the null hypothesis that rule-sensitivity + interaction = 0. So, there is an average effect of rule-sensitivity on tax avoidance, but it is mostly driven by unequal opportunity treatment. This treatment makes the norm against avoidance more salient, and thus individuals who are sensitive to following rules, in general, respond by reducing tax. We then drop the interaction term in further regressions.

In the fourth column, we control for risk preference, personal norms about tax avoidance, and fairness of the tax system.[9] The coefficient on the gender dummy is not significant, whereas the coefficient on age is significant and negatively related to tax avoidance. Even after controlling for these variables, the coefficient on rule sensitivity is still significant at the 5 percent level of significance. Finally, we control for session fixed effects and find rule sensitivity to continue to be significantly negatively related to avoidance. The R^2 improves from 2.6 percent (column 1) to 17 percent (column 5).

1.5 Expectation Management

In these treatments, we manage subjects' empirical and normative expectations using data from other experimental sessions. We provide subjects with a message informing them about the majority of the subjects' actual choice (empirical expectation) and (or) their most approved choice (normative expectation) in another session of a similar experiment. Before doing so, we ask them to guess which of the six possible actions was the majority's choice. We further tell them that they will earn £2 if their guess turned out to be correct and £0 otherwise. Thus, we incentivize them to give their true expectation. This design choice enabled us to obtain subjects' initial beliefs about the actual and approved level of tax avoidance. Once they

[9]To estimate the risk preferences, we ask subjects to rate their willingness to take risk in life, in general (Dohmen et al. (2011). We also measure personal norms about avoidance by asking them to rate if they think it can always (10) or never be justified (1) to avoid taxes.

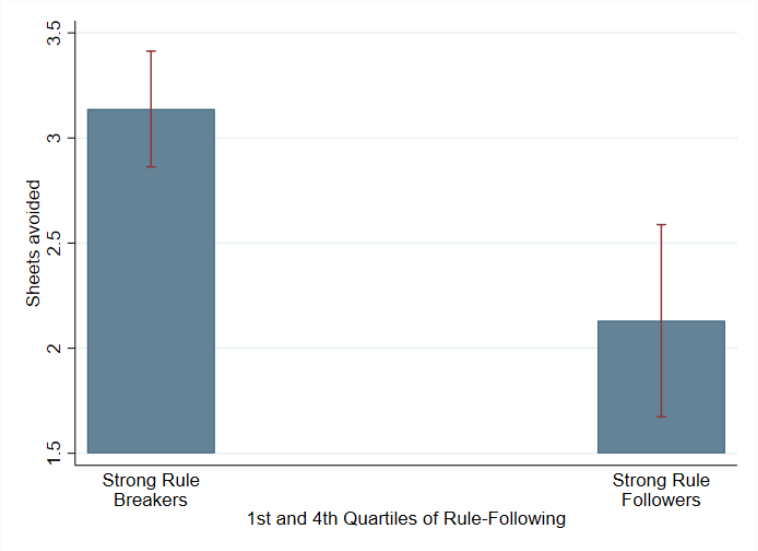


Figure 1.6: Strong rule-breakers (followers) are subjects among the 1st (4th) quartile of rule-following. The bars indicate the average sheets avoided by the two groups of subjects from tax payments.

	Use of Loophole				
	(1)	(2)	(3)	(4)	(5)
rule-sensitivity	-0.973**	-0.901**	-0.136	-0.740**	-0.943**
	(0.243)	(0.277)	(0.576)	(0.254)	(0.335)
Unequal Opportunity		-0.369	0.388	-0.304	0.0547
		(0.479)	(0.466)	(0.539)	(0.144)
rs × *UnequalOpportunity*			-1.210		
			(0.837)		
Constant	3.075***	3.233***	4.205**	4.323**	4.273***
	(0.207)	(0.219)	(1.233)	(1.313)	(0.787)
Observations	109	109	109	109	109
R-squared	0.027	0.034	0.105	0.096	0.172
Demographics			YES	YES	YES
Risk			YES	YES	YES
TaxFair			YES	YES	YES
Justified			YES	YES	YES
Session FE					YES

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 1.5: OLS regression: Dependent variable is the Use of Loophole. Rule-sensitivity $\in [0, 1]$. *Demographics* include age and gender. *Unequal Opportunity* is the treatment dummy (=1 under unequal opportunity to earn condition). *Risk* is the risk preference measured using the method by Dohmen et al. (2011). *TaxFair* measures whether subjects think the tax system applied in the experiment was justified or not. *Justified* measures whether subjects think it can always or never be justified to avoid taxes. *Session FE* controls for session fixed effects.

confirm their choice, subjects proceed to the next screen. Here, we provide them with the true value of these expectations. Table 1.2 shows the six treatments on managing subjects' expectations and the corresponding messages sent to the subjects.

Some norms are widely followed, and an individual's normative and empirical expectations often push behaviour in the same direction. Under such norms, violations are highly unlikely. However, there are also largely unfollowed norms, indicating an underlying contradiction between empirical and normative expectations. There is strong evidence of this inherent inconsistency in the example of rampant corruption in many developing countries. While formal laws and informal social norms are condemning corruption, the widespread prevalence of bribe-taking, nepotism, and kickbacks push people to expect that the majority of their compatriots engage in acts of corruption. There is, however, a lack of scholarly investigation into the relative importance of these expectations, especially in cases when they conflict. In a 2015 study on public attitude towards tax avoidance, HMRC detected this conflict in normative and empirical expectations. The majority of the respondents (63 percent) felt that tax avoidance schemes were widespread (high empirical expectation). In contrast, most of the respondents (61 percent) also believed that it is never acceptable

to use tax avoidance schemes (low normative expectation).[10] Wenzel (2005) used questionnaires and administrative data to show a pattern of pluralistic ignorance among Australian taxpayers.[11] One previous study that looked at the relative importance of empirical and normative expectations is by Bicchieri and Xiao (2009). They use data from Dictator games to manage subjects' empirical and normative expectations towards fairness or selfishness. They find empirical expectations to be the primary determinant of norm obedience. We aim to apply their methodology in a tax avoidance context to check three things. First, do empirical expectations continue to be the dominant force affecting behaviour? Second, what are the implications of the conflicting expectations regarding tax avoidance behaviour found in the HMRC study? Should the messages from HMRC continue to focus on the lack of social approval of tax avoidance among the public, or should they also address the misperception of widespread avoidance?

	Fair Choice	**Selfish Choice**	**Fair Belief**	**Selfish Belief**	**Fair Belief & Selfish Choice**	**Selfish Belief & Fair Choice**
DEMOGRAPHICS						
Mean Age	22.74	21.37	23.30	21.60	24.84	22.23
Share female	0.71	0.80	0.64	0.55	0.71	0.48
Share student	0.94	0.97	0.98	0.93	0.84	0.94
DECISIONS						
Rule-following	0.59	0.69	0.51	0.58	0.56	0.59
Sheets Avoided	2.56	3.94	2.23	4.24	3.10	2.90
Income Evaded	369.12	311.57	561.17	416.67	550.00	533.87
% Income Evaded	0.18	0.16	0.28	0.20	0.26	0.27

Table 1.6: Descriptive Statistics for Study 2.

1.5.1 Expectations About Other participants

We also measured subjects' expectations about other participants. To elicit empirical expectations of subjects, we asked them how many participants in their session they believed made particular use of the loophole? They had to report a percentage for all six possible actions (see Appendix A.4.7). At the end of the experiment, we randomly draw one of the six possible actions and pay subjects £2 if their answer is the same as the actual number of subjects who made that choice during that session.
We also ask subjects, 'do you think one ought to use an x amount of the loophole?'. To elicit their normative expectations, we ask how many participants they think said 'yes' to the previous question in their session. We randomly draw one of the six questions on normative expectations and pay them £2 if their response matched the actual number (see Appendix A.4.8).

Empirical Expectations

In the belief elicitation stage, each subject answers the question, "what percentage of people in today's session do you think made a certain use of the loophole?". They have to provide an estimate for each of the six possible actions satisfying $EE(x) \geq 0$ and $\sum_{x=0}^{5} EE(x) = 1$. These estimates provide us a subjects' empirical expectation ($EE_j^i(x)$) for each of the six possible actions. We calculate the average of the empirical expectations for each action under each treatment. Figure 1.7 (a) provides subjects' beliefs regarding actual tax avoidance for the treatment involving the management of empirical information. In comparison with their initial beliefs, providing a fair (selfish) choice increases (decreases) their expectation about no tax avoidance. The opposite is true for initial beliefs and empirical expectations about 'maximal tax avoidance' across the two treatments. We can see that the expectation of tax avoidance increases (decreases), though not monotonically, with the level of tax avoidance in the case of selfish (fair) choice treatments. Knowledge of fair (unfair) use by others leads subjects to expect that most of their compatriots make a fair (unfair) use of the loophole.
Figure 1.7 (b) shows the same beliefs of subjects for the treatments involving management of normative expectations. Belief management was successful in shifting empirical expectations across the two treatments. In the fair belief treatment, far more people expect their compatriots not to avoid taxes than those in the selfish belief treatment did (36.5 percent against 7.1 percent; t-test; $p - value = 0.0001$). In the selfish belief

[10] Observations of this study are noteworthy as the use of tax avoidance schemes is not widespread in the UK. A 2017 study by HMRC shows that tax avoidance accounted for only 8 percent of the total tax gap (0.5 percent of total theoretical tax liabilities) in 2013-14.

[11] which refers to the situation where most group members privately reject (accept) a group norm but believe most others accept (reject) it.

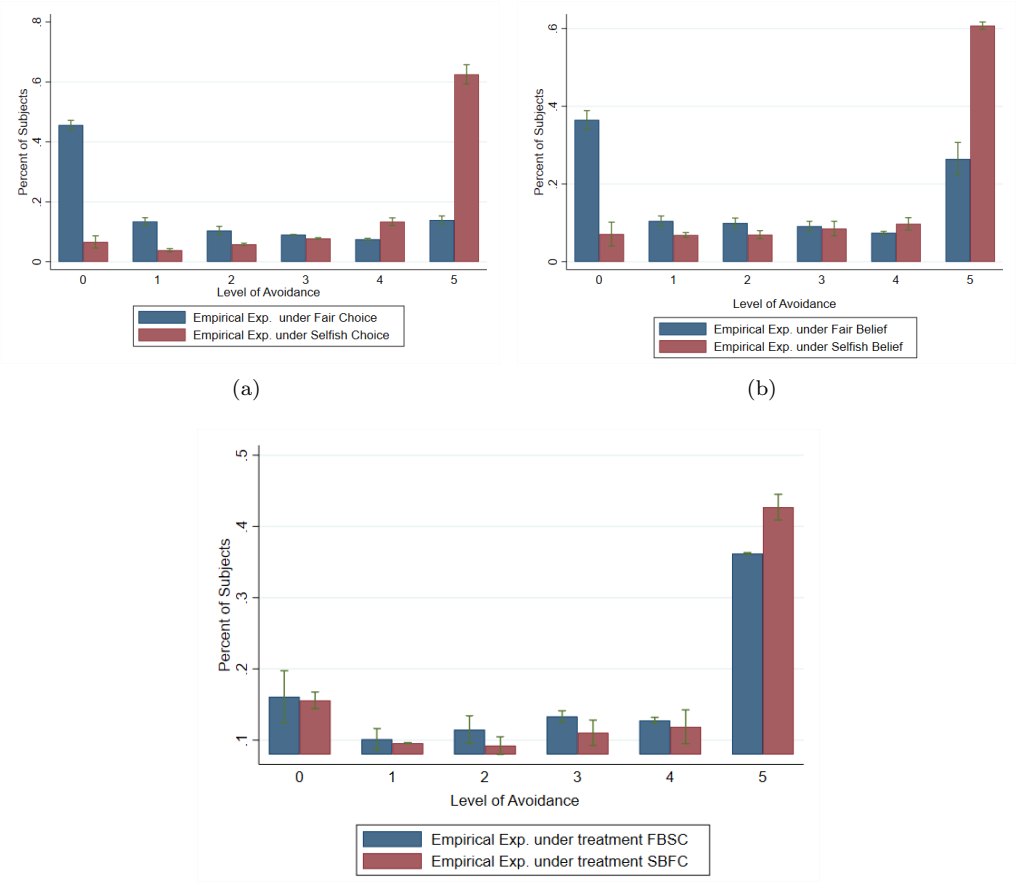


Figure 1.7: Subjects' belief of level of tax avoidance during their session for treatments involving **(a)** empirical information, **(b)** normative information, and **(c)** conflicting information. Bars indicate the average of the subject's expectations about the percentage of people who make the respective use of the tax loophole.

treatment, far more people expect their fellow participants to undertake maximal tax avoidance than those in the fair belief treatment did (60.7 percent against 26.4 percent; t-test; $p - value < 0.0001$).Again, we see that the normative information leads to the desired expectation among subjects about other participants. Subjects expect a higher fraction of members not to avoid taxes when they observe that most subjects in another session did not avoid taxes. Similarly, they expect an overwhelming majority of members to avoid taxes maximally when they find that most subjects in the other session avoided taxes maximally. The data from treatments involving conflicting information shows that the subjects' do not hold significantly different beliefs about the use of tax avoidance across the two treatments (see Figure 1.7 (c)). Only when they estimate the percentage of people who undertake maximal tax avoidance is where we see the difference. Subjects who participated in the treatment involving selfish belief (SB+FC) expect higher use of maximal tax avoidance than those who participated in the treatment involving fair belief (FB+SC).

Normative Expectations

We also elicit normative expectations in an incentivized manner (see Appendix A.4.8). For each subject, we get an estimate of the normative expectation ($EE^i_j(x)$) regarding each use of the tax loophole. We calculate the average of the normative expectations (NE) in each treatment from these estimates.

$$NE^i(x) = \frac{\sum_{j=1}^{n_t} NE^i_j(x)}{n_i}$$

where $x \in \{0, 1, 2, .., 5\}$ represents the use of loophole, $i \in \{SC, FC\}$ represents the treatment, $j \in \{1, 2, ..., n\}$ represents the individual subjects in treatment i, and n_i represents the total number of subjects in treatment i.

Figure 1.8 (a) shows the distribution of normative expectations for the fair and selfish choice treatments.

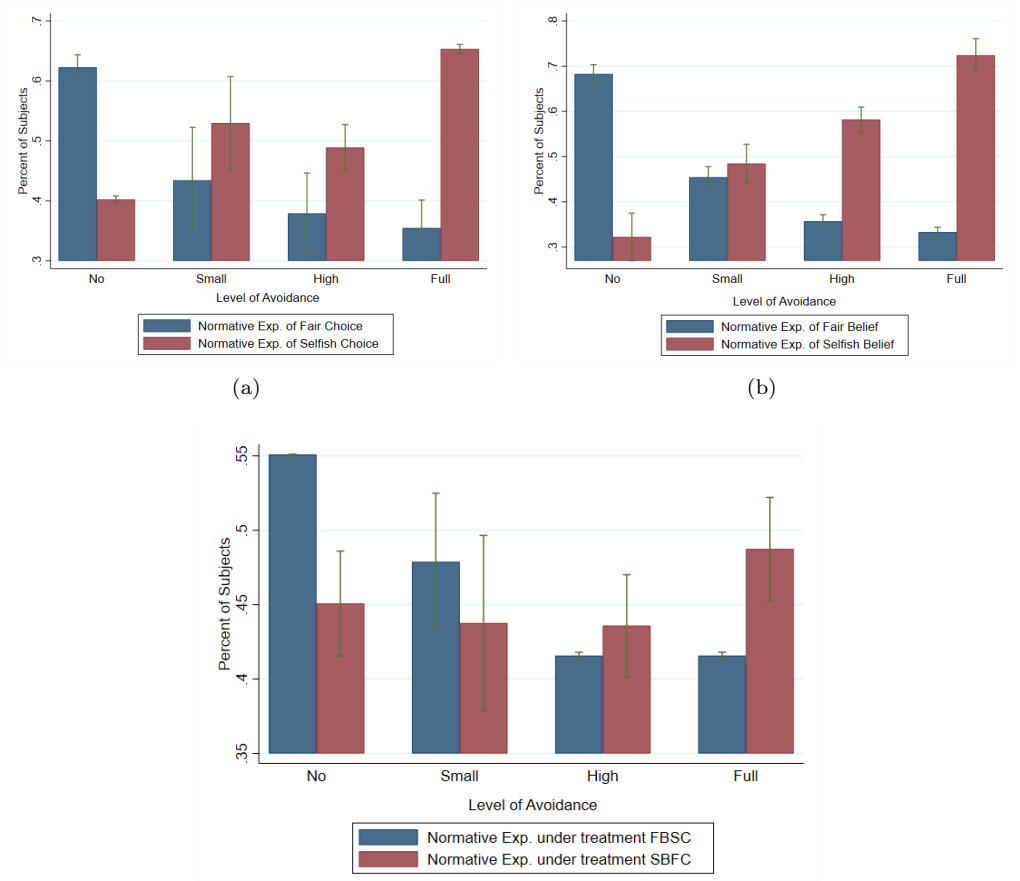


Figure 1.8: Subjects' belief of level of tax avoidance during their session for treatments involving **(a)** empirical information, **(b)** normative information, and **(c)** conflicting information. Bars indicate the average of the subject's expectations about the percentage of people who approve of the respective use of the tax loophole.

Subjects who participated in the selfish choice treatment expect more people to consider maximal tax avoidance as the appropriate action. In contrast, subjects who participated in the fair choice treatment expect more people to consider no tax avoidance as the appropriate action. If we look at the distribution of EEs and NEs together, two things emerge. First, social information has an intuitive impact on NEs. Upward (selfish) social information makes subjects think others regard higher levels of tax avoidance more appropriate. They believe more people expect others to undertake maximal tax avoidance than any other level of avoidance. The opposite holds for downward (fair) social information where subjects believe that more people expect others not to avoid taxes than undertaking any positive level of tax avoidance. Nonetheless, they still believe 35 percent of subjects expect others to undertake maximal avoidance. Providing relevant (downward) social information does not create a strong norm against tax avoidance.[12]

In treatments involving normative informational content, subjects who participated in the selfish belief treatment expect more people to consider higher levels of tax avoidance as more appropriate (see Figure 1.8 (b)). Subjects who participated in the fair belief treatment expect more people to consider less tax avoidance as more appropriate. As in the case of empirical expectations, we can see that the expectation of acceptance increases (decreases) monotonically with the level of tax avoidance under selfish (fair) choice treatments. These results show that subjects' beliefs are sensitive to the provided information. Their beliefs move towards the norm that is salient under the available information.

In Figure 1.8 (c), we look at the data from our conflicting information treatments. We find that subjects who participated in the treatment involving a fair belief (FB+SC) expect more of their co-participants to consider no or small tax avoidance to be more appropriate than those who participated in the treatment

[12]In future, it will be worthwhile to see if information backed by sanctions can create an even stronger norm against tax avoidance.

involving a selfish belief (SB+FC). Also, fewer fellow-subjects consider high and full tax avoidance to be more appropriate.
If we look at the distribution of empirical and normative expectations together, two things emerge. First, social information has an intuitive impact on normative expectations. Upward (selfish) social information makes subjects think that others consider a higher level of tax avoidance more appropriate. They believe more people expect others to undertake maximal tax avoidance than less or no tax avoidance. Though, they still believe a considerable percentage of subjects expect them not to avoid taxes. The opposite holds for downward (selfish) social information where subjects believe that more people expect others to not undertake tax avoidance at all than undertaking any positive level of tax avoidance. Nonetheless, they still believe a considerable fraction of subjects expect them to undertake maximal tax avoidance. Subjects beliefs are susceptible to the salience of information. Irrespective of their treatment, some subjects believe others do not approve of tax avoidance. It will be worthwhile to see if information backed by sanctions can create a strong norm against avoidance in future work. [13]

1.5.2 Tax Avoidance

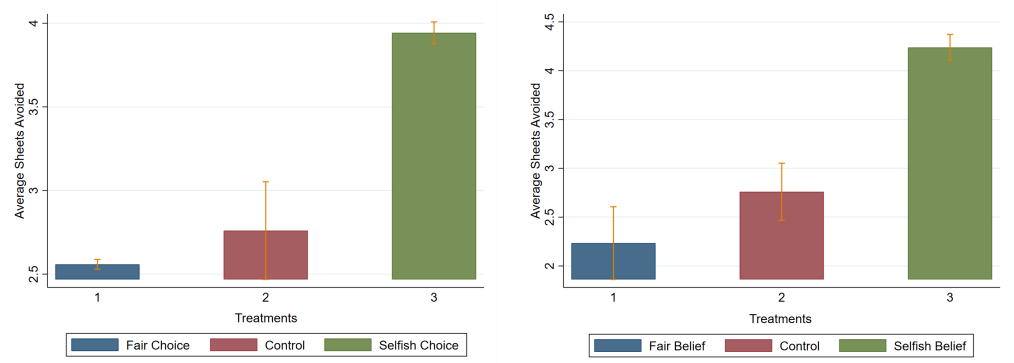
We first look at the average level of tax avoidance under the fair and selfish choice treatments. As shown in Figure 1.9 (a), mean tax avoidance is 3.94 in the selfish choice treatment and 2.55 in the fair choice treatment. In line with conjecture 2, this difference is statistically significant (p-value < 0.01). While tax avoidance in fair choice treatment is not statistically different from that in the control group, the difference is statistically different under selfish choice treatment (t-test gives p-values of 0.6767 and 0.0091 respectively). This result supports conjecture 3, such that the impact of upward (selfish) information is stronger than the impact of downward (fair) information.

We now look at tax avoidance across selfish and fair belief treatments. As shown in Figure 1.9 (b), subjects on average avoided taxes on 2.23 sheets in the fair belief treatment and on 4.23 sheets in the selfish belief treatment. Tax avoidance in the fair belief treatment is not statistically different from that in the control group (t-test; p-value = 0.2351). In the selfish belief treatment, tax avoidance is significantly different from the control group (t-test; p-value = 0.0002). These results support conjecture 3 and indicate that subjects respond more strongly to information on empirical and normative expectations that best serve their narrow self-interest than information that nudges them towards a socially desirable behaviour. When the information has either normative or empirical content, subjects react more strongly to normative information.
Looking at aggregate tax avoidance under conflicting information (Figure 1.9 (c)), we find that subjects are pulled more towards the unfair expectation. Average sheets avoided are 3.1 in the FB+SC treatment and 2.9 in the SB+FC treatment. While subjects avoid more taxes in the treatment involving selfish choice management than in the treatment involving fair choice management, this difference is not statistically significant at any level of significance. Our finding is that when subjects are provided pro-social information (FC or FB), it leads to an insignificant reduction in tax avoidance. However, when they are given pro-self information (SC or SB), it results in a significant increase in tax avoidance. This indicates that subjects respond more to information (expectations) that are in line with their narrow self-interest than those that are in line with social interest. When they receive two types of signals that provide conflicting information then their behaviour is again in the direction of their self-interest, though the differences are insignificant.

1.6 Tax Avoidance and Managed Expectations

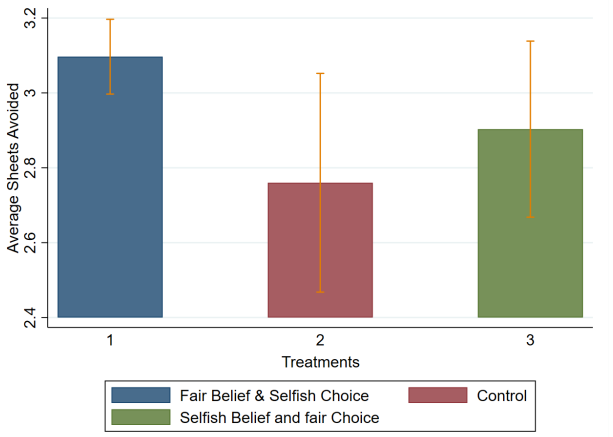
We observed that in isolation, normative expectations create more variation in tax avoidance than empirical expectations. It is, however, worthwhile to see their relative effect in situations where subjects have both normative and empirical information at hand and the two conflict. In Figure 1.10, we look at average tax avoidance across the six information managements. First, we run a Kruskal-Wallis test to check if average tax avoidance is different across the six treatments. The test reveals a statistically significant difference in tax avoidance across the six variations ($\chi^2 = 26.803$ and $p < 0.001$).
We now confine our analysis only to those subjects who acted in accordance with the social norm (most socially approved action). To look more closely at the relative importance of empirical and normative expectations, we compare norm compliance in the conflicting information treatments with their respective single information variants. Mean sheets avoided in the SC+FB treatment are significantly more as compared with 'FB alone' treatment (3.09 vs 2.23; t-test; $p = 0.0851$) and significantly less as compared with 'SC alone'

[13] One way to do so is to inform subjects that in addition to disapproving tax avoidance, participants in a previous session also put monetary sanctions on those who avoided taxes.



(a)

(b)



(c)

Figure 1.9: Subjects' level of tax avoidance during their session for treatments involving **(a)** empirical information, **(b)** normative information, and **(c)** conflicting information. Bars indicate, for each treatment, the average number of sheets avoided from tax payment.

treatment (3.09 vs 3.94; t-test; $p = 0.0894$) albeit only at 10 percent level of significance. Although the conflicting treatment has a significantly different average use of the tax avoidance opportunity than its single information versions, the difference is less pronounced under the 'SC alone' treatment. Mean avoidance in the FC+SB treatment is insignificantly more in comparison with the 'FC alone' treatment(2.9 vs. 2.55; t-test; $p = 0.5393$) and significantly less than the 'SB alone' treatment (2.9 vs. 4.24; t-test; $p = 0.006$). Together, these results imply that average behaviour in the conflicting treatments is more comparable to average behaviour in treatments with only the empirical information than in treatments with just the normative information. In other words, subjects in the conflicting treatments hedge their choice more on empirical content of the information than on its normative content.

We also plot fair (no) use of avoidance opportunity against its empirical and normative expectations under all six treatments. In Figure 1.11 (a), we get a comparison of the usage of fair use and the associated empirical expectations. We use a t-test to compare the percentage of fair choices and the mean of empirical expectations. Out of all six comparisons, three show no significant difference at a 1 percent level of significance. In Figure 1.11 (b), we get a comparison of fair choice and the associated normative expectations. Again, we use a t-test to compare the percentage of fair choices and the mean of normative expectations. All six comparisons show a significant difference between the means at a 1 percent significance level. These results show that subjects' use of avoidance opportunity is more in line with their belief about what others actually do (empirical expectation) than their belief about what others think others should do (normative expectation). When we say that people respond more to empirical expectation than normative expectation it is only for norm compliant subjects (who do not avoid any taxes).

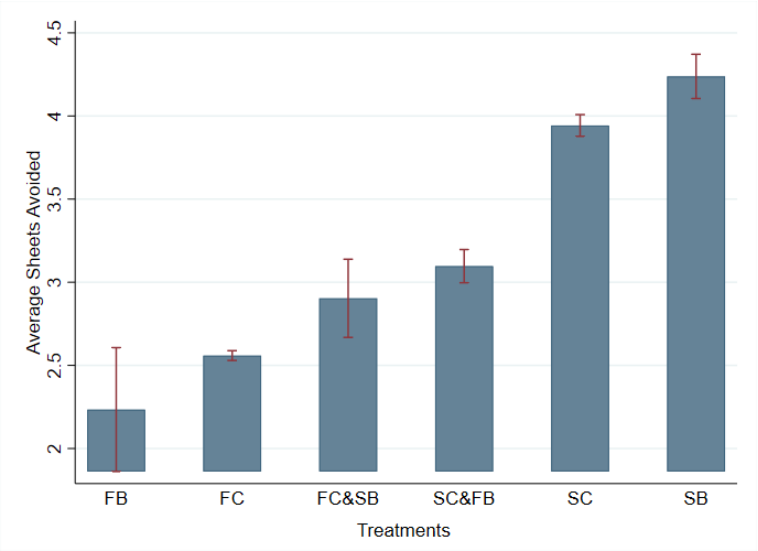
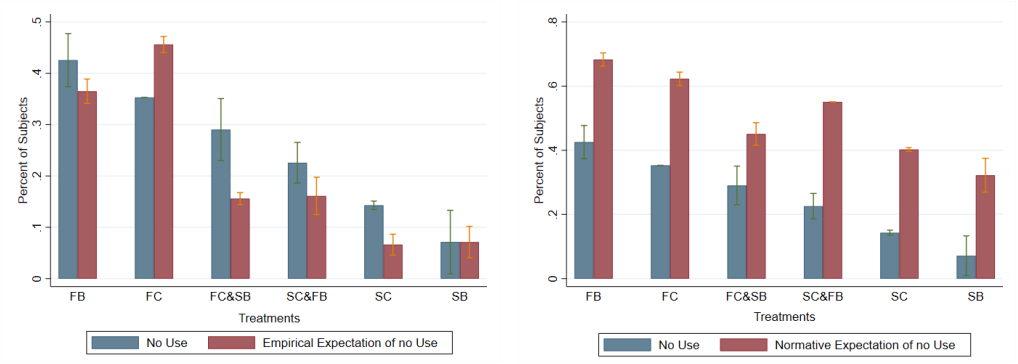


Figure 1.10: Subjects' level of tax avoidance by treatment. Bars indicate, for each treatment, the average number of sheets avoided from tax payment.



(a) (b)

Figure 1.11: Percentage of people who did not avoid taxes plotted against **(a)** the reported expectation about percentage of people who had not avoided taxes in their session, and **(a)** the reported expectation about percentage of people who approved of no tax avoidance in their session. Navy bars represent the actual behaviour and maroon bars represent expectations about **(a)** empirical expectation, and **(a)** normative expectation.

1.6.1 Regression Analysis

Empirical Expectations

The preliminary results show that social information leads to a difference in behaviour even when the norm under consideration is weak. In Table 1.7, we present a series of OLS regressions where we control for various factors. These regressions compare the treatments involving empirical expectations with the control group (no information). The dependent variable is the number of sheets subjects avoid tax on. In the first specification, we use *Selfish Choice* and *Fair Choice* dummies as the independent variables where the control group is the baseline. Selfish Choice dummy takes a value of 1 for the treatment where subjects are informed about a selfish Choice and 0 otherwise. *Fair Choice* dummy takes a value of 1 if the treatment is where the subjects are informed about a fair choice and 0 otherwise. Baseline is the control group where subjects do not receive any information. We find the coefficient on the *Selfish Choice* dummy to be moderate, positive, and significant. Thus encountering a selfish choice increases tax avoidance significantly. On average, subjects avoid 1.183 more sheets than the control group. The coefficient on the *Fair Choice* dummy is small, negative and insignificant. Qualitatively, providing a fair choice decreases tax avoidance, but this reduction is not significantly different from the control group. In the second specification, we add demographics such as age

and gender. We also add other controls such as subjects' risk preferences, whether subjects think the tax system applied in the experiment was fair, and whether subjects think it can be justified to avoid taxes. The coefficient on the *Selfish Choice* dummy is still significant while that on the *Fair Choice* dummy continues to be insignificant. Among demographic variables, only the coefficient on *age* is significant such that tax avoidance decreases with the age of subjects. These results are robust to adding additional controls like session fixed effects. R-squared improves from 7.4 percent in the first regression to 16.3 percent in the third regression.

	Use of Loophole		
	(1)	(2)	(3)
Selfish Choice	1.183***	1.202***	0.589**
	(0.265)	(0.322)	(0.188)
Fair Choice	-0.201	-0.0638	-0.739***
	(0.261)	(0.311)	(0.120)
Constant	2.760***	3.455**	4.393***
	(0.260)	(0.974)	(1.115)
Observations	119	119	119
R-squared	0.074	0.150	0.163
Demographics		YES	YES
RISK		YES	YES
JUSTIFIED		YES	YES
TAXFAIR		YES	YES
SESSION FE			YES

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 1.7: Dependent variable is the Use of Loophole. Demographics include age and gender. Risk is the risk preference measured using the method by Dohmen et al. (2011). TaxFair measures whether subjects think the tax system applied in the experiment was justified or not. Justified measures whether subjects think it can always or never be justified to avoid taxes. Session FE controls for session fixed effects.

Normative Expectations

	Use of Loophole		
	(1)	(2)	(3)
Selfish Belief	1.478***	1.688***	2.402**
	(0.280)	(0.230)	(1.038)
Fair Belief	-0.526	-0.477	0.383
	(0.413)	(0.283)	(0.866)
Constant	2.760***	0.506	0.924
	(0.255)	(1.375)	(1.198)
Observations	139	139	139
R-squared	0.148	0.194	0.211
Demographics		YES	YES
RISK		YES	YES
JUSTIFIED		YES	YES
TAXFAIR		YES	YES
SESSION FE			YES

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 1.8: Dependent variable is the Use of Loophole. Demographics include age and gender. Risk is the risk preference measured using the method by Dohmen et al. (2011). TaxFair measures whether subjects think the tax system applied in the experiment was justified or not. Justified measures whether subjects think it can always or never be justified to avoid taxes. Session FE controls for session fixed effects.

In Table 1.8, we present a series of OLS regressions comparing treatments providing information on normative expectations with the control group (no information). Again, the dependent variable is the number of sheets subjects avoid tax on. In the first specification, we use *Selfish Belief* and *Fair Belief* dummies as the independent variable where the control group is the baseline. *Selfish Belief* dummy takes a value of 1 if the treatment is where subjects receive a selfish belief and 0 otherwise. Similarly, *Fair Belief* dummy takes a value of 1 if the treatment is where the subjects receive a fair belief and 0 otherwise. Baseline is the treatment where subjects are not provided with any information. We find the coefficient on the *Selfish Belief* dummy to be positive, moderate and significant. Therefore, encountering a selfish belief increases tax avoidance significantly. On average, subjects avoid 1.478 more sheets than the control group. On the other hand, the coefficient on the *Fair Belief* dummy is negative and insignificant. Qualitatively, providing a fair belief decreases tax avoidance, and this reduction is not significant. The second specification involves the

addition of demographics like age and gender. We also control subjects' risk preferences, whether subjects think the tax system applied in the experiment was fair, and whether subjects think it can be justified to avoid taxes. The coefficient on the *Selfish Belief* dummy is still significant while that on the *Fair Belief* dummy continues to be insignificant. Among demographic variables, only the coefficient on *age* is significant such that tax avoidance decreases with the age of subjects. Finally, we add session fixed effects in the third specification and still find the *Selfish Belief* dummy to be significant. R-squared improves from 14.8 percent in the first specification to 21.1 percent in the third.

1.7 Discussion and Conclusion

We find that tax avoidance is affected by the management of empirical and normative expectations. Influence on avoidance is more decisive under information that focuses on unfair content than on fair content. When information draws on only one of the expectations, normative expectation comes out as a stronger influence than the empirical expectation. However, taxpayers hedge on the empirical expectation to resolve that conflict when information conflicts.

There is a lot of experimental evidence on the role of payoff relevant information affecting behaviour (Fehr and Schmidt, 2001). Still, there is compelling evidence that in economic interactions, payoff-independent information about the behavior of other subjects can influence behaviour in general and norm obedience (Bichieri and Xiao, 2009; Krupka and Weber, 2008). Our paper adds to the latter strand of literature. It shows that norm obedience is due to the mediating role of perceived empirical and normative expectation of behaviour in our population of interest. We also add to the literature on norm-dependent tax compliance (Traxler, 2010). We find that tax behaviour is conditional on observed social norms within and between groups. Our study, thus, also provides a role for belief management as an essential policy instrument, especially when such beliefs are at odds with reality.

Providing subjects with different information about the prevalent use of tax avoidance opportunities triggers a varied response. If they observe that most people do not avoid taxes, they avoid fewer taxes. On the contrary, if they find that most people fully avoid taxes, they undertake a higher level of tax avoidance - on average. Initial beliefs of subjects are such that they expect a lot of people to hold a norm against avoidance and at the same time expect many of them to not act upon such a norm. One reason could be the absence of formal/informal sanctions against norm violation in our experiment. Subjects respond more strongly to unfair (upward) social information than fair (downward) social information. One possible reason for this effect is that both the norm and self-interest pull them in the same direction in upward social information. On the other hand, under downward social information norm pulls them in one direction, and self-interest pulls in the other direction (Croson and Shang, 2008).[14]

In addition to empirical expectations, we find that differences in normative expectations also lead to significant differences in tax behaviour. In our experiment, normative information created a greater difference in the average level of tax avoidance than empirical information. If subjects' beliefs are managed in the upward (unfair) direction, they undertake a higher level of tax avoidance (on average). In contrast, if beliefs are managed in the downward (fair) direction, they avoid fewer taxes (on average). The difference in tax avoidance between treatments is larger under normative management than empirical management. So, subjects react more strongly if the information has only an injunctive or normative content than if it only has descriptive or empirical content.

Our conflicting information treatments provide a precise analysis of which of the two expectations weigh more in decision-making. We find that subjects put their weight on empirical expectations to decide in situations where the two expectations conflict. Their use of avoidance opportunity is more in line with their expectations about what others in their session do than their expectations about what others think others should do. Thus they avoid taxes if they believe sufficient others also avoid taxes, even when they believe that the fellow members will not approve such behavior. Our results speak to the policymakers, especially in domains where they aim to use measures to contain anti-social behaviour. We find that containing undesirable behaviour takes more than just leveraging on the illegitimacy of such behaviour. It will help much more the cause of public policy if interventions highlight that most people act according to the norm. Tax collecting institutions like the HMRC already incur many administrative costs. They can decrease tax violations cost-efficiently by disseminating information about a significantly large fraction of norm-abiding taxpayers.

[14] Bicchieri (2005) argues that norm compliance is conditional upon holding a belief that most others (a) perform that behaviour, (b) expect others to perform that behaviour and, (c) are willing to sanction those who do not perform that behaviour.

Bibliography

Agarnov, M., Healy, P.J. and Nielsen, K. *Randomization Across Domains: Testing Theories of Probability Matching and Convex Preferences.* Working Paper

Allingham, M.G. and Sandmo, A., 1972. *Income tax evasion: A theoretical analysis.* Journal of public economics, 1(3-4), pp.323-338.

Alm, J., 1988. *Compliance costs and the tax avoidance-tax evasion decision.* Public Finance Quarterly, 16(1), pp.31-66.

Alm, J., McClelland, G.H. and Schulze, W.D., 1992. *Why do people pay taxes?.* Journal of public Economics, 48(1), pp.21-38.

Alm, J., Jackson, B.R. and McKee, M., 1993. *Fiscal exchange, collective decision institutions, and tax compliance.* Journal of Economic Behavior Organization, 22(3), pp.285-303.

Alm, J. and Torgler, B., 2006. *Culture differences and tax morale in the United States and in Europe.* Journal of economic psychology, 27(2), pp.224-246.

Alm, J., McClelland, G.H. and Schulze, W.D., 1999. *Changing the social norm of tax compliance by voting.* Kyklos, 52(2), pp.141-171.

Alm, J., Bahl, R. and Murray, M.N., 1993. *Audit selection and income tax underreporting in the tax compliance game.* Journal of development Economics, 42(1), pp.1-33.

Alm, J., 1999. *Tax compliance and administration.* Public Administration and Public Policy, 72, pp.741-768.

Andreoni, J., Erard, B. and Feinstein, J., 1998. *Tax compliance.* Journal of economic literature, 36(2), pp.818-860.

Arcand, J.L. and Graziosi, G.R., 2005. *Tax compliance and rank dependent expected utility.* The Geneva Risk and Insurance Review, 30(1), pp.57-69.

Becker, G.S., 1968. *Crime and punishment: An economic approach.* In The economic dimensions of crime (pp. 13-68). Palgrave Macmillan, London.

Becker, W., Büchner, H.J. and Sleeking, S., 1987. *The impact of public transfer expenditures on tax evasion: An experimental approach.* Journal of Public Economics, 34(2), pp.243-252.

Bicchieri, C., 2005. *The grammar of society: The nature and dynamics of social norms.* Cambridge University Press.

Bicchieri, C. and Xiao, E., 2009. *Do the right thing: but only if others do so.* Journal of Behavioral Decision Making, 22(2), pp.191-208.

Bicchieri, C. and Dimant, E., 2019. *Nudging with care: The risks and benefits of social information.* Public choice, pp.1-22.

Bicchieri, C., Dimant, E., Gächter, S. and Nosenzo, D., 2020. *Social Proximity and the Erosion of Norm Compliance.*

Binmore, K., 2005. *Natural justice.* Oxford University Press.

Blaufus, K., Hundsdoerfer, J., Jacob, M. and Sünwoldt, M., 2016. *Does legality matter? The case of tax avoidance and evasion.* Journal of Economic Behavior Organization, 127, pp.182-206.

Bock, O., Baetge, I. and Nicklisch, A., 2014. *hroot: Hamburg registration and organization online tool.* European Economic Review, 71, pp.117-120.

Bordignon, M., 1993. *A fairness approach to income tax evasion.* Journal of Public Economics, 52(3), pp.345-362.

Braithwaite, V., 2003. *Dancing with tax authorities: Motivational postures and non-compliant actions.* Taxing democracy, pp.15-39.

Braithwaite, V.A., 2009. *Defiance in taxation and governance: Resisting and dismissing authority in a democracy.* Edward Elgar Publishing.

Castro, L. and Scartascini, C., 2015. *Tax compliance and enforcement in the pampas evidence from a field experiment.* Journal of Economic Behavior Organization, 116, pp.65-82.

Cialdini, R.B. and Trost, M.R., 1998. *Social influence: Social norms, conformity and compliance.*

Cialdini, R.B., Reno, R.R. and Kallgren, C.A., 1990. *A focus theory of normative conduct: recycling the concept of norms to reduce littering in public places.* Journal of personality and social psychology, 58(6), pp.1015.

Cialdini, R.B., Kallgren, C.A. and Reno, R.R., 1991. *A focus theory of normative conduct: A theoretical refinement and reevaluation of the role of norms in human behavior.* In Advances in experimental social psychology (Vol. 24, pp. 201-234). Academic Press.

Cialdini, R.B. and Goldstein, N.J., 2004. *Social influence: Compliance and conformity.* Annu. Rev. Psychol., 55, pp.591-621.

Cowell, F.A., 1990. *Tax sheltering and the cost of evasion.* Oxford Economic Papers, 42(1), pp.231-243.

Croson, R. and Shang, J.Y., 2008. *The impact of downward social information on contribution decisions.* Experimental Economics, 11(3), pp.221-233.

Cummings, R.G., Martinez-Vazquez, J., McKee, M. and Torgler, B., 2009. *Tax morale affects tax compliance: Evidence from surveys and an artefactual field experiment.* Journal of Economic Behavior Organization, 70(3), pp.447-457.

Cooter, R., 1998. *Expressive law and economics.* The Journal of Legal Studies, 27(S2), pp.585-607.

Dhami, S. and Al-Nowaihi, A., 2007. *Why do people pay taxes? Prospect theory versus expected utility theory.* Journal of Economic Behavior Organization, 64(1), pp.171-192.

Dohmen, T., Falk, A., Huffman, D., Sunde, U., Schupp, J. and Wagner, G.G., 2011. *Individual risk attitudes: Measurement, determinants, and behavioral consequences.* Journal of the European Economic Association, 9(3), pp.522-550.

Dubin, J.A., Graetz, M.J. and Wilde, L.L., 1990. *The effect of audit rates on the federal individual income tax, 1977-1986.* National Tax Journal, pp.395-409.

Elffers, H., Weigel, R.H. and Hessing, D.J., 1987. *The consequences of different strategies for measuring tax evasion behavior.* Journal of Economic Psychology, 8(3), pp.311-337.

Elster, J., 1989. *Social norms and economic theory.* Journal of economic perspectives, 3(4), pp.99-117.

Elster, J., 2015. *Explaining social behavior: More nuts and bolts for the social sciences.* Cambridge University Press.

Erkal, N., Gangadharan, L. and Nikiforakis, N., 2011. *Relative earnings and giving in a real-effort experiment.* American Economic Review, 101(7), pp.3330-48.

Etzioni, A., 2010. *Moral dimension: Toward a new economics.* Simon and Schuster.

Fehr, E. and Schmidt, K.M., 2001. *Theories of fairness and reciprocity-evidence and economic applications.*

Fellner, G., Sausgruber, R. and Traxler, C., 2013. *Testing enforcement strategies in the field: Threat, moral appeal and social information.* Journal of the European Economic Association, 11(3), pp.634-660.

Fischbacher, U., 2007. *z-Tree: Zurich toolbox for ready-made economic experiments.* Experimental economics, 10(2), pp.171-178.

Fortin, B., Lacroix, G. and Villeval, M.C., 2007. *Tax evasion and social interactions.* Journal of Public Economics, 91(11-12), pp.2089-2112.

Frey, B.S. and Meier, S., 2004. *Social comparisons and pro-social behavior: Testing 'conditional cooperation' in a field experiment.* American Economic Review, 94(5), pp.1717-1722.

Frey, B.S., 1997. *A constitution for knaves crowds out civic virtues.* The Economic Journal, 107(443), pp.1043-1053.

Frey, B.S. and Torgler, B., 2007. *Tax morale and conditional cooperation.* Journal of comparative economics, 35(1), pp.136-159.

Gächter, S. and Renner, E., 2014. *Leaders as role models for the voluntary provision of public goods.*

Gordon, J.P., 1989. *Individual morality and reputation costs as deterrents to tax evasion.* European economic review, 33(4), pp.797-805.

Gürdal, M.Y., Torul, O. and Vostroknutov, A., 2019. *Norm compliance, enforcement, and the survival of redistributive institutions.*

Haidt, J. and Joseph, C., 2004. *Intuitive ethics: How innately prepared intuitions generate culturally variable virtues.* Daedalus, 133(4), pp.55-66.

Hammar, H., Jagers, S.C. and Nordblom, K., 2009. *Perceived tax evasion and the importance of trust.* The Journal of Socio-Economics, 38(2), pp.238-245.

HMRC, (2017). *Measuring Tax Gaps* available at: see https://www.gov.uk/government/uploads/system/uploads/attachment
measuring-tax-gaps-2017.pdf

Kim, Y., 2003. *Income distribution and equilibrium multiplicity in a stigma-based model of tax evasion.* Journal of Public Economics, 87(7-8), pp.1591-1616.

Kimbrough, E.O. and Vostroknutov, A., 2016. *Norms make preferences social.* Journal of the European Economic Association, 14(3), pp.608-638.

Kimbrough, E.O. and Vostroknutov, A., 2018. *A portable method of eliciting respect for social norms.* Economics Letters, 168, pp.147-150.

Kirchler, E., 2007. *The economic psychology of tax behaviour.* Cambridge University Press.

Krupka, E. and Weber, R., 2008. *The focusing and observational effects of social norms.*

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

Malik, S., Mihm, B. and Timme, F., 2018. *An experimental analysis of tax avoidance policies.* International Tax and Public Finance 25.1 : 200-239.

McGee, R.W., Ho, S.S. and Li, A.Y., 2008. *A comparative study on perceived ethics of tax evasion: Hong Kong vs the United States.* Journal of Business Ethics, 77(2), pp.147-158.

Murphy, K., 2003. *An examination of taxpayers attitudes towards the Australian tax system: Findings from a survey of tax scheme investors.* Australian Tax Forum, 18, 208–241.

Murphy, R., 2012. *The EU tax gap.* Technical report.

Myles, G. D. and Naylor, R. A., 1996. *A model of tax evasion with group conformity and social customs.* European Journal of Political Economy, 12(1), 49-66.

Neck, R., Wächter, J.U. and Schneider, F., 2012. *Tax avoidance versus tax evasion: on some determinants of the shadow economy* Int Tax Public Finance pp 104-117

Office of Tax Simplification, 2015. *Complexity Index Methodology Paper.*

Orviska, M. and Hudson, J., 2002. *Tax evasion, civic duty and the law abiding citizen.* European Journal of Political Economy, 19 (1), 83-102.

Pommerehne, W.W. and Frey, B.S., 1992. *The effects of tax administration on tax morale* (No. 191). Diskussionsbeiträge: Serie II, Sonderforschungsbereich 178 'Internationalisierung der Wirtschaft', Universität Konstanz.

Posner, R., 1981. *The Economics of Justice.* Harvard University Press, Cambridge, MA.

Ryan, R. M. and Deci, E. L., 2000. *Intrinsic and extrinsic motivations: Classic definitions and new directions.* Contemporary educational psychology, 25(1), pp 54-67.

Shang, J. and Croson, R., 2006. *The impact of social comparisons on nonprofit fund raising.* In Experiments investigating fundraising and charitable contributors (pp. 143-156). Emerald Group Publishing Limited.

Slemrod, J., 1985. *An Empirical Test for Tax Evasion.* The Review of Economics and Statistics, 67, issue 2, pp 232-38,

Slemrod, J., Blumenthal, M. and Christian, C., 2001. *Taxpayer response to an increased probability of audit: evidence from a controlled experiment in Minnesota.* Journal of public economics, 79(3), pp 455-483.

Slemrod, J. B., 2016. *Tax Compliance and Enforcement: New Research and Its Policy Implications.* Ross School of Business, Paper No. 1302. Available at SSRN: https://ssrn.com/abstract=2726077

Snow, A. and Warren, R.S., 2005 *Ambiguity about audit probability, tax compliance, and taxpayer welfare.* Economic Inquiry, 43(4), pp 865-871.

Stiglitz, J.E., 1985. *The general theory of tax avoidance.* Natl. Tax J. 38 (3), pp 325–337. Team, Behavioural Insights, 2012. *Applying behavioural insights to reduce fraud, error and debt.* Cabinet Office, London.

Thomsson, K. M. and Vostroknutov, A., 2017. *Small-world conservatives and rigid liberals: Attitudes towards sharing in self-proclaimed left and right.* Journal of Economic Behavior Organization pp 181-192.

Torgler, B. and Schneider, F., 2007. *What shapes attitudes toward paying taxes? Evidence from multicultural European countries.* Social Science Quarterly, 88(2), 443-470.

Traxler, C., 2010. *Social norms and conditional cooperative taxpayers.* European Journal of Political Economy 26.1 pp 89-103.

Turner, J. C., Hogg, M. A., Oakes, P. J., Reicher, S. D. and Wetherell, M. S., 1987. *Rediscovering the social group: A self-categorization theory.* Basil Blackwell.

Wenzel, M. (2004). *An analysis of norm processes in tax compliance.* Journal of economic psychology, 25(2), pp 213-228.

Wenzel, M., 2005. *Misperceptions of social norms about tax compliance: From theory to intervention.* Journal of Economic Psychology, 26(6), pp 862-883.

Yitzhaki, S., 1974. *Income tax evasion: A theoretical analysis.* Journal of Public Economics, 3, issue 2, pp 201-202.

Chapter 2

Incentives for whistle-blowing: An Experiment

Zia Ul Hassan Khan[1]

Abstract

We investigate the effectiveness of financial rewards in encouraging optimal whistle-blowing. To do so, we conduct a laboratory experiment that features a two-player bribery game between a public official and a citizen. First, the public official decides whether (and how much) to ask for a bribe. Then, the citizen jointly decides whether to pay the bribe and whether to report the public official for corruption. We use a between-subject experimental design that varies the accuracy of the judicial system (defined as the probability of type-1 and type-2 errors) and the evidence threshold required for proving guilt (defined as the cost of reporting). Our results indicate that the underlying judicial environment influences the impact of financial rewards on bribe incidence and whistle-blowing. The percentage of officials who ask for a bribe and the average bribe are both significantly lower under a high judicial accuracy. We also find a significant decrease in false reporting by citizens under a high evidence threshold. Taken together, our results suggest that optimal whistle-blowing occurs in environments with a high judicial accuracy and a high evidence threshold.

Keywords: Corruption, Harassment Bribes, Whistleblowing, Judicial Accuracy, Evidence Threshold.
JEL: C92, D73, D9, K4

[1]Zia-Ul-Hassan Khan, Department of Economics, University of Essex, United Kingdom. email: zuhkha@essex.ac.uk

2.1 Introduction

Corruption is a widespread phenomenon that affects all countries. Given its adverse impact on economic growth, investment, and social disparity (Lambsdorff, 2005), a lot can be achieved through in-depth investigations of the causes of corruption and the policies through which it can effectively be reduced (Bardhan, 1997). The answers to these questions vary depending on the context of corruption. Bribery is one major form of corruption. Citizens often offer bribes to get undue favors in contracts and services from public servants. Such bribes are therefore distortionary in nature. There are also instances where citizens are forced to pay bribes to obtain services they are legally entitled to. These types of bribes have been named '*harassment bribes*' in the literature and are the interest of our study.[2]

It is argued that higher financial rewards can motivate individuals to report illegal activities. Countries vary considerably in legislation against corruption and in the use of whistle-blowing incentives.[3]The impact of any reward scheme to incentivize whistle-blowing will be mediated by the accuracy of the judicial system and by the adopted level of evidence threshold required to indict someone. Judicial accuracy is an indicator of the quality of the judicial system and therefore determines the overall error rate of the judicial system. Two errors can arise. First, a false positive (type-1 error) occurs when an innocent individual is convicted under a false accusation of bribery. Second, a false negative (type-2 error) occurs when a guilty individual is exonerated after a truthful report. Evidence threshold or burden of proof is the minimum amount of evidence required by the court to convict the accused. It determines how many cases get passed on for prosecution. If the threshold level is high, individuals may not invest in gathering evidence. Hence, fewer cases (true or false) will reach the prosecution stage. The opposite holds if the evidence threshold is low. A high evidence threshold means the error is distributed such that false negatives are high and false positives are low. A low evidence threshold means the error is distributed such that false positives are high and false negatives are low.

Spagnolo et al. (2017) show that introducing a whistle-blower reward scheme in a country where courts do not perform well can be harmful. Citizens either falsely report or do not report at all, depending on how the system distributes the two errors. Both scenarios reduce social efficiency. In such a case, improving judicial accuracy is the only policy that unambiguously improves the gains from whistle-blower reward schemes. Thus, agents are more likely to report only when a corrupt exchange occurs. Through our experiment, we aim to find the effectiveness of reward schemes in contexts where (1) citizens can report honest officials, (2) evidence threshold is either high or low, and (3) judicial accuracy is either high or low.

Our paper is the first attempt to assess the effectiveness of reward schemes in curbing harassment bribes. In the domain of harassment bribes, previous work has chiefly concentrated on the role of full or partial leniency policies. We extend the literature by allowing individuals to make false reports to see whether the citizens exploit the reward schemes or not. Whistleblowing here arises in the context of interactions between individuals and not between corporations or government institutions. Initiation of bribes in such interactions is rampant in the developing world. Evidence threshold is introduced as a monetary cost of reporting the official. We avoid using the real effort tasks to keep things simple because it is difficult to quantify these costs in such a setting. We introduce variation in judicial accuracy by varying the exogenous probability of conviction and exoneration under true and false reports.

We find that more bribes are initiated if the accuracy of the judicial system is low. The average bribe asked is also significantly high under the less accurate judicial system. Citizens report a higher fraction of officials when the evidence threshold is low. With the increase in the evidence threshold, there is a significant decrease in false reports. However, we do not find a significant decrease in honest reporting. Finally, whistle-blowing accurately identifies corrupt behaviour when improved judicial accuracy complements a high evidence threshold.

The rest of the paper is organized as follows. Section 2 provides a review of the theoretical and experimental literature. Section 3 describes the experimental design and stages. Section 4 provides the results, and Section 5 discusses the results. Finally, Section 6 concludes the paper and provides suggestions for future work.

[2]The term was popularised by Kaushik Basu in his note to India's Ministry of Finance while he was serving as the chief economist of the Indian government. In that note, he made a point that for a particular class of bribes, i.e. harassment bribes, the act of bribe giving should be legalized.

[3]For instance, Singapore awards officials for refusing bribes and exposing their clients (Svensson, 2005). In countries like the United States, the United Kingdom, Germany, and India, bribe-giver and bribe-taker are equally culpable and face punishment. China, Russia and Japan use asymmetric liabilities where bribe-giver gets a milder punishment than bribe-taker. Most of the support for financially rewarding the whistle-blowers comes from the success of the US, where whistle-blower reward schemes have been intact for a long time. In the UK, regulators are still very skeptical about the effectiveness of such programs, which is contrary to empirical evidence (Dyck, Morse, and Zingales, 2010).

2.2 Previous Literature

Empirical investigations of corruption and whistleblowing involve many difficulties. Indexes are used to rank a country based on the prevalence of corruption. These rankings are often based on subjective assessments, prone to biases and misreporting, like anecdotal evidence and survey studies (Svensson, 2005). The correlation between the indices and the individuals' and firms' actual experiences of corruption is also not that high (Treisman, 2007). It is in the interest of those involved in corruption to keep their actions hidden. So the object of study is not observable to the researcher, which renders the actual level of corruption to be higher than the observed level of corruption. The opportunities to blow whistles that arise from illegal exchanges are also underestimated. Measurement and identification challenges raise eyebrows on empirical studies as researchers can only observe corruption after it is detected and whistles after they have been blown. There has been an increase in the use of experimental methods to study whistleblowing and corruption as they provide a better alternative. Laboratory experiments enable researchers to observe corruption and whistleblowing directly. The controlled environment of such studies is helpful to identify the impact of various policy instruments such as changes in monetary and non-monetary incentives, protection against retaliation, and differences in judicial systems.

Initial experimental investigations of corruption focused on individual determinants of corruption. Frank and Schulze (2000) find that economics students take more bribes than other students, and this difference is probably due to self-selection than indoctrination. In another paper, the authors study the effect of deterrence on the intrinsic motivation of agents (Frank and Schulze, 2003). They find that, although monitoring reduces corrupt exchanges, it crowds out the intrinsic motivation for honesty. Some studies model a bribery scenario into the decision-making environment. Due to the inherently reciprocal nature of bribe exchanges, most studies use a version of trust games (Abbink, 2006). Abbink et al. (2002) use an interactive corruption game in the laboratory and find that reciprocity can lead to bribery exchanges. The possibility of a negative externality of a corrupt act has no impact on bribe incidence. However, the threat of punishment has a strong deterrent effect and significantly decreases corrupt exchanges. Further studies, Abbink (2004, 2005), find that governments, firms, and aid agencies can use staff rotation as a policy to reduce bribery. They do not find evidence for the 'fairness hypothesis' that officials paid less are more likely to be corrupt as they deem they have been treated unfairly.

González et al. (2004) explore a situation with two public officials independently deciding to approve a citizens application where one official can also delay the decision. They find that the officials with delaying power tend to demand a higher share and delay the decision when they receive a lower share than the other official. Greasing is found to be moderately efficient in speeding up the service.[4]

Several studies explore corruption and whistleblowing jointly. Studies on whistle-blowing can be divided into two distinct strands based on the relationship of the whistle-blower to illegal activity. One strand of literature focuses on whistle-blowing by bystanders who did not participate in the crime (Spagnolo, 2008). Breuer (2013) uses a tax frame to look at the role of rewards on individual willingness to blow the whistle. He investigates the role of whistle-blower rewards by comparing a tax regime without any whistle-blowing mechanism to three other regimes with a mechanism that differs in the whistle-blowing incentives. He finds that pecuniary rewards for whistle-blowing lead to an increase in tax evasion reporting. The results show that individuals are more likely to report in the presence of reward schemes, and this likelihood increases with the size of the reward. He does not find any crowding-out effect of the reward schemes on the intrinsic motivation of whistle-blowers. They further find that subjects correctly anticipate this behaviour, and thus, tax evasion is significantly lower in the presence of such reward schemes.

In an experimental setting, Makowsky and Wang (2018) investigate the relationship between organizational shape and corruption. They find that the rates of corruption and whistleblowing increase with the number of levels in the organizational structure. Moreover, flat (no hierarchy) and pyramid structures (where leaders lead by example) effectively reduce corruption than other structures.

The paper by Reuben and Stephenson (2013) investigates the motivation of individuals to report group members who cheat. The setting is such that each group member, including honest, bears the cost of whistleblowing. They find that a sufficiently large number of individuals report lies that in fixed groups lying becomes unprofitable. When individuals can select group members, they ignore those who report lies which gives rise to groups where lying is common, and reporting is absent. Carpenter et al. (2017) experimentally create a corporate setting to investigate the whistle-blowing behavior of peers in the same firm. They find that a firm can motivate co-workers to report shirking employees by sharing profits with the whistle-blowers. Abbink and Wu (2017) study the role of betrayal or fear of betrayal in preventing collusive

[4]Grease payment is a sum of money, typically a harassment bribe, paid to a government official or a business person to facilitate or expedite some decision or transaction. The payment is not intended to influence the outcome of the official's action, only its timing.

bribery. They investigate bribery under regimes where both parties can self-report, or only one party can do so. The client or official can get amnesty and a financial reward by self-reporting on illegal activities. They find that allowing both to self-report reduces the exchange of bribes and deters corruption. The effect, though still positive, is not that strong when the two parties are aware of a possible future interaction. Allowing only one to self-report does not significantly reduce corruption.

Most relevant to our paper is the study by Abbink et al. (2014) on the role played by leniency, introduced as asymmetric liability, in deterring harassment bribes. They find that providing legal impunity only to bribe-givers reduces corruption. However, asymmetric liability may face challenges as incentives for bribe-givers are often not strong, and bribe-takers can retaliate against any whistle-blowing. We extend the work of Abbink et al. (2014) by allowing for spiteful behavior by citizens whereby they can use asymmetric rewards to report honest officials. In our experiment, we use rewards rather than leniency to encourage reporting. Full leniency returns the bribe amount to the bribe-giver and reduces the bribe-asker's payoff by the bribe amount.

Dufwenberg and Spagnalo (2011) develop a theoretical model to test the effectiveness of leniency policies involving asymmetric liability. They find that the quality of indigenous institutions strongly influences the effectiveness of such policies. In countries where the legal system and bureaucracy are efficient, i.e. where the cost of reporting is low, and the bureaucrat's opportunity cost of shirking is high, leniency policies involving asymmetric liability can effectively deter bribes. Our treatments on evidence threshold and judicial accuracy can shed light on the differences in the effectiveness of rewards schemes under different judicial systems.

2.3 Experimental Design

The experiment consists of four stages, revealed to the subjects in succession. Sequentiality ensures that their decisions in the earlier stages are not affected by strategic considerations regarding the later stages.

2.3.1 Stage 1: Encryption Task

In the first stage, subjects have the opportunity to earn a prize worth 600 Lab Sterling by completing an encryption task (see Appendix B.2.1 and Appendix B.3.1). The task is due to Malik et al. (2018), a modified version of the task introduced by Erkal and others (2011). Subjects are required to encrypt five words in five minutes to earn the prize. The computer screen presents them an encryption table assigning a unique number to each of the 26 letters of the English alphabet. On the same screen, they use the table to encrypt words of English alphabets. To win the prize, they must complete at least five encryptions. In order to contain the sample size, the task is designed to be easy enough so that everyone can win the prize. Once they have correctly encrypted a word, the screen presents the next word. At the end of the task, they move to the payoff screen, which displays their prize earnings. This stage aims to provide subjects with an income for the bribery game. Rather than giving them the endowment, we used a real effort task to create a sense of entitlement among subjects. This sense of belonging is crucial for our experiment so that any bribe asked is considered harassment.

2.3.2 Stage 2: Bribery Game

The next stage is a stylized harassment bribery game developed in the spirit of Abbink et al. (2012). We use a neutral frame in our instructions to make sure the wording is not suggestive of a right answer (see Appendix B.2.2 and B.3).[5] In all references to a bribe, we use the word *charge* instead. Before describing the game, consider two real-life contexts where harassment bribes may occur. First, imagine a situation where a citizen is owed a rebate on their tax return. They have to follow a bureaucratic procedure that requires them to interact with a government official to get this money. The official has absolute discretionary powers to provide the rebate. Second, imagine a citizen who wants to travel abroad and makes an application for their passport. They have to interact with a specific government official who has discretionary powers to reject the application or delay the decision. Public officials may seek bribes to provide the required service in these situations.

Figure 2.1 describes the game. At the end of the encryption task, we show subjects their prize earnings and divide them into two-member groups. In each group, one member is assigned the role of *player A* and the other the role of *player B*. Player A mimics the role of a government official, whereas player B mimics the role of a citizen. The game is an interaction of sequential decisions where a citizen is matched with a government official. Prizes for both members are given to the official who can ask the citizen to pay a

[5] Though, Abbink and Henning-Schmidt (2002) show that a loaded frame does not lead to a significantly different level of corruption.

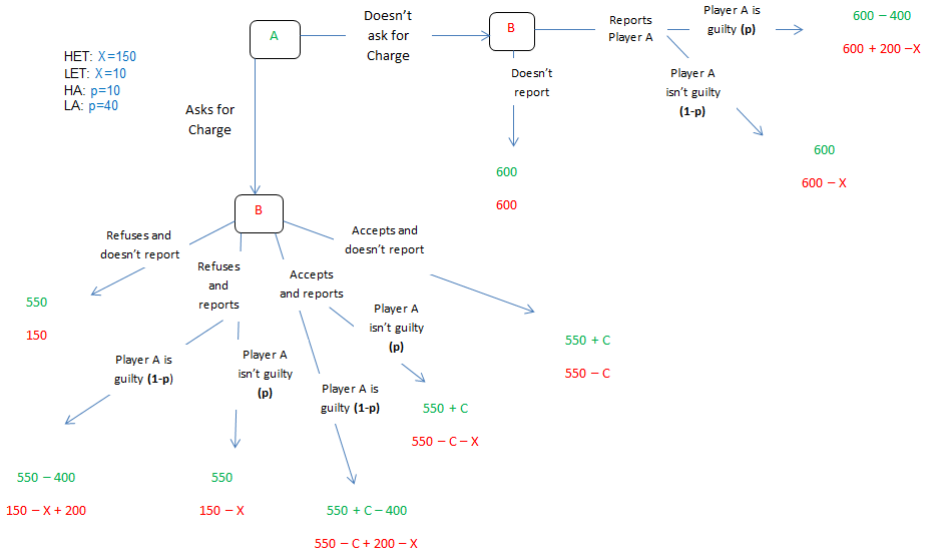


Figure 2.1: Decision tree for the bribery game.

charge to receive their prize. The Official moves first and is given the option to demand a charge ($C \in \{0,1,..,200\}$) in increments of *1* from the citizen. The citizen moves next and is informed if the official has demanded a charge or not. If demanded, they are also informed about the size of the charge. When a charge is demanded, the citizen can choose one of the four possible actions: (a) refuse the charge and do not report, (b) refuse the charge and report, (c) accept the charge and do not report, and (d) accept the charge and report. The exchange takes place if the citizen accepts to pay the charge. If the official does not ask for a charge, the citizen can choose between two options: (a) report the official, and (b) do not report the official. Reporting is costly for the citizen and depends on the level of evidence threshold. Under a high evidence threshold the cost of reporting is $X = 150$ Lab Units. Whereas, under a low evidence threshold the cost of reporting is $X = 10$ Lab Units. Evidence threshold represents the amount of effort and resources the citizen must exhaust to take up the case with the court of law. A higher threshold means more effort and resources are required and therefore entails a higher cost paid by the citizen. Once the evidence threshold is reached, i.e. when the citizen pays the cost, a random draw implements the decision. The probability of success and failure depends on the accuracy of the judicial system. In a more accurate judicial setting, the probability of type-1 (convicting the innocent) and type-2 (exonerating the guilty) error rates is lower than in the less accurate setting. Specifically, the error rates are 10 percent under a more accurate judicial system and 40 percent under a less accurate judicial system. If the official is declared guilty after being reported, the system imposes a fine of 400 Lab units on the official and bestows a reward of 200 Lab units to the citizen for blowing the whistle. If the citizen refuses to pay the bribe, their payoff reduces by three times as much as that of the official. If the citizen does not receive the product/service, there will be an extreme and immediate adverse effect on their utility. The payoff of the official decreases slightly to capture the social efficiency losses due to corruption (Abbink et al., 2002).
The game is played for 12 rounds. Players' roles (official or citizen) remain the same throughout the experiment. In each round, we randomly rematch an official with a citizen. We reveal the repeated nature of the game only after the end of round 1, which enables us to analyze behaviour in the stage game and the repeated game.

2.3.3 Stage 3: Norm Elicitation Task

In this stage, we use a framework introduced by Krupka and Weber (2013) to directly elicit norms for the two treatments (see Appendix B.3.8 and B.3.9). In this task, we present subjects with three scenarios in which an anonymous individual 'X' has to choose among several available actions. Subjects provide their social appropriateness ratings corresponding to each possible action on a four-level Likert scale: very socially appropriate, somewhat socially appropriate, somewhat socially inappropriate, and very socially inappropriate. We incentivize their responses using a coordination game. At the end of the task, a random draw selects one of the three scenarios and one of the possible actions. Subjects earned £2 if their response was the same as the one selected by most other subjects in their session. Otherwise, they earned £0.

2.3.4 Stage 4: Questionnaire

At the end of the experiment, subjects filled a questionnaire on demographics, risk preferences, institutional and inter-personal trust, attitudes towards bribery, and political orientation (see Appendix B.1.1).

Treatment	Label	Cost of Reporting	Rate of error
1	HIGH-JA & LOW-ET	10 Lab Units	10 percent
2	LOW-JA & HIGH-ET	150 Lab Units	40 percent
3	HIGH-JA & HIGH-ET	150 Lab Units	10 percent
4	LOW-JA & LOW-ET	10 Lab Units	40 percent

Table 2.1: Treatment Parameters. Cost of reporting is an indicator of the evidence threshold and rate of error is an indicator of the quality of the judicial decision making.

2.3.5 Treatments

We use a 2x2 between-subject design that varies the judicial accuracy (high or low) and the evidence threshold (high or low). The probability of type-1 and type-2 errors determine the judicial system's accuracy, and the cost of reporting an official determines the evidence threshold. Both factors have two levels, i.e. high and low. Table 1 summarizes the treatment parameters. 2.1 We now explain the four treatments.
In the treatment with low evidence threshold and high judicial accuracy (treatment 1), citizens pay 10 lab units to make sure the court takes up their report. The probability of type 1 and type 2 errors is 10 percent. Under these treatment conditions, player B should always report irrespective of whether player A demands a bribe or not. Player A, in anticipation, should not ask for a charge. Under high evidence threshold and low judicial accuracy (treatment 2), citizens pay 150 lab units to ensure the court takes up their report. Here, the probability of type 1 and type 2 errors is 40 percent. Under these conditions, the citizen should never report the official whether the latter asks for a charge or not. Knowing this, the official should then demand the highest bribe.
In the treatment with high evidence threshold and high judicial accuracy (treatment 3), citizens pay 150 lab units to ensure the court takes up their report. The probability of type 1 and type 2 errors is 10 percent. The payoff structure is such that the citizen should only report when the official demands a bribe. Then the best response of the official is not to demand the bribe. The last treatment involves a low evidence threshold and a low judicial accuracy (treatment 4). Here, the cost of reporting for the citizen is 10 lab units and the probability of type 1 and type 2 errors is 40 percent. Under these conditions, the citizens should always report, and the official, in anticipation, should demand the highest bribe.
These treatments aim to estimate the effect of evidence threshold and judicial accuracy on bribes demanded and officials reported. Treatments 1 and 3 jointly represent the high judicial accuracy environment, whereas treatments 2 and 4 jointly represent the low judicial accuracy environment. Similarly, treatments 2 and 3 represent the high evidence threshold situation and treatments 1 and 4 represent the low evidence threshold situation. The following section provides an equilibrium analysis of the stage game and identifies equilibrium predictions.

2.3.6 Other Details

As standard in the literature, we assume risk-neutral players. However, the equilibrium predictions remain the same for moderate to high levels of risk aversion (Holt and Laury, 2002).[6] We use backward induction to find the equilibrium. In the bribery game, the citizen makes the last move. If player A does not ask for a charge, then the utility of player B from reporting player A is

$$U_B(R, C = 0) = p \times (600 + 200 - x) + (1 - p) \times (600 - x)$$

and the utility of player B from not reporting player A is

$$U_B(NR, C = 0) = 600$$

Therefore, player B reports player A if and only if

[6] Assuming constant relative risk aversion and utility specification, $U(x) = \frac{x^{(1-r)}}{1-r}$, the predictions hold up to r=0.97.

$$x < 200 \times p$$

If player A asks for a charge, then it is never optimal for player B to refuse the charge. The utility of player B from accepting the charge and reporting player A is

$$U_B(AR, C > 0) = p \times (550 - C - x) + (1 - p) \times (550 - C + 200 - x)$$

and the utility of player B from accepting the charge and not reporting player A is

$$U_B(ANR, C > 0) = 550 - C$$

Therefore, player B reports player A if and only if

$$x < 200 \times (1 - p)$$

In Table 2.2, we provide the best response function of the citizen (player B). When the evidence threshold is low, the citizen always reports the official irrespective of the accuracy of the judicial decisions. Under the high evidence threshold, the citizen does not report at all when judicial accuracy is low. Under the high evidence threshold and high judicial accuracy, citizen reports only after being asked for a charge.
[H2.3.6] Whistle-blowing is high when the evidence threshold is low (treatments 1 and 4) and low when the evidence threshold is high (treatments 2 and 3). We now turn to the behaviour of player A. If the evidence threshold is low and the judicial accuracy is low, then player B chooses RR, i.e. reports whether asked for a charge or not. Player A should then choose to ask for a charge as below

$$0.4 \times (550 + C) + 0.6 \times (550 + C - 400) > 0.4 \times (600 - 400) + 0.6 \times (600)$$
$$\forall \quad C > 130$$

When cost of reporting is sufficiently large and accuracy is low, player B chooses NrNr. Player A then asks for a charge as

$$(550 + C) > 600$$
$$\forall \quad C > 50$$

When cost of reporting is sufficiently small and accuracy is high, player B chooses RR. Player A then does not ask for a charge as

$$0.1 \times (600 - 400) + 0.9 \times (600) > 0.1 \times (550 + C) + 0.9 \times (550 + C - 400)$$
$$\forall \quad C < 200$$

When cost of reporting is sufficiently large and accuracy is high, player B chooses NrR. Player A then does not ask for a charge as

$$600 > 0.1 \times (550 + C) + 0.9 \times (550 + C - 400)$$
$$\forall C < 200$$

Thus player A should ask for a charge when the judicial accuracy is low and should not ask for a charge when the judicial accuracy is high. As the player's utility is monotonically increasing in the amount of charge, he/she should ask for the maximum amount of charge under the less accurate judicial system.
[H2.3.6] Irrespective of the evidence threshold, bribe incidence is higher under a judicial environment with low accuracy (treatments 2 and 4) than under a judicial environment with high accuracy (treatments 1 and 3).

Table 2.2 provides the complete equilibrium profile for the four treatments. The level of evidence threshold affects the probability of occurrence of false positives and false negatives. If the threshold is set high, fewer innocent officials will be reported, which also leads to a fall in the number of guilty officials reported. There is a possibility that we could move from a situation where citizens always report officials to a situation

	High Judicial Accuracy	*Low Judicial Accuracy*
High Evidence Threshold	(C = 0 , NrR)	(C = 200, NrNr)
Low Evidence Threshold	(C = 0, RR)	(C = 200, RR)

Table 2.2: Equilibrium profile of all treatments. C=0 means the official does not ask for a bribe, C=200 means the official asks for the maximum possible bribe, RR means citizen reports the official irrespective of whether they ask for a bribe or not, NrNr means citizen does not report the official whether they ask for a bribe or not, and NrR means they report the official only if they ask for a bribe.

where they never report officials. According to the expected utility framework, this occurs when the underlying judicial system is prone to a high error rate. This can be overcome by improving the quality of the judiciary such that the citizens only report when the official demands a bribe. [H2.3.6] Whistle-blowing is optimal, i.e. it identifies bribery when the evidence threshold and judicial accuracy are both high (treatment 3).

	LOW-ET HIGH-JA	**HIGH-ET HIGH-JA**	**LOW-ET LOW-JA**	**HIGH-ET LOW-JA**
Mean Age	21.95	24.61	23.57	22.97
% Female	0.66	0.92	0.65	0.55
% Student	0.97	0.57	0.88	0.65

Table 2.3: Sample Characteristics.

		HIGH-ET	**LOW-ET**	**OVERALL**
Average Bribe Asked	HIGH-JA	110.00	136.33	125.80
	LOW-JA	96.67	121.00	105.36
	OVERALL	99.09	126.75	110.74
% Officials Asked Bribe	HIGH-JA	7.14	9.67	8.47
	LOW-JA	21.95	14.71	18.67
	OVERALL	15.94	12.30	14.17
% Citizens Who Report	HIGH-JA	7.14	32.26	20.34
	LOW-JA	17.07	44.11	29.33
	OVERALL	13.04	38.50	25.37
% False Reports	HIGH-JA	0.00	22.58	20.34
	LOW-JA	9.76	35.29	29.33
	OVERALL	5.80	29.23	17.16

Table 2.4: Descriptive Statistics.

2.4 Results

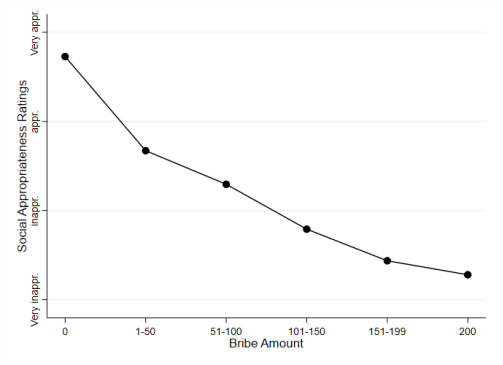
The experiment was programmed using z-tree and was conducted at ESSEXLab, a social sciences research laboratory at the University of Essex (Fischbacher, 2007). Subjects were recruited through the ESSEXLab Online Recruitment System. We ran a total of 14 sessions where 268 subjects participated in the experiment. Table 2.3 shows sample characteristics for these subjects.

2.4.1 Perceived Norm

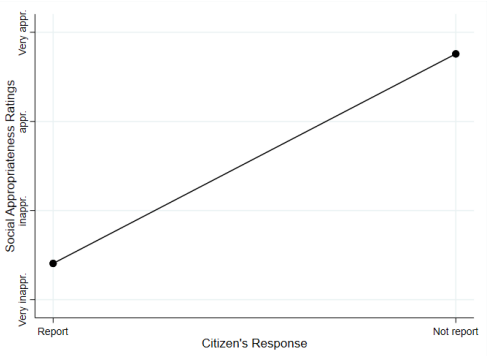
The results from the norm elicitation stage are provided in Figure 2.2. Figure 2.2(a) gives us the perceived norm related to the official's behaviour. It emerges that subjects believe that not asking for a bribe is the most socially appropriate action. The approval ratings go down with the bribe amount such that choosing the maximum possible bribe amount is rated as the most socially inappropriate action.

Figure 2.2(b) gives the perceived norm related to the citizen's behaviour when the official did not ask for a bribe. Citizens had two possible actions to choose from: report the official and not report the official. The mean appropriateness ratings for reporting and not reporting are 0.83 and -0.72, respectively. The figure shows a considerable agreement among subjects that not reporting the official is the most appropriate action in this situation.

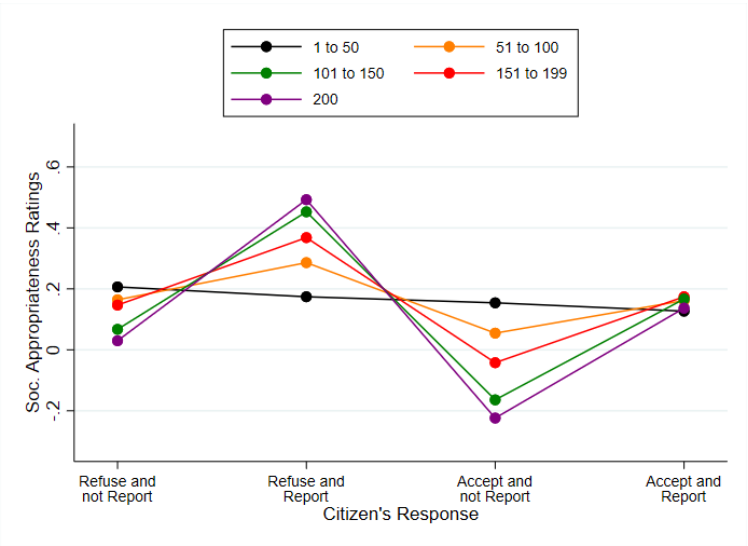
We also elicit norms related to citizens' responses for a few possible bribe ranges. In order to be succinct,



(a) Officials' actions



(b) Citizens' response when not asked for a bribe



(c) Citizens' response when asked for a bribe

Figure 2.2: **(a)** subjects rate the actions of officials: *0* means the official did not ask for a bribe, and *200* means the official asked for the maximal bribe. **(b)** subjects rate the actions available to citizens when the official had asked for a bribe. **(c)** subjects rate the actions available to citizens for different possible level of bribes asked by the official.

we plot them all together in Figure 2.2(c). It can be seen that for the lowest bribe range, i.e. from 1 to 50, there is no apparent difference between the appropriateness of each action. We can see that not reporting the bribe is considered less appropriate than reporting the bribe for higher bribe ranges. However, these differences are insignificant, and we cannot say that subjects perceive a distinct norm under these scenarios.

2.4.2 Official's Behaviour

We start with the official's behaviour and look at the level of bribe incidence. About 14.2 percent of officials ask for a bribe in all four treatments. One reason for such a small fraction of bribe demands is that subjects had to decide at two decision points (Soman et al., 2010).[7] First, they had to make a binary choice whether to ask for a bribe or not. Those who had answered in the affirmative then stated the bribe amount. [8]

According to the model presented in section 3.6, we expected a high incidence of bribe demands in a setting with low judicial accuracy compared to those with high judicial accuracy. Treatments 1 and 3 correspond to a high judicial accuracy setting, whereas treatments 2 and 4 correspond to the setting with low judicial

[7]Soman and others (2010) find that partitioning the decision into more decision points leads consumers to pause and deliberate on their decisions at each decision point. In our case, subjects had to decide between two choices at the first decision node. If they decided to ask for a bribe, then the foregone option was not to ask a bribe. However, if they had only one decision node and decided to ask a bribe x, there are $200 - x$ potentially more unfair choices.

[8]This contrasts to a decision environment with only one decision point where they choose the bribe amount from 0 to 200.

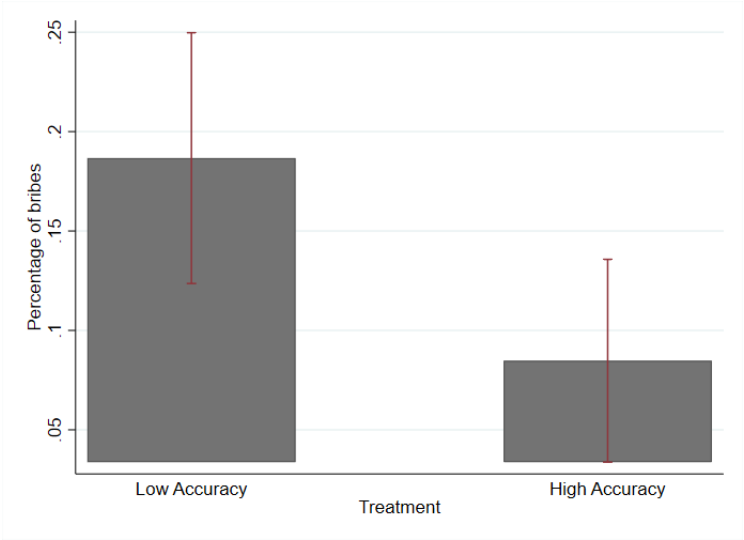


Figure 2.3: Bar graph showing the percentage of bribes initiated under low and high judicial accuracy treatments.

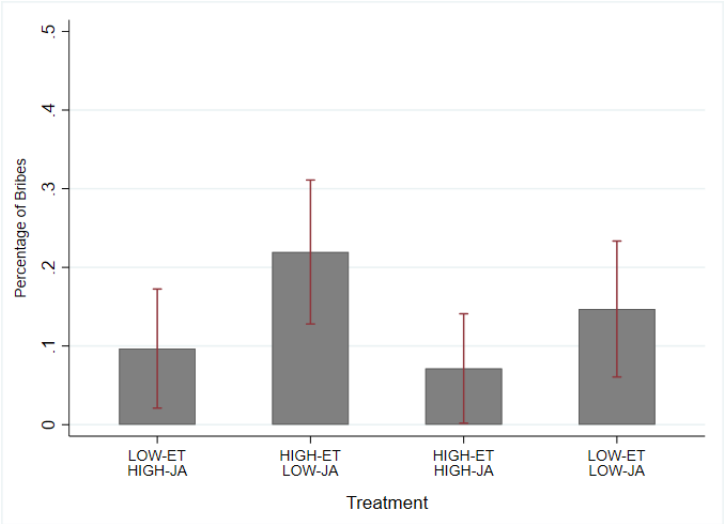


Figure 2.4: Bar graph showing the percentage of bribes initiated under all four treatments.

accuracy. Figure 2.3 shows the bar graph for the high and low judicial accuracy treatments. In our data, 18.7 percent of subjects ask for a charge under the low judicial accuracy treatment, and only 8.5 percent ask for a bribe under the high judicial accuracy treatment. The difference is weakly significant according to the fisher's exact test ($p = 0.075$).
We also present the percentage of bribes initiated across all four treatments (see Figure 2.4). We can see that the differences are greater when we vary the accuracy of judicial decisions than the evidence threshold. It shows that officials' behaviour is more responsive to judicial accuracy than the evidence threshold. The results are in line with the theoretical prediction. The average amount of bribes asked under the high accuracy and the low accuracy treatments are 10.6 and 19.6, respectively. A Wilcoxon rank-sum test shows that the difference in bribe amounts is statistically insignificant ($z = 1.600; p = 0.109$). We also find that subjects do not choose between all possible bribe amounts, and there are only nine different amounts of bribes they choose. So, we also perform a chi-squared test and find that the difference in various bribe levels chosen across the two treatments is also statistically insignificant ($\tilde{\chi}^2 = 11.27$; $p = 0.187$).

2.4.3 Citizen's Behaviour

Now we turn to the behaviour of player B, who performed the role of a citizen in our game. In the experiment, treatments 1 and 4 correspond to the low evidence threshold setting, whereas treatments 2 and 3 correspond to the high evidence threshold setting. In order to test our hypothesis related to citizens'

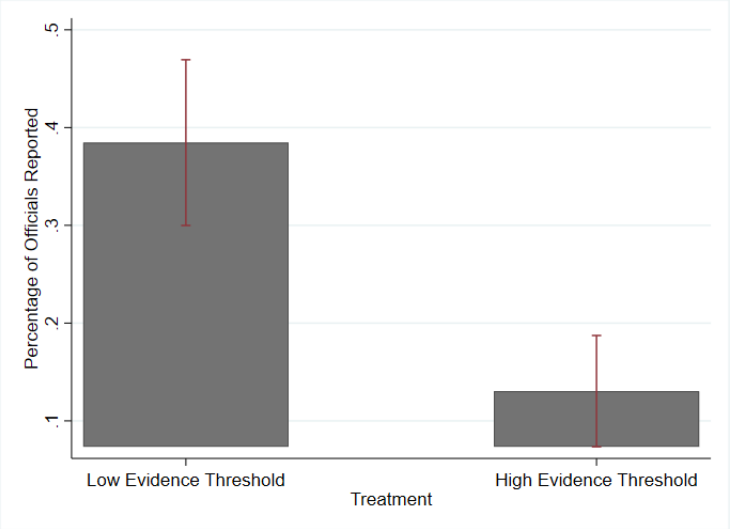


Figure 2.5: Bar graph showing the percentage of officials reported under low and high evidence threshold treatments.

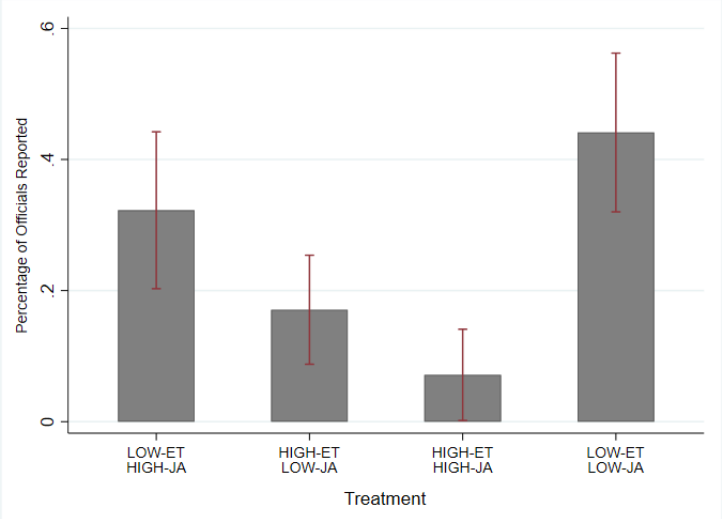
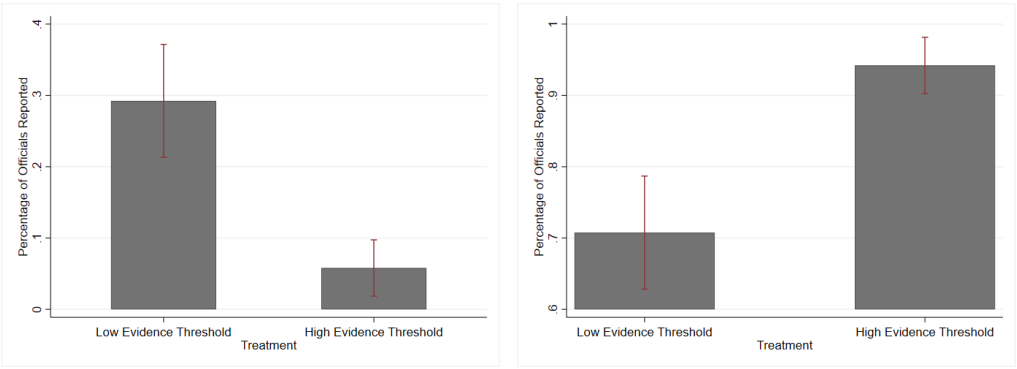


Figure 2.6: Bar graph showing the percentage of officials reported across all four treatments.

reporting behaviour, we merge treatments 1 and 4 into one treatment and treatments 2 and 3 into another treatment. In Figure 2.5, we provide the reporting behaviour across the merged treatments. It shows that about 38.5 percent of citizens report officials under the low evidence threshold, whereas only 13 percent report the officials under the high evidence threshold (Fisher's exact; $p = 0.001$).

We also present the percentage of officials reported across all four treatments (see Figure 2.6). We can see that the differences are greater when we vary the evidence threshold than when we vary the accuracy of judicial decisions. It shows that citizens' behaviour is more responsive to the evidence threshold than judicial accuracy. However, all of the reporting is not justified. We would thus like to differentiate between truthful (reporting when asked for a bribe and not reporting when not asked) and false reporting (not reporting when asked for a bribe and reporting when not asked). Figures 2.7(a) and 2.7(b) respectively provide the percentage of true and false reports when the evidence threshold is low or high. A Fisher's exact test performed separately for the two cases reveals that the false reporting significantly decreases (from 29.2 percent to 5.7 percent; $p < 0.001$) when the evidence threshold is increased. On the contrary, truthful reporting increases significantly (70.7 percent to 94.2 percent ; Fisher's exact $p < 0.001$) when the evidence threshold increases.

One of the standard utility model predictions was that reporting would be optimal, i.e. it identifies bribery when high evidence threshold is accompanied by improvement in judicial accuracy. Figure 2.8 provides a bar graph showing the percentage of honest reports across the four treatments. The honest report takes the value 1 when the citizen reports a corrupt official and does not report an honest official. Reporting is the



(a) False Reports (b) Honest Reports

Figure 2.7: **(a)** Percentage of officials falsely reported **(b)** Percentage of officials honestly reported.

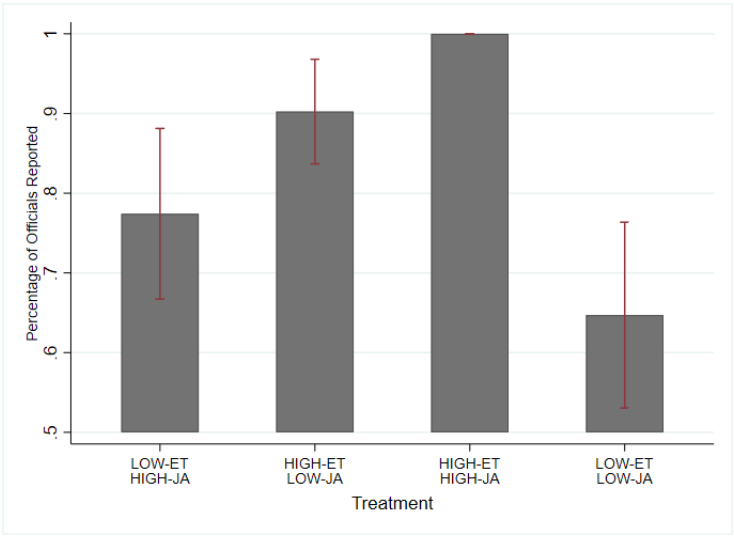


Figure 2.8: Bar graph showing the percentage of honest reports under all four treatments.

least honest when citizens have to pay a high cost of reporting, and the judicial accuracy is low. However, reporting is the most honest, or it perfectly identifies bribery under the treatment where an improved judicial system accompanies high evidence threshold. A chi-squared test performed to check if honest reporting is significantly different across the treatments confirms our predictions ($\tilde{\chi}^2 = 15.88$; $p = 0.001$). This provides evidence for the theoretical prediction of our expected utility model and that of Spagnolo et al. (2017). They show that given the same fine to rewards ratio, when an increase in the burden of proof accompanies an improvement of judicial accuracy, we move from a point where there is over-reporting to a point where officials are only reported if they perform a corrupt act.

2.4.4 Regression results:

Probit regression in Table 2.5 affirms the effectiveness of improvement in Judicial accuracy to act as a deterrent for officials. The coefficient on the treatment dummy is significant and negative. In specification 2, we control for age, gender, and risk preferences. The effect of judicial accuracy persists and improves . In specification 3, we also control for additional factors such as whether

(a) they think it is justified for an official to ask for a bribe (official bribe),

(b) they think it is justified for a citizen to pay a bribe to obtain a favour (citizen bribe), and

(c) whether they were ever asked to pay a bribe (paid a bribe).

The coefficients on the latter two are insignificant. In contrast, the variable 'official bribe' is positive and moderately significant ($p = 0.08$), which means that those subjects who said it was justified for an official

to ask for a bribe are weakly more likely to ask for a bribe themselves.

In Tables 2.6 and 2.7, we report the results of regressions that look at citizens' behaviour. The theoretical

	Bribe Asked		
	(1)	(2)	(3)
HIGH-JA	-0.1057*	-0.1543**	-0.1484*
	(0.0598)	(0.0572)	(0.0696)
Age		0.0034	0.0045
		(0.0029)	(0.0028)
Gender		-0.0175	-0.0286
		(0.0628)	(0.0530)
Risk		0.0113	0.0126
		(0.0156)	(0.0184)
Official bribe			0.0993*
			(0.0411)
Constant	-0.8902***	-1.3568**	-2.2324***
	(0.167)	(0.538)	(0.873)
Paid a bribe			YES
Citizen bribe			YES
Observations	134	112	112
Pseudo R-squared	0.0270	0.0638	0.1489
Log-likelihood	-53.2238	-46.2276	-42.0252

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 2.5: Probit regression of bribe asked (binary: 1 means the official asked for a bribe) on treatment dummy ('HIGH-JA'), demographics, and other controls.'HIGH-JA' is equal to 1 when the judicial accuracy is high. 'Gender' is equal to 1 for female. The variable '*Official bribe*' asks subjects whether it is acceptable for an official to ask for a bribe. The variable '*Citizen bribe*' asks whether it is acceptable for a citizen to pay a bribe to the official for a service they are entitled to receive. They both take the values from 1 to 5 where 1 means never justified and 5 means always justified. The variable '*Paid a bribe*' is a dummy variable which equals 1 if the subject was ever asked, in real life, to pay a bribe.

prediction for the citizen's behaviour was that reporting would decrease under the high evidence threshold. We create a variable (Report) that takes a value of 1 if the citizen reported the official and 0 if the citizen did not report the official. In Table 2.6, we produce the results from a probit regression. In the first column, we regress the variable measuring the reporting decision on the treatment dummy (which takes the value 1 if the evidence threshold is high and the value 0 if the evidence threshold is low). The treatment dummy is found to be significant at a 1 percent level of significance ($p = 0.007$). The direction of the effect is in line with the theoretical prediction such that reporting decreased significantly with the increase in the evidence threshold. When we control for factors (in specification 2) such as age, gender, and risk preferences, the coefficient on the treatment dummy is still significant though only at a 5 percent level of significance. The treatment effect is robust to the inclusion of additional factors such as the subject's political orientation, whether they ever paid a bribe, and whether they think it can ever be justified for an official (citizen) to ask (pay) a bribe (see specification 3).

In Table 2.7, we filter the data to look only at the cases where citizens falsely reported the officials. We find that the treatment effect is negative and strong. In the first column, we regress the reporting dummy (which is equal to 1 if the report is false) on the treatment dummy (which takes the value 1 if the evidence threshold is high and the value 0 if the evidence threshold is low). The treatment dummy is found to be significant at a 1 percent level of significance ($p = 0.006$). The direction of the effect is again in line with the theoretical prediction such that false reporting decreases significantly with the increase in the evidence threshold.[9] When we control for factors (in specification 2) such as age, gender, and risk preferences, the coefficient on the treatment dummy is still significant though only at a 5 percent level of significance. The treatment effect is also robust to the inclusion of additional factors such as the subject's political orientation, whether they ever paid a bribe, and whether they think it can ever be justified for an official (citizen) to ask (pay) a bribe (see specification 3; $p = 0.012$). The coefficient on political orientation is significantly and negatively related to the reporting decision of the citizens. Those who put themselves on the right side of the political spectrum are less likely to falsely accuse an official of demanding a bribe. This relationship between political orientation and the reporting decision was not present when we looked at overall reporting (true as well as false).

[9]The Conversely, honest reporting increases when we increases the evidence threshold.

	Citizen Reports		
	(1)	(2)	(3)
HIGH-ET	-0.2447***	-0.1772**	-0.1745**
	(0.0813)	(0.0799)	(0.0798)
Age		0.0121	-0.0099
		(0.0082)	(0.0077)
Gender		-0.2215***	-0.2114***
		(0.0648)	(0.0723)
Risk		0.0123	0.0087
		(0.0156)	(0.0171)
Polit			0.0095
			(0.0474)
Constant	-0.2933	0.7479	0.0282
	(0.1989)	(0.7954)	(0.6706)
Paid a bribe			YES
Citizen bribe			YES
Observations	134	112	112
Pseudo R-squared	0.0774	0.1179	0.1297
Log-likelihood	-70.0257	-59.8976	-59.0964

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 2.6: Probit regression of 'Citizen Reports' (binary: 1 means the citizen reports the official) on treatment dummy ('HIGH-ET'), demographics, and other controls. 'HIGH-ET' is equal to 1 when the evidence threshold is high 'Gender' is equal to 1 for female. The variable 'Citizen bribe' asks subjects whether it is acceptable for Citizens to pay a bribe to receive a government service. It takes the value from 1 to 5 where 1 means never justified and 5 means always justified. The variable 'Paid a bribe' is a dummy variable which equals 1 if the subject was ever asked to pay a bribe. The variable 'Polit' measures the political orientation of subjects. It takes the value from 0 to 5 where 0 means on the far left and 5 means on the far right of the political spectrum.

	False Reports		
	(1)	(2)	(3)
HIGH-ET	-0.2321***	-0.1934**	-0.1843**
	(0.0786)	(0.0789)	0.0722
Age		0.0044	0.0040
		(0.0052)	0.0042
Gender		-0.1111**	-0.0977
		(0.0523)	0.0631
Risk		0.0407**	0.0372**
		(0.0179)	0.0187
Polit			-0.0632***
			(0.0226)
Constant	-0.5466**	-1.7826***	-1.4557**
	(0.2109)	(0.5830)	0.7092
Paid a bribe			YES
Citizen bribe			YES
Observations	134	112	112
Pseudo R-squared	0.1121	0.1407	0.1897
Log-likelihood -54.5463	-48.8649	-46.0791	

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 2.7: Probit regression of 'False Reports' (binary: 1 means the citizen falsely reports the official) on treatment dummy ('HIGH-ET'), demographics, and other controls. 'HIGH-ET' is equal to 1 when the evidence threshold is high 'Gender' is equal to 1 for female. The variable 'Citizen bribe' asks subjects whether it is acceptable for Citizens to pay a bribe to receive a government service. It takes the value from 1 to 5 where 1 means never justified and 5 means always justified. The variable 'Paid a bribe' is a dummy variable which equals 1 if the subject was ever asked to pay a bribe. The variable 'Polit' measures the political orientation of subjects. It takes the value from 0 to 5 where 0 means on the far left and 5 means on the far right of the political spectrum.

2.4.5 Multiple Period Interaction

We also exploited the option to look at continued interaction between pairs of subjects. The bribery stage was played for a total of twelve rounds.[10] The role of subjects remained the same. However, at the beginning of each round, they were randomly re-matched into groups of two to emulate an environment that is comparable to everyday experience where citizens and officials interact on more than one occasion. This

[10]In the first four sessions, we only had eight rounds of the bribery stage which were later increased to 12 to get a clear pattern of choices.

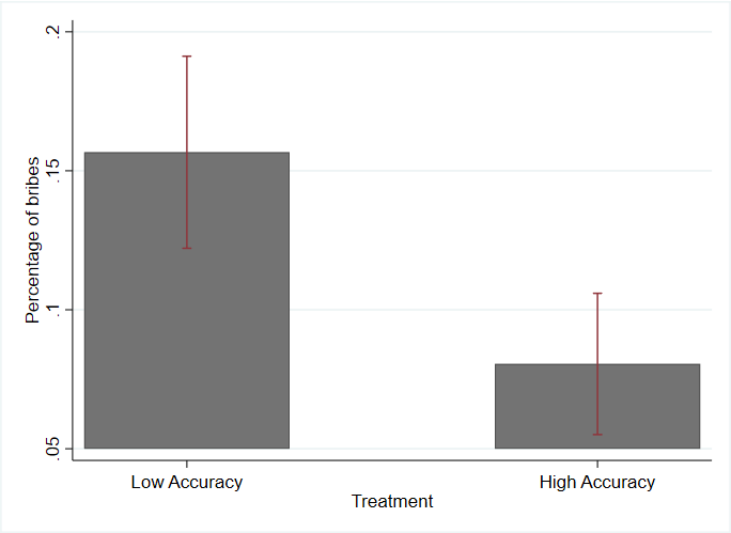


Figure 2.9: Bar graph showing the percentage of bribes initiated under low and high judicial accuracy treatments across all rounds.

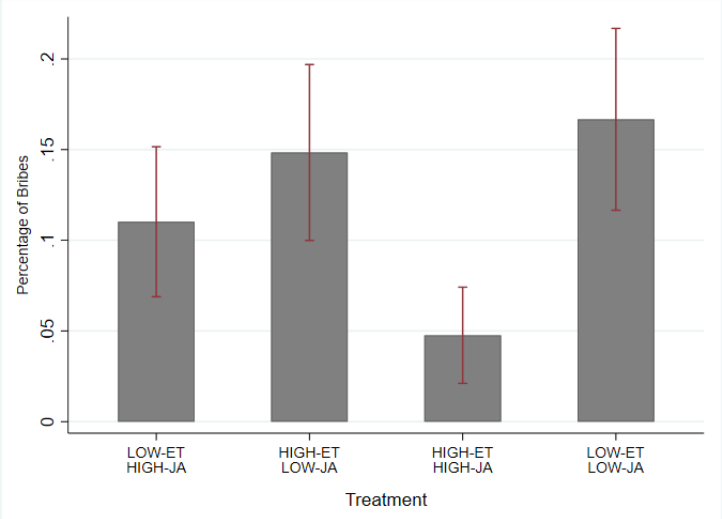


Figure 2.10: Bar graph showing the percentage of bribes initiated under all four treatments across all rounds.

allows us to learn how the experience of interactions among officials and citizens affect their future choices. We can also test if the behaviour in the stage game is sustained throughout the duration of the game or it evolves through a different trend.

Figure 2.9 shows the bar graph for the high and low judicial accuracy treatments. Across the twelve rounds, 15.16 percent of subjects ask for a charge under the low judicial accuracy treatment, and only 7.23 percent ask for a bribe under the high judicial accuracy treatment. The difference is strongly significant according to the fisher's exact test ($p < 0.001$). The results are in line with the theoretical prediction. In Figure 2.10, we provide the incidence of bribes across all four treatments to show that officials' behaviour responds more to the accuracy of judicial decisions than the evidence threshold.

In Figure 2.11, we look at the reporting behaviour of citizens across all 12 rounds. It shows that about 42.93 percent of citizens report officials under the low evidence threshold, whereas only 13.7 percent report the officials under the high evidence threshold (Fisher's exact; $p < 0.001$). In Figure 2.12, we provide the percentage of officials reported to have asked for a bribe across all four treatments. It shows that citizen's behaviour responds more to the evidence threshold than the accuracy of judicial decisions.

However, all of the reporting is not justified. We would thus like to differentiate between truthful (reporting when asked for a bribe and not reporting when not asked) and false reporting (not reporting when asked for a bribe and reporting when not asked). Figures 2.13(a) and 2.13(b) respectively provide the percentage of true and false reports when the evidence threshold is low or high. A Fisher's exact test performed separately for the two cases reveals that the false reporting significantly decreases (from 32.88 percent

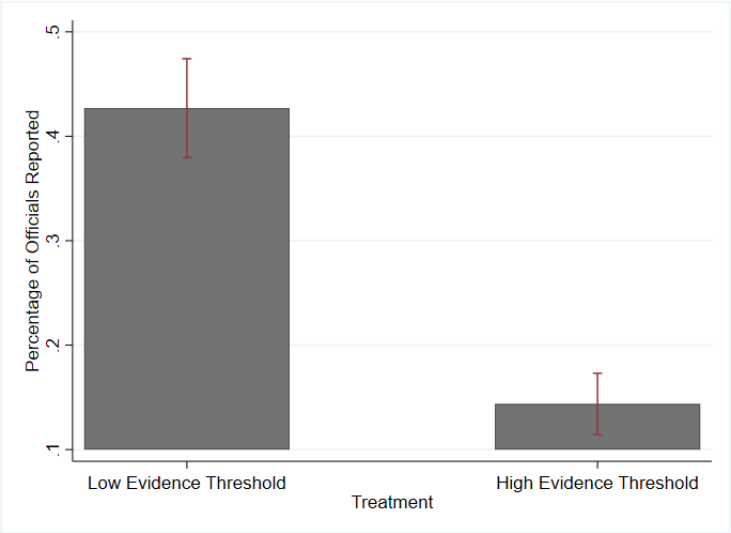


Figure 2.11: Bar graph showing the percentage of officials reported under low and high evidence threshold treatments across all rounds.

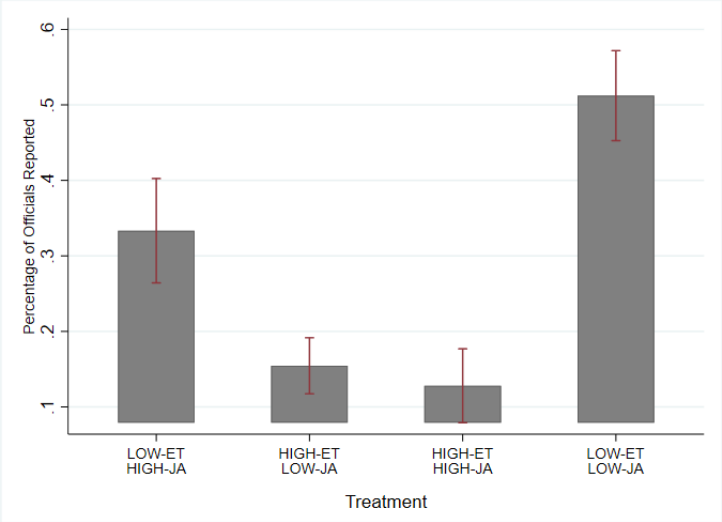
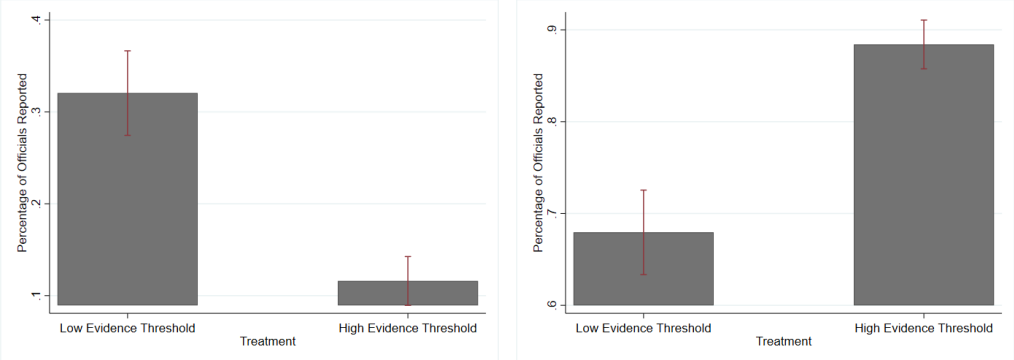


Figure 2.12: Bar graph showing the percentage of officials reported under all four treatments across all rounds.



(a) False Reports

(b) Honest Reports

Figure 2.13: **(a)** Percentage of officials falsely reported across all rounds. **(b)** Percentage of officials honestly reported across all rounds.

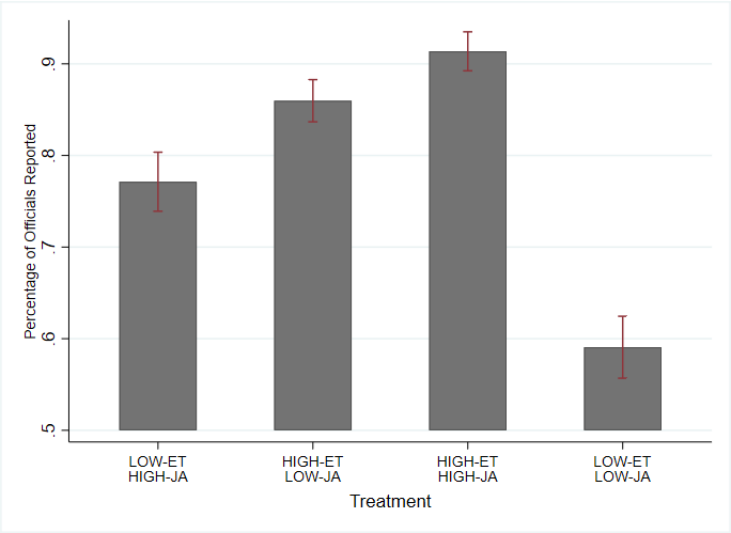


Figure 2.14: Bar graph showing the percentage of honest reports under all four treatments across all rounds.

to 11.63 percent; $p < 0.001$) when the evidence threshold is increased. On the contrary, truthful reporting increases significantly (67.12 percent to 88.26 percent ; Fisher's exact $p < 0.001$) when the evidence threshold increases.

Figure 2.14 provides a bar graph showing the percentage of honest reports across the four treatments. The honest report takes the value 1 when the citizen reports a corrupt official and does not report an honest official. Reporting is the least honest when citizens have to pay a low cost of reporting and the judicial accuracy is low. Our modeling showed that reporting is the most honest, or it perfectly identifies bribery under the treatment where an improved judicial system accompanies high evidence threshold. A chi-squared test performed to check if honest reporting is significantly different across the treatments confirms our predictions ($\tilde{\chi}^2 = 136.05$; $p < 0.001$). This provides evidence for the theoretical prediction of our expected utility model and that of Spagnolo et al. (2017).

In Figure 2.15, we look at the trend of bribe initiation between the treatments involving high judicial accuracy and low judicial accuracy. We would expect bribe incidence to decrease over time under high judicial accuracy. It emerges from the data that the difference between the levels of bribe incidence decreases in the first few rounds. It then starts increasing in the later rounds. More bribes are initiated under the low judicial accuracy treatment every round than under the high judicial accuracy treatment. Almost no bribes (0 percent and 2 percent) are initiated in the final few rounds under high judicial accuracy. In contrast, bribe incidence is comparatively higher under low judicial accuracy (15 percent and 10 percent, respectively). A Wilcoxon rank-sum test over 12 periods shows that bribe incidence is significantly lower under high judicial accuracy than low judicial accuracy ($p < 0.001$). The average bribe level is also significantly different across the two treatments (15 percent and 6 percent; $p < 0.001$).

The results show that a less accurate but functional judicial system can deter bribes over time to some extent. However, as the quality or accuracy of judicial decision making improves, this deters bribes to a much greater extent. The overall level of bribe incidence in our experiment is minimal. Even under the setting with high type 1 and type 2 errors, no more than 20 percent of officials ask for a bribe in any period.

We now look at the reporting decisions of citizens over the 12 periods under the treatments with low and high evidence thresholds. Figure 2.16(a) shows the reporting decisions of citizens under the setting with high and low evidence thresholds. About 38 (13) percent of officials are reported in the first period when the evidence threshold is low (high). The number of officials reported remains higher under the low evidence threshold than the high evidence threshold. By the last round, only 11 percent of officials were reported under the high evidence threshold, in contrast to 37 percent reported under the low evidence threshold.

It is worthwhile to distinguish between honest and false reports. Figure 2.16(b) shows the occurrences where the citizens made an honest report. In the stage game, the percentage of honest reports is 70.7 (94.2) percent under the low (high) evidence threshold. Over time, the reporting rate decreases under the high evidence threshold and reaches 85 percent in the last round. Under the low evidence threshold, the reporting rate is not very different in round 12 (70..4 percent) than in round 1 (70.7 percent).

Finally, Figure 2.16(c) shows the fraction of false reports made by citizens. Under the low evidence threshold treatment, false reports increase from 29.2 percent (first round) to 42.6 percent (tenth round) before declining to 29.2 percent (final round). Under the high evidence threshold treatment, 5.8 percent of the filed reports

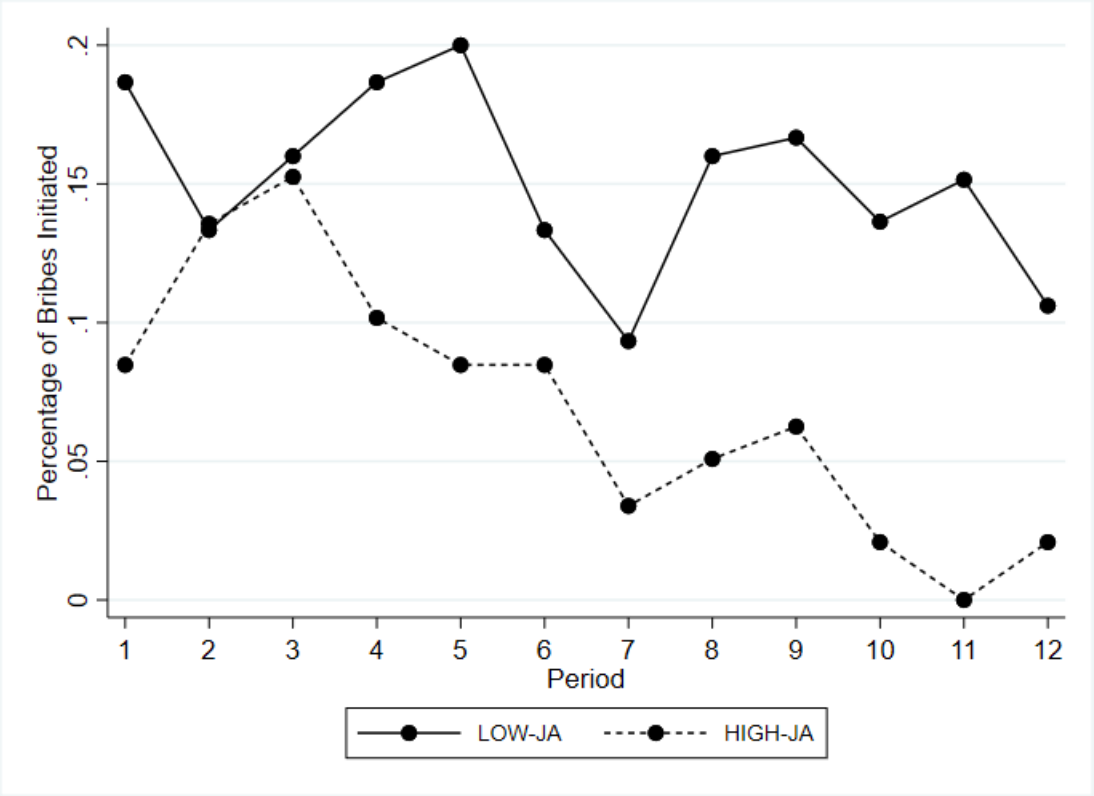


Figure 2.15: Bribe Incidence over time between treatments with low and high judicial accuracy.

were false in the first round, increasing to 15 percent in the last round.

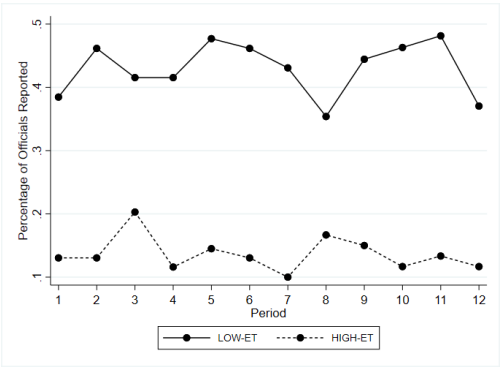
2.5 Discussion

The results from our experiment show that the underlying judicial environment influences the impact of financial rewards on bribe incidence and whistle-blowing. Overall, there is less bribery in our experiment, where only 14.2 percent of officials asked for a bribe across the four treatments. One explanation for such a low fraction of bribes is the theory of decision points (Soman et al., 2010). Instead of directly choosing an amount from 0 to 200, participants in our study first decide if they should demand a bribe or not. If they decide to demand a bribe, they must specify the bribe amount (from 1 to 200). Another reason is less exposure to corruption among our sample of subjects. In the post-experiment questionnaire, we asked subjects whether they had ever been asked to pay a bribe. Only 18.75 percent of subjects responded in the affirmative. Consistent with our hypothesis (H1), fewer officials ask for a bribe when the underlying judicial system is highly accurate. The average amount of bribes asked is significantly less when judicial accuracy is high.

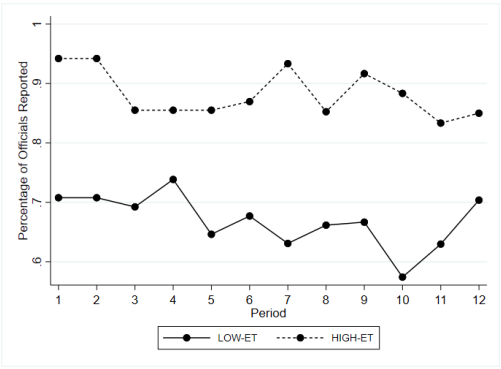
If the evidence threshold is increased, then fewer officials are reported. Our hypothesis (H2) was that there would be less overall reporting (true as well as false) under high evidence threshold. We find that there is a significant decrease in reporting by citizens under a higher evidence threshold. An interesting picture emerges when we distinguish between true and false reporting. We find that the fall in reporting is only significant in the situation where the official had not asked for a bribe. So, a higher evidence threshold only leads to a fall in false reporting. We had also hypothesized that whistle-blowing would be optimal when high evidence threshold is accompanied by an improvement in judicial accuracy (treatment 1). We find that the percentage of honest reports in treatment 1 is significantly higher than any other treatment.

It is in the interest of policymakers to consider the judicial parameters like evidence threshold and quality of decision making when drafting a whistle-blowing reward scheme. If the evidence threshold is kept low, introducing a whistle-blowing reward scheme will result in over-reporting. This will overburden the judiciary in less developed countries with a stockpile of unresolved cases. There should be much more emphasis on incorporating judicial reforms into whistle-blowing reward programs to overcome this problem.

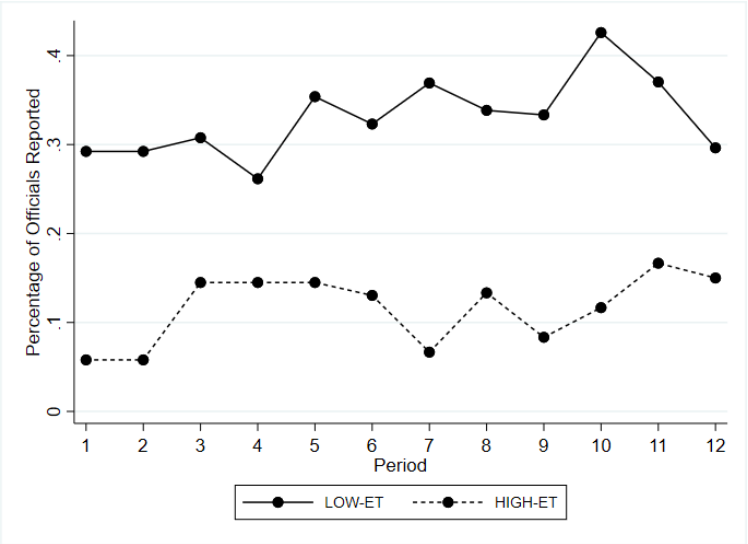
The evidence from our study can be extended in at least two dimensions. First, it would be worthwhile to see the effects of our treatments using subjects who have greater exposure to corruption. The study



(a) Citizen Reports - All



(b) Citizen Reports - Honest



(c) Citizen Reports - False

Figure 2.16: **(a)** Trend of percentage of reports by citizens whether honest or false. **(b)** Trend of percentage of honest reports by citizens. **(c)** Trend of percentage of false reports by citizens.

can be replicated in countries that perform poorly on the Corruption Perception Index of the Transparency International or Control of Corruption Index of the World Bank.

Another possible extension is to study the effect of social norms in influencing the decisions of officials and citizens. Norm information can either be descriptive (what most people do) or injunctive (what most people consider appropriate) in nature. It would be especially interesting to see how people make decisions when the two pieces of information conflict. Previous work has shown that descriptive information is more powerful in influencing behavior than injunctive information (Bicchieri and Xiao, 2009; Bicchieri and Dimant, 2019, Bicchieri et al., 2021).

In our experiment, we symmetrically increase or decrease type-1 and type-2 errors. It can be argued that a decrease in false positives can only be achieved at the cost of an increase in false negatives. Under such a trade-off, treatments can be designed with asymmetric changes in the two errors.

Bibliography

Abbink, K., Irlenbusch, B. and Renner, E., 2002. An experimental bribery game. Journal of Law, economics, and organization, 18(2), pp.428-454.

Abbink, K. and Hennig-Schmidt, H., 2002. Neutral versus loaded instructions in a bribery experiment (No. 23/2002). Bonn Econ Discussion Papers.

Abbink, K., 2006. 14 Laboratory experiments on corruption. International handbook on the economics of corruption, p.418.

Abbink, K. and Wu, K., 2017. Reward self-reporting to deter corruption: An experiment on mitigating collusive bribery. Journal of Economic Behavior Organization, 133, pp.256-272.

Abbink, K., Dasgupta, U., Gangadharan, L. and Jain, T., 2014. Letting the briber go free: an experiment on mitigating harassment bribes. Journal of Public Economics, 111, pp.17-28.

Bardhan, P., 1997. Corruption and development: a review of issues. Journal of economic literature, 35(3), pp.1320-1346.

Basu, K., 2011. *Why, for a Class of Bribes, the Act of Giving a Bribe should be Treated as Legal.*

Breuer, L., 2013. Tax compliance and whistleblowing–The role of incentives. The Bonn Journal of Economics, 2(2), pp.7-44.

Buckenmaier, J., Dimant, E. and Mittone, L., 2018. Effects of institutional history and leniency on collusive corruption and tax evasion. Journal of Economic Behavior Organization.

Butler, J.V., Serra, D. and Spagnolo, G., 2020. Motivating whistleblowers. Management Science, 66(2), pp.605-621.

Carpenter, J., Robbett, A. and Akbar, P.A., 2018. Profit sharing and peer reporting. Management Science, 64(9), pp.4261-4276.

DuFwenberg, M. and Spagnolo, G., 2011. Legalizing bribes. Eller College of Management Working Paper, (11-09).

Dyck, A., Morse, A. and Zingales, L., 2010. Who blows the whistle on corporate fraud?. The journal of finance, 65(6), pp.2213-2253.

Erkal, N., Gangadharan, L. and Nikiforakis, N., 2011. Relative earnings and giving in a real-effort experiment. American Economic Review, 101(7), pp.3330-48.

Fischbacher, U., 2007. *z-Tree: Zurich toolbox for ready-made economic experiments.* Experimental economics, 10(2), pp.171-178.

Frank, B. and Schulze, G.G., 2000. Does economics make citizens corrupt?. Journal of economic behavior organization, 43(1), pp.101-113.

Frank, B. and Schulze, G.G., 2003. Deterrence versus intrinsic motivation: Experimental evidence on the determinants of corruptibility. Economics of governance, 4(2), pp.143-160.

Graf Lambsdorff, J., 2005. Consequences and causes of corruption: What do we know from a cross-section of countries? (Vol. 34, No. 05). Passauer Diskussionspapiere-Volkswirtschaftliche Reihe.

Holt, C.A. and Laury, S.K., 2002. Risk aversion and incentive effects. American economic review, 92(5), pp.1644-1655.

Malik, S., Mihm, B. and Timme, F., 2018. An experimental analysis of tax avoidance policies. International Tax and Public Finance, 25(1), pp.200-239.

Reuben, E. and Stephenson, M., 2013. Nobody likes a rat: On the willingness to report lies and the consequences thereof. Journal of Economic Behavior Organization, 93, pp.384-391.

Soman, D., Xu, J. and Cheema, A., 2010. *Decision points: A theory emerges.* Rotman Magazine, pp.64-68.

Giancarlo, Spagnolo, 2008. Leniency and whistleblowers in antitrust. Handbook of Antitrust Economics. The MIT Press, Cambridge, Massachusetts.

González, L., Güth, W. and Levati, M.V., 2002. Speeding up bureaucrats by greasing them: an experimental study. Max-Planck-Inst. for Research into Economic Systems, Strategic Interaction Group.

Svensson, J., 2005. Eight questions about corruption. Journal of economic perspectives, 19(3), pp.19-42.

Treisman, D., 2007. What have we learned about the causes of corruption from ten years of cross-national empirical research?. Annu. Rev. Polit. Sci., 10, pp.211-244.

Chapter 3

Diversity in Committees

Niall Hughes, Zia-Ul-Hassan Khan & Friederike Mengel[1]

Abstract

We investigate the potential of diversity to influence committee decision-making. We present a model in which committee members receive private information on the state of the world, deliberate and vote. Committee members belong to one of two groups which may differ along two dimensions: (a) their preferences and (b) their information structures. We think of diversity as more nuanced than ethnic or gender-based diversity an. It comes either in the form of informational diversity or preference diversity. Diverse committees outperform other committees if information structures are correlated within groups and preferences are aligned. With heterogeneity in preferences, homogeneous committees do better than some, but not all diverse committees. We test the model's key predictions in a laboratory experiment. The model's predictions are largely borne out. Diverse committees outperform homogeneous committees when information structures are correlated, and preferences are aligned, but they do not do worse even when preferences are misaligned. The reason is that participants reveal more information than theory predicts and update imperfectly.

Keywords: Committee Decision-Making, Experiments, Diversity, Information Aggregation, Deliberation, Voting, Strategic Communication.

JEL Classification Number: C92, D71, D72.

[1]We thank Andrea Albertazzi for excellent research assistance. Friederike Mengel thanks the European Research Council for financial support (ERC Starting Grant DYNNET 801517).

3.1 Introduction

The value of diversity in organisations and society is a recurring and controversial topic in public debate. The widely cited "business case" for diversity endorsed by many business leaders and politicians is that increased diversity in teams and organisations will lead to increased profits.[2] Typical arguments include (a) diverse organisations are more in tune with the preferences of their diverse customers (b) diverse teams have a wider set of skills than homogeneous teams, and (c) those with different backgrounds have different perspectives and access to different information (Thomas and Ely, 1996). Diversity is also seen as a way to tackle inherent bias and groupthink (Hamilton et al., 2012) and to lead to better decisions (Hong and Page, 2001). As a result of this, many organisations have championed the cause of diversity. In a 2020 survey of 693 US public company directors undertaken by PwC, 94% of directors agreed that diversity brought unique perspectives, 83% claimed it enhanced board performance, 85% considering that it improved relations with investors and 72% claimed it improved the performance of the company.

While the business case for diversity seems intuitive and is widely endorsed, the academic literature paints a much more mixed picture. In some models, diversity leads to increased discrimination (Becker, 1957), increased free-riding (Kandel and Lazear, 1992), and less information aggregation (Coughlan, 2000). In other models, diversity provides access to more skills (Hong and Page, 2001) and more information acquisition (Che and Kartik, 2009). In yet another set of models, diversity is either a boon or a drawback depending on conditions (Lazear, 1999; Prat, 2002; Kets and Sandroni, 2021; Jensen, 2021). The empirical literature is no less ambiguous. Diverse committees and company boards have been associated with better investments and financial outcomes in some studies (Gompers and Kovvali, 2018; Gomez and Bernet, 2019), but other researchers have not found evidence that increased diversity on company boards would generate higher profits for the firm (Ahern and Dittmar, 2012; Ely and Thomas, 2020). Diversity has also been shown to be detrimental if it is associated with preference heterogeneity (Alesina and Ferrara, 2005) or when groups have group dependent social preferences (Chen and Chen, 2011). The literature has also found mixed effects when it comes to assessing how the diversity of committees affects outcomes for minorities (Bagues et al., 2017; Sommers, 2006; Bagues and Esteve-Volart, 2010; Malmstrom et al., 2018; D'Acunto et al., 2021). When it comes to team productivity, the evidence is similarly mixed. While diverse teams have shown better performance in some studies (Apesteguia et al., 2012; Hoogendoorn et al., 2013), team diversity has been found to hurt productivity in others (Hjort, 2014; Lyons, 2017). In a field experiment in Kenya, Marx et al. (2021) find that diversity within teams reduced productivity while diversity between teammates and their supervisor increased productivity. Finally, (Weidmann and Deming, 2021) find that there is no correlation between team performance and ethnic or gender diversity.[3]

What lessons should we draw from this mixed bag of evidence? We believe there are two key insights we can draw. First, diversity is a multifaceted object - while a homogeneous group is alike on all dimensions, a diverse group may differ on one or several dimensions. Therefore, it is essential to specify what dimensions of diversity we are investigating. In this paper, we think of diversity in terms of information or preferences. Second, diversity will have different impacts in different conditions, so these details matter.[4] With this in mind, we analyse the impact of diversity in committee decision-making setting where members receive private information on the state of the world, deliberate, and vote. We focus on this setting for several reasons. First, many real-world committees fit that description - juries, hiring committees, monetary policy committees. Indeed, there is a vast literature on information aggregation in committees going back to Condorcet's (1785) Jury Theorem. Second, there has been much debate and growing movement towards increased diversity in these committees (D'Acunto et al., 2021). Third, if members were fully informed in these committees, all

[2] See for example: Boston Consulting Groups' 2018 report "How Diverse Leadership Teams Boost Innovation", McKinsey's 2020 report "Diversity Wins", and the Institute for Business Ethics 2020 report "The Ethics of Diversity"

[3] See (Azmat, 2014) for a review of the literature on diversity in teams.

[4] See Ely and Thomas (2020) for a discussion of how simplistic models of diversity and unrealistic expectations have set back its success

would agree on the correct decision. That is, adding diversity does not mechanically create unresolvable conflict. While there is ample evidence in the economics literature on how diversity in preferences affects committee performance (Coughlan, 2000); Austen-Smith and Feddersen, 2006) - there is very little work on the role played by informational diversity and how it interacts with preference diversity. In this paper, we seek to address this gap by jointly studying the role of preference diversity and informational diversity in committee decision-making. Studying them together will help us identify when diverse committees should be expected to outperform other committees and when this is not the case.

We study a group decision-making problem where members vote by majority rule to implement one of two options where the optimal choice depends on an unknown binary state of the world. Before voting, each member gets a partially informative signal about the state of the world, and the committee deliberates. Each committee member belongs to one of two groups, which is publicly observable. We may think of these groups as male/female, white/non-white, or any other two groups. Committee members from different groups potentially differ along two dimensions. First, those in different groups may differ in their utility gain for a correct decision in each state of the world. This is the standard heterogeneity of preferences already present in the literature (Coughlan, 2000); Austen-Smith and Feddersen, 2006; Quement and Marcin, 2019). Second, members in different groups obtain their signals about the state of the world from distinct, though equally informative, information structures. Moreover, signals are potentially correlated within groups. As far as we are aware, we are the first to employ such informational diversity in a model of committee decision-making.[5]

We can think of several reasons why the information of those within the same group may be correlated. First, it may be that those from the same group have overlapping networks and so are more likely to encounter the same information (Chandrasekhar et al., 2020; Grimm and Mengel, 2020). Second, it may be that those from the same group follow similar search strategies, e.g. they read the same newspapers (Levy and Razin, 2015b; Pogorelskiy and Shum, 2019). Third, it may be that those from the same group interpret a piece of evidence one way while another group interprets it differently, due, for example, to different life experiences as part of that group (Kets and Sandroni, 2021).[6] We believe that both forms of diversity - in preferences and in information - are important in real-world committees. We vary whether preferences in the two groups are aligned or misaligned, while also varying whether information structures are correlated or not within groups. This allows us to isolate the two independent effects of each dimension of diversity and the joint effect.

The model provides several testable predictions. It predicts that with aligned preferences and correlated information, diverse committees outperform homogeneous committees in terms of both decision accuracy and welfare. This is because a diverse committee has access to more independent sources of information. The model further predicts that when information is independent but the preferences of the two groups are not aligned - so that one group is moderate while the other is extremist - then moderate homogeneous committees outperform any kind of diverse committee. This is driven by the fact that fully truthful communication is not achievable in a committee with diverse preferences. As heterogeneity in information favours diverse committees while heterogeneity in preferences favours (moderate) homogeneous committees, it was unclear ex ante whether having differences on both dimensions would favour one committee type or the other. Out final key prediction of our model is that, with correlated information and misaligned preferences, diverse committees where extremists are in the majority generate higher welfare for moderates than any other committee. Not only does the choice of a homogeneous or diverse committee matter for welfare - but the type of diversity matters. Moderates do best when decisions are made by a diverse committee with extremists in the majority. The reason for this result is that all members of extremist-majority committees can truthfully

[5]Bardhi and Bobkova (2021) make a related distinction between informational diversity and demographic diversity, but their paper focuses on information acquisition in an advisory board where there is no private information and no voting.

[6]There is an increasing body of empirical evidence summarised in (Kets and Sandroni, 2021) that shows the differences in cognitive processes exist across groups. How a person responds to a contextual cue or what they pay attention to is often a function of their group identity.

share their signals while those in moderate-majority committees cannot. Truthful communication also occurs in homogeneous committees, but their information sources are more correlated and thus lower quality than those of the extremist-majority diverse committees.

These predictions are based on an equilibrium refinement, as well as assumptions of flawless belief updating of committee members and fully strategic behaviour in deliberation. To deal with the multiplicity of equilibria, we follow the cheap talk literature in focusing on the most informative equilibrium. In subsection 3.4.1, we discuss reasons why this equilibrium selection and our assumption of fully strategic behaviour might fail to predict behaviour well in all treatments. We then take our hypotheses to the laboratory in what is the first experimental test of theoretical predictions on committee deliberation and decision-making with diverse information structures.

We find strong evidence in favour of our hypothesis regarding the difference between homogeneous and diverse committees when preferences are aligned but information structures are correlated. The resulting differences in committee error rates confirm our first key hypothesis - there are indeed more correct group decisions in diverse committees. As predicted, the difference is driven by diverse committees making significantly more correct decisions in one state of the world. We further find that the data supports our second theoretical result - with misaligned preferences but independent information, moderate-majority committees outperform all other committees. Finally, we do not find support for our last hypothesis, that extremist-majority diverse committees outperform other committee types. In our experiment, moderate-majority diverse committees perform best. This departure from the theory is because subjects truthfully reveal their signals in cases when it is not individually optimal to do so. This lack of strategic lying, means that there is little difference between the various diverse committees. Taken together our theory and experimental results suggest there are great benefits to diversity in committees. The only case in which homogeneous committees perform better is when preferences are misaligned and there is no correlated information within groups.

Our research contributes to a large literature on deliberation and information aggregation in committee decision-making.[7] This literature has focused on how information aggregation and truth-telling vary with the voting rule, the communication protocol, and the degree of preference diversity in the committee. Coughlan (2000) shows that if preferences are homogenous, there is always an equilibrium where members share their signals truthfully and then vote sincerely. Instead, in diverse committees, truthful communication is stifled, leading to inferior decision making vis a vis homogeneous committees. In a lab experiment with free-form deliberation, Goeree and Yariv (2011) show that homogenous committees indeed outperform diverse committees - though the difference is not as stark as in theory. Those in diverse committees share their signals with much higher frequency than theory predicts. This excessive truth-telling is in line with that found in other experiments on cheap talk communication (Cai and Wang, 2006; Battaglini and Makarov, 2014; Fehrler and Hughes, 2018). Quement and Marcin (2019) explore this result further and find that excessive truth-telling is driven by cognitively heterogenous agents rather than social preferences or lying aversion. In all of this literature, diversity is only in terms of preferences - all members still receive conditionally independent signals on the state of the world. Our key contribution to this literature is that we model diversity of information as well as of diversity of preferences.

The applied and experimental literature has mostly focused on whether diversity in a committee (e.g. a higher share of women) benefits female or minority candidates in hiring or promotion decisions (Bagues et al., 2017; Sommers, 2006; Bagues and Esteve-Volart, 2010; De Paola and Scoppa, 2015; Deschamps, 2020), in performance evaluations (Mengel, 2021) or in the allocation of venture capital (Malmstrom et al., 2018). The literature has also considered the diversity of company boards and executive committees and how it relates to indicators of firm success with mixed results (Gompers and Kovvali, 2018; Gomez and Bernet, 2019; Ahern and Dittmar, 2012; Ely and Thomas, 2020). Last, researchers have convincingly demonstrated that central bankers' individual characteristics shape voting behaviour on monetary policy

[7] See Gerling et al. (2005) and Austen-Smith and Feddersen (2009) for reviews of this literature.

committees (Goehlmann and Vaubel, 2007; Malmendier et al., 2021). This literature has by and large not been able to study and distinguish mechanisms by which diversity leads or does not lead to desired outcomes. This is a big advantage of our experimental data collection. We are able to model and cleanly separate the effect of two key dimensions of diversity: diversity of preferences and diversity of information.

In Section 3.2, we present the model. Section 3.3 provides the main results of the model. We proceed with the experimental design in Section 3.4 before discussing the aggregate and individual level results in Section 3.5. In Section 3.6, we conclude.

3.2 Model

A committee of three members must make a decision $D \in \{P(urple), Y(ellow)\}$ by simple majority rule. The objective is to match the decision to the unknown state of the world $S \in \{P, Y\}$ where states are equally likely ex-ante. Each member either belongs to group 1 or group 2, and this is publicly observable. Committee members from different groups differ potentially along two dimensions. First, members in different groups obtain their signals about the state of the world from distinct information structures. Nature draws a signal $s_i \in \{p, y\}$ for each member i about the true state. In addition, Nature draws a signal $s_g \in \{p, y\}$ for each group $g \in \{1, 2\}$. These signals are equally informative: a signal is correct with probability $\sigma \in (0.5, 1)$.[8] Committee member i from group g receives s_i with probability q and the group signal s_g with probability $1 - q$. Let $\tilde{s}_{i,g}$ denote the signal a member receives, where $E(\tilde{s}_{i,g}) = qs_i + (1 - q)s_g$. A member does not know whether they have received their individual signal or the group signal. Whether a member receives an individual or group signal is independent of the type of signal any other member receives. Second, those in different groups may differ in their utility gain for a correct decision in each state of the world. A member from group g gains x_g when $D = S = P$, gains $(1 - x_g)$ when $D = S = Y$, and nothing when $D \neq S$ where $x_g \in [0, 1]$.

The timing of the game is as follows. Each committee member receives a signal $\tilde{s}_{i,g}$ and thus learns their type $\{p_1, p_2, y_1, y_2\}$, where e.g. p_1 refers to a member from group 1 that receives a signal p. Then, in a simple straw poll, each member simultaneously announces a message $m_i \in \{m_p, m_y\}$, i.e. raises their hand in favor of P or Y. Finally, each member casts a vote $v_i \in \{v_P, v_Y\}$, the group decision is taken, the state is revealed and payoffs are realised.

A committee member's strategy consists of a messaging strategy and a voting strategy. A messaging strategy $\boldsymbol{m^i}$ is a mapping from type into a probability distribution over messages $\{m_p, m_y\}$. A voting strategy $\boldsymbol{v^i}$ is a mapping from type and messages exchanged in the straw poll into a probability distribution over votes $\{v_P, v_Y\}$. We study symmetric perfect Bayesian equilibria. As is standard in voting games, we restrict attention to strategies that are not weakly dominated. As talk is cheap, there will be many equilibria in each of the three settings. We follow the cheap talk literature (Crawford and Sobel, 1982; Ottaviani and Sørensen, 2001; Chen et al., 2008; Fehrler and Hughes, 2018) in focusing on the most informative equilibrium, i.e. in each setting we look for equilibria where the maximal amount of information is shared in the communication stage. We also ignore equilibria with inverted language.

We compare the equilibrium performance of homogenous committees and diverse committees in four settings: (i) when group preferences are aligned ($x_1 = x_2$) and the information structure is independent ($q = 1$), (ii) when group preferences are aligned ($x_1 = x_2$) and the information structure is correlated within groups ($q < 1$), (iii) when group preferences are misaligned ($x_1 < x_2$) and the information structure is independent ($q = 1$), (iv) when group preferences are misaligned ($x_1 < x_2$) and the information structure is correlated within groups ($q < 1$). [9]

[8] That is, $Pr(s_i = p|S = P) = Pr(s_g = p|S = P) = Pr(s_i = y|S = Y) = Pr(s_g = y|S = Y) = \sigma \in (0.5, 1)$

[9] If $q < 1$, individuals know the information structure but not whether they have received an individual or group signal. This is distinct from correlation neglect, where individuals assume $q = 1$ even though $q < 1$.

3.3 Model Results

In what follows, we make the following assumption:

Assumption 1. $0.5 < Pr(S = Y|yyp, hcorr) < x_1 < \sigma$

The term $Pr(S = Y|yyp, hcorr)$ is the posterior of a homogenous committee that observes two yellow signals and one purple signal under correlated information.[10] The assumption means that individuals from group 1 have a weak preference for P but will be swayed by a single signal in favour of Y, if it is known to be independent but not if it may be correlated. We study aligned and misaligned preferences in turn.

3.3.1 Aligned Preferences

With $x_1 = x_2$, it is always an equilibrium for players to truthfully reveal signals in the straw poll.[11] Committee behaviour depends on its composition and the information structure. With an independent information structure, there is no difference between homogenous and diverse committees in terms of the information or preferences. They have identical welfare, decision accuracy and committee member behaviour. In the most informative equilibrium, each member truthfully reveals their signal and the committee implements the decision with a majority of signals.

Proposition 1. *With an independent information structure and aligned preferences, there is no difference between homogenous committees and diverse committees neither in terms of decision accuracy nor in terms of welfare.*

The proofs of all propositions are in Appendix C.3. The proposition is trivial but serves as a useful benchmark with which to compare our other settings. Under a correlated information structure, diverse committees obtain better quality information than homogenous committees as there is less chance of correlated signals within a diverse committee. In a homogenous committee, all three members may have correlated signals; or any pair of member signals may be correlated. In a diverse committee, at most the signals of those in the majority group can be correlated. We can now state our main result under aligned preferences and correlated information.

Proposition 2. *With a correlated information structure and aligned preferences, diverse committees have higher overall decision accuracy and generate higher welfare than homogenous committees.*

The difference in decision accuracy and welfare stems from (i) more independent information in diverse committees, which leads to (ii) different choices in these respective committees for the same profile of signals. The reduction in information quality for a homogenous committee when moving from independent to correlated information means they no longer implement Y after a bare majority of yellow signals. Diverse committees face a smaller drop off in information quality - they still implement P after a bare majority of signals as long as the two y signals come from members of different groups. If both y signals come from the same group, the diverse committee selects P. Though diverse committees outperform homogenous committees in this correlated information setting, decision accuracy and welfare is higher yet again under the independent information setting of Proposition 1.

3.3.2 Misaligned Preferences

In this section, we study what happens when we no longer have $x_1 = x_2$. Though there are many ways we could deviate from aligned preferences to misaligned preferences, we examine the case where $x_1 < x_2$, assuming the following

[10] As we show in the proof of Proposition 1, $Pr(S = Y|yyp, hcorr) < \sigma$ is true for any $q < 1$.

[11] Only a truth-telling equilibrium maximises the probability of the group decision being correct.

Assumption 2. $\sigma < x_2 < Pr(S = Y|yyy, hcorr)$

The preferences of those in group 1 remain unchanged, while those in group 2 are now more in favour of implementing P for any signal realisation. The assumption means that one independent signal in favour of Y is not enough to overcome the preference for P, but if all signals unanimously favour Y, even if correlated, those in group 2 will implement Y. Though both groups favour P ex-ante, group 1 are "moderates" while group 2 are "extremists". From now on we will use these terms interchangeably with group names. Assumption 2 ensures that the divergence in preference between the two groups is large enough that behaviour is not identical for both groups after all signal realisations. We kept the preferences of those in group 1 fixed to allow a close comparison to the aligned preference case, both theoretical and in the experiment. An alternative way to generate misaligned preferences would be to set $x_2 < 0.5 < x_1$ so that groups have biases in the opposite direction. We examine this case in Appendix C.2 but do not study it experimentally.

Analysing welfare is not straightforward when preferences are misaligned. The composition of a committee affects not only the quality of signals and truthfulness but also it also changes the preferences of committee members. A good outcome for a moderate may not be good for an extremist, so the quality of decisions will depend on the group composition. In order to make clean comparisons across treatments, we focus on the welfare of a moderate from any particular committee composition and information structure. One way to think of this is that a moderate principal must decide whether and how much (extremist) diversity to add to his committee. This focus on moderate welfare also allows us to compare welfare in aligned and misaligned settings.

The following proposition shows that moderate homogenous committees outperform diverse committees when information is independent and preferences are misaligned.

Proposition 3. *With an independent information structure and misaligned preferences, moderate homogenous committees have higher overall decision accuracy and generate higher welfare for moderate committee members than any other committee.*

Clearly, moderate homogenous committees behave as they do in the case of aligned preferences. The same would be true for moderate-majority diverse committees if all members truthfully revealed their signals in the straw poll. However, the introduction of misaligned preferences within the committee means truth-telling is not incentive compatible for the extremist member. He always has an incentive to report a purple signal. As they are in the majority, the moderates dictate the group decision. They do so based solely on their two signals, ignoring the message of the extremist. In the most informative equilibrium, the moderates tell the truth while the extremist babbles; the committee implements Y only if both moderates have y signals. In an extremist-majority committee truth-telling in the straw-poll is not incentive compatible for the moderate member. He always has an incentive to report a y signal. As they are in the majority, the extremists dictate the group decision. They do so based solely on their two signals, ignoring the message of the moderate. In the most informative equilibrium, the extremists tell the truth while the moderate babbles; the committee implements Y only if both extremists have y signals. Thus, the minority member babbles in both diverse committees. Furthermore, the committee decision after each signal realisation is the same in each diverse committee, meaning decision accuracy and moderates' welfare is the same. Next, we combine our two forms of heterogeneity across groups - correlated information structures and misaligned preferences.

Proposition 4. *With a correlated information structure and misaligned preferences, diverse committees have higher overall decision accuracy than homogenous committees. Extremist-majority diverse committees generate the highest welfare for moderate committee members followed by moderate homogenous committees, then moderate-majority diverse committees.*

Not only does the choice of a homogenous or diverse committee matter for welfare - but the *type* of diversity matters. Moderates do best when decisions are made by a diverse committee with extremists in the majority.

The utility of a moderate type is monotonic in the number of moderates on the committee when information is independent. With correlated information, there is a benefit to diversity as the committee gets more independent information. One might conjecture that welfare for moderates would now be highest under a moderate-majority diverse committee. This is indeed the case if all committee members truthfully reveal their signals. However, once again, truth-telling is not an equilibrium in moderate-majority diverse committees. The benefit that diversity brings in terms of more information is outweighed by the cost of a reduction in truthfulness. Welfare is higher for moderates under moderate homogenous committees than moderate-majority diverse committees.

One might expect the same problem to arise with extremist-majority diverse committees. After all, truth-telling was not an equilibrium in these committees under independent information. As it turns out, truth-telling is an equilibrium here for extremist-majority diverse committees. Under independent information, when there is a bare majority of y signals, and they come from the same group, the moderate has an incentive to lie by reporting a y signal in order to induce a group decision of Y. With the same realisation of signals under correlated information, the preferences of the extremists and moderate are aligned - both want to implement P. Therefore, there is no incentive for the moderate to deviate from truth-telling. Extremist-majority diverse committees outperform all others here because they obtain the informational benefit of diversity without the drawback of reduced truthfulness.[12] This is not the case if such a committee has independent information. This means that welfare is higher with a correlated information structure than an independent one, conditional on having an extremist-majority diverse committee.

3.3.3 Choosing committees under uncertainty

When deciding on the composition of a committee, a principal may not know in which environment it will operate. If it is a standing committee, it will have many different decisions to make over time. Perhaps information is correlated within groups for some decisions but not for others. Similarly, member preferences may be aligned for some tasks but not for others. Should a principal tend towards homogenous or diverse committees if he is unsure of the environment? From the propositions above, we can give the following advice.

If preferences are aligned but the information structure may be independent or correlated, a diverse committee will outperform a homogenous committee. We know from Proposition 1 that there is no difference when information is independent and from Proposition 2 that diverse committees perform better when information is correlated.

Suppose the information structure is known to be independent but the principal doesn't know whether preferences will be aligned or not. In that case, the optimal choice for a moderate principal is to form a moderate homogenous committee. We know from Proposition 1 that there is no difference when preferences are aligned and from Proposition 3 that moderate homogenous committees generate higher utility for moderates when preferences are misaligned.

If the information structure is known to be correlated but the principal doesn't know whether preferences will be aligned or not, the optimal committee is an extremist-majority diverse committee. This follows from the fact that by Proposition 2 diverse committees outperform homogenous committees when preferences are aligned, and the fact that by Proposition 4 extremist-majority diverse committees outperform all other committee types when preferences are misaligned.

Finally, if preferences are sure to be misaligned but information may be independent or correlated, the optimal committee type will depend on the probability that information is independent. We know from Proposition 3 that moderate homogenous committees general higher utility for moderates when information is independent. In contrast, from Proposition 4 we know that extremist-majority diverse committees perform

[12] Welfare is also higher for extremists in extremist-majority diverse committees than in extremist homogenous committees. This is because both committees truthfully reveal their signals, but the diverse committee has better quality signals.

best when information is correlated. If the ex-ante probability of independent information is above a certain threshold, homogenous committees are better. Otherwise, the principal should choose extremist-majority diverse committees. The threshold is a function of σ, q and x_1. For our chosen parameters in the lab, the threshold is approximately 0.36; so if the probability of independent information is lower than 0.36, extremist-majority committees are the best choice for a moderate principal.

3.4 Experiment

To test the predictions of our model, we ran a laboratory experiment in a 2x2 design, varying whether the information structure was independent or correlated, and varying whether preferences were aligned or misaligned. Before describing the experimental setup in detail, we first discuss possible reasons the model's predictions may fail.

3.4.1 Why model predictions may fail in the lab

Truth-telling, Strategic Behaviour & Open Chat In our model, as there are multiple equilibria, we focused on the most informative equilibrium in each setting, following the cheap talk literature (Crawford and Sobel, 1982; Ottaviani and Sørensen, 2001; Fehrler and Hughes, 2018). One issue that may arise is that committee members play a less truthful equilibrium. If diverse committees play less truthful equilibria than homogenous committees, then our results on the benefits of diversity may no longer hold. We know that equilibrium truthfulness is, in fact, low in diverse committees when preferences are misaligned, but it may also be the case that these committees share less information than homogenous committees even when preferences across groups are aligned.

A related issue is that of excess truth-telling. Our predictions on homogenous committees outperforming diverse committees in Proposition 3 rely on the minority member in each diverse committee not truthfully revealing their signal. Similarly, the difference between the two types of diverse committees in Proposition 4 depends on differences in equilibrium truth-telling. There is a large body of experimental research documenting the fact that subjects in the lab typically have a preference for truth-telling.[13] This is particularly so in experiments involving free-form communication such as ours (see Goeree and Yariv (2011). In our theory, the benefit of homogenous committees over diverse committees is that the former can always communicate truthfully. If de facto all committees are truthful - and every member knows they are truthful, then the principal faces no tradeoff in choosing between diverse and homogenous committees - diversity is always weakly better.

Bayesian Updating & Correlation Neglect There is now widespread evidence that individuals in the lab struggle to update beliefs according to Bayes' rule.[14] In our setting, updating is quite simple when information is independent, but it becomes much more difficult when signals correlated within groups. Our results in Proposition 2 and Proposition 4 rely on players understanding how the correlated information structure affects player incentives. A player needs to understand that his optimal choice may change when moving from independent to correlated information and that this change depends on the composition of the committee. For example, under aligned preferences, if a committee has two y signals and an independent information structure, they prefer to implement Y, but for the same realisation of signals in a homogenous committee with correlated information, the optimal decision is P. Further complicating things, a diverse committee will prefer P or Y depending on exactly which members received the two y signals. If subjects struggle with Bayesian updating in simple environments with independent signals, we may expect them to perform even worse when information is correlated. In particular, subjects in the lab may neglect the

[13] See for example Cai and Wang (2006), Goeree and Yariv (2011), Battaglini and Makarov (2014), and Fehrler and Hughes (2018).

[14] See for example Charness and Levin (2005), Charness et al., (2007), and Levin et al., (2016).

		N_C	% correct committee decisions					
			$S = P$		$S = Y$		Overall	
			Theory	Exp	Theory	Exp	Theory	Exp
AL-IND	DIVERSE	685	78.4	75.9	78.4	66.4	78.4	70.6
	NOT	215	78.4	71.5	78.4	66.6	78.4	68.8
AL-CORR	DIVERSE	683	86.0	75.7	60.0	59.1	73.2	68.5
	NOT	217	91.5	74.7	44.0	49.0	67.9	62.2
MIS-IND	DIV-MOD	267	91.0	81.4	49.0	65.9	70.0	69.5
	DIV-EXT	302	91.0	74.7	49.0	52.5	70.0	69.5
	NOT-MOD	161	78.4	80.5	78.4	72.3	78.4	75.7
	NOT-EXT	146	97.3	80.7	34.3	52.9	65.8	67.8
MIS-CORR	DIV-MOD	290	86.0	80.4	54.0	67.1	70.0	73.1
	DIV-EXT	341	96.0	79.6	42.0	63.7	70.0	71.5
	NOT-MOD	121	91.5	80.4	44.0	64.0	67.9	68.0
	NOT-EXT	136	91.5	76.9	44.0	56.3	67.9	68.0

Table 3.1: Basic Summary Statistics Experiment. Number of committees (N_C) as well as the percentage of correct committee decisions depending on the state and the type of committee.

correlated structure of information entirely, and instead, behave as if signals were all independent. Several recent papers in economics have examined the prevalence of correlation neglect and its effect on aggregate outcomes (Levy and Razin, 2015a; Enke and Zimmerann, 2018). However, there is little or no research on correlation neglect in committees. We examine this case of correlation neglect in Appendix C.1 and show that under correlation neglect diverse committees still outperform homogenous committees. Nonetheless, we cannot rule out a scenario where, in the lab, correlation neglect is more severe in diverse committees than homogenous committees: When preferences are aligned the optimal behaviour of diverse committees is more nuanced than homogenous committees; when preferences are misaligned, those in diverse committees need to work out their preference following each collection of messages, and whether these messages are actually truthful. Finally, for a principal choosing a committee type under uncertainty as in subsection 3.3.3, correlation neglect dampens the benefits of choosing a diverse committee over a homogeneous committee when preferences are known to be misaligned.

3.4.2 Experimental Design

We designed a lab experiment to test the basic predictions of the theory (Propositions 1-4) and to understand to which extent committees suffer from correlation neglect. Ultimately we are interested in when diversity in committees is better and when it is worse for information aggregation.

Table 3.1 shows the basic treatment structure of our experiment together with some summary statistics for each treatment. Treatment **AL-IND** implements the case where preferences are aligned and information is independent, treatment **AL-CORR** the case where it is correlated and treatments **MIS-IND** and **MIS-CORR** study the corresponding cases when preferences are misaligned. We used the following parameter values in the experiment.

Parameter Values We set $\sigma = 0.7$ and in the correlated treatments $q = 0.5$. (In the independent treatments, of course, $q = 1$. That is, the purple urn has 7 purple balls and 3 yellows ones, while the yellow urn has 7 yellow balls and 3 purple ones; and in correlated treatments each participant has a 50% of getting their group signal and a 50% chance of getting an independent individual signal. We set $x_1 = x_2 = 0.66$ in the aligned treatments and $x_1 = 0.66$, $x_2 = 0.75$ in the misaligned treatments. In the purple state the payoff for a correct decision is 10 GBP for a moderates but 12 GBP for an extremist. In the yellow state the payoff for a correct decision is 5 GBP for a moderates but 3 GBP for an extremist.

In each treatment participants were assigned a group (1 or 2, in the experiment labelled S or T) at

the beginning of the experiment. Participants stayed in the same group in the entire experiment. They then played the following voting game for 12 rounds. At the beginning of each round an urn was randomly selected by nature. Then each participant received a signal using the process (correlated or uncorrelated) explained above. Afterwards committees of three participants were randomly formed and participants were shown the group of each of the committee members. Hence committee composition changed each round. Committee members were then given 45 seconds to chat using an on-screen chat window. Last committee members simultaneously voted. Experimental Instructions can be found in Appendix C.4 and screenshots in Appendix C.5. At the end of the experiment one round was randomly selected for each participant and participants received the number of points earned in that round in GBP. In addition all participants received a 5 GBP show up fee and 3 GBP for answering the post-experimental questionnaire.

	AL-IND	**AL-CORR**	**MIS-IND**	**MIS-CORR**
Mean Age	23.48	22.69	25.79	22.68
Share female	0.66	0.65	0.58	0.63
Share student	0.81	0.81	0.79	0.81
Nationality				
UK	0.42	0.32	0.53	0.43
Europe (non-UK)	0.36	0.45	0.34	0.41
Other	0.22	0.23	0.13	0.16

Table 3.2: Sample Characteristics.

We received ethical approval from the social sciences ethical subcommittee at the University of Essex with number ETH1920-0015. The experiment was pre-registered at the AEA trial registry with number AEARCTR-0005402. As there are no prior comparable studies it was hard to determine an expected empirical effect size for the difference in correct committee decisions between diverse and non-diverse committees. We hence conducted two sessions of treatment **AL-CORR** with 54 participants to determine an expected effect size. In these two sessions we detected an effect size of 0.107 for the difference in correct committee decisions. We then conducted a power analysis using these effect sizes. A target sample of 216 participants per treatment across 8 clusters allows us to detect a committee level effect of 0.107 with 78.47% power (at the 5% level). As show up for experimental sessions can be unpredictable, we decided to proceed running sessions until the threshold of 216 participants has been reached in all treatments. This resulted in 225 participants in treatments **AL-IND** and **AL-CORR**, 219 participants in **MIS-IND** and 222 participants in **MIS-CORR** (see also Table 3.1). Sample characteristics for these participants are shown in Table 3.2.

3.5 Experimental Results

3.5.1 Results: Aligned Preferences

We first show the experimental results testing the content of Propositions 1 and 2. In a second step we dig deeper into the decision-making process and ask how well diverse and non-diverse committees solve the aggregation problem and to which extent they suffer from correlation neglect. We will study the question of truth telling by looking at our chat data.

The two rightmost columns of Table 3.1 show the percentage of correct committee decisions as predicted by standard theory (with bayesian decision-makers) and in the experiment. The table shows that in both types of committees the percentage of correct committee decisions is somewhat lower than predicted by theory ($p < 0.1$), albeit not by a huge amount. The table also shows that with independent information diverse and non-diverse committees perform very similarly, while with correlated information diverse committees seem to do better. Table 3.3 investigates this point more systematically using LPM regressions where we regress binary variable indicating that committee decision was correct on the dummy "DIVERSE" indicating whether the committee was diverse.

	Independent			*Correlated*		
	(1)	(2)	(3)	(4)	(5)	(6)
DIVERSE	0.018	0.018	0.015	0.070**	0.071**	0.075**
	(0.044)	(0.045)	(0.042)	(0.027)	(0.029)	(0.029)
Constant	0.653***	0.615***	0.660***	0.526***	0.441***	0.517***
	(0.031)	(0.058)	(0.054)	(0.022)	(0.058)	(0.103)
Observations	900	900	900	900	900	900
R-squared	0.015	0.017	0.033	0.022	0.031	0.059

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 3.3: **Percent of Correct committee decisions. Aligned Preferences**. LPM regression of binary variable indicating that committee decision was correct on dummy indicating whether committee was diverse. Columns (1)-(3) show independent information and Columns (4)-(6) correlated information. All regressions include session fixed effects. Columns (2) and (5) linearly control for period and columns (3) and (6) have period fixed effects.

Table 3.3 shows that diverse committees are on average 7-8 percentage points more likely to reach a correct decision. Given the baseline of 62% correct decisions in non-diverse committees this represents an $11-12\%$ increase in correct decisions when the committee is diverse. This is a substantial improvement in decision making quality. The effect size is in line with what we would have expected from theory which - given our parameters - predicts an $\approx 8\% improvement for the diverse committee with Bayesian decision makers and an \approx 14\% improvement with correlation neglect. Appendix Table C.6.1 shows the percentage of individually correct decisions. Also here diverse committees when information is correlated.$

With correlated information the percentage of correct committee decisions is expected to depend on the state. Specifically, we would expect correct decisions 86 (91) percent of the time in the purple state and 60 (44) percent of the time in the yellow state with the diverse (non-diverse) committee. Hence with fully rational bayesian decision-makers we would expect the diverse committee to outperform the non-diverse committee when the state is yellow, but not when it is purple. In the experiment we find that the diverse committee always outperforms the non-diverse committee at least weekly. In the purple state 76 (75) percent of decisions are correct and in the yellow state 59 (49) percent of decisions are correct in the diverse (non-diverse) committee. Only the latter difference is statistically significant (t-test, $p < 0.05$).

	Independent			*Correlated*		
	(1)	(2)	(3)	(4)	(5)	(6)
DIVERSE	0.249	0.249	0.246	0.732***	0.756**	0.813***
	(0.334)	(0.333)	(0.312)	(0.207)	(0.230)	(0.203)
Constant	4.514***	4.531***	4.824***	3.937***	2.653***	5.106**
	(0.238)	(0.742)	(1.099)	(0.169)	(0.676)	(1.551)
Observations	900	900	900	900	900	900
R-squared	0.051	0.051	0.126	0.029	0.048	0.125

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 3.4: **Welfare. Aligned Preferences**. LPM regression of binary variable indicating that committee decision was correct on dummy indicating whether committee was diverse. Columns (1)-(3) show independent information and Columns (4)-(6) correlated information. All regressions include session fixed effects. Columns (2) and (5) linearly control for period and columns (3) and (6) have period fixed effects.

Table 3.4 shows the effect of diversity in the committee on welfare (payoffs). The effect is positive throughout, but statistically significant only with correlated information. In this case payoffs are about 18% higher in diverse compared to non-diverse committees (column (4)). As payoffs are directly proportional to correct decisions in each state this effect should not be surprising. The reason that the effect on payoffs is somewhat higher than the effect on correct decisions is that the purple state was somewhat more frequent in the experiment.

In sum, as suggested by Propositions 1 and 2, diverse committees make better decisions if and only if information is correlated. In this case the difference in both correct decisions and welfare is highly statistically

significant.

We now dig somewhat deeper into individual decisions in **AL-CORR**. First we address the question of correlation neglect. Both the Bayesian model and the model with correlation neglect predict that people should vote yellow when the signal distribution is YYY or YPY (i.e. if the minority group member has a yellow signal) and that people should vote purple when there are at least two purple signals. The only case where the two models disagree is YYP, i.e. the case where both majority group members have a yellow signal and the minority group member a purple signal. A bayesian decision maker would vote purple in this case as the two yellow signals are correlated and payoffs are higher if the purple state is guessed correctly. With correlation neglect decision makers would vote yellow as for they do not take the correlation into account. In our data there is little evidence that participants account for correlation in the committee. Overall 69% of decisions in diverse committees are consistent with Bayesian reasoning. It is more informative, however, to consider the case YYP, where both majority group members have a yellow signal and the minority group member a purple signal. In this case only 38% of decisions are in line with Bayesian reasoning, while 62% of decisions show correlation neglect. In non-diverse committees around 62% of decisions are consistent with Bayesian reasoning.

Chat Data We analyze chat data to understand better communication in committees. We first ask whether participants reveal their signals truthfully. For each chat and participants we define a dummy "Truth" which takes the value 1 if participants reveal their signal truthfully, i.e state "purple" ("yellow") if their signal was purple (yellow). The dummy "Lie" takes the value 1 if participants explicitly lie about their signal, i.e. state "yellow" when the signal was purple or vice versa. The dummy "No Mention" takes the value 1 if participants don't mention their signal in the chat. Classification was done by one research assistant and one co-author (Zia UH Khan) independently and revealed consistent results (deviations are less than 2 percentage points in all cases).[15] The figures and tables below are based on the classification produced by the research assistant.

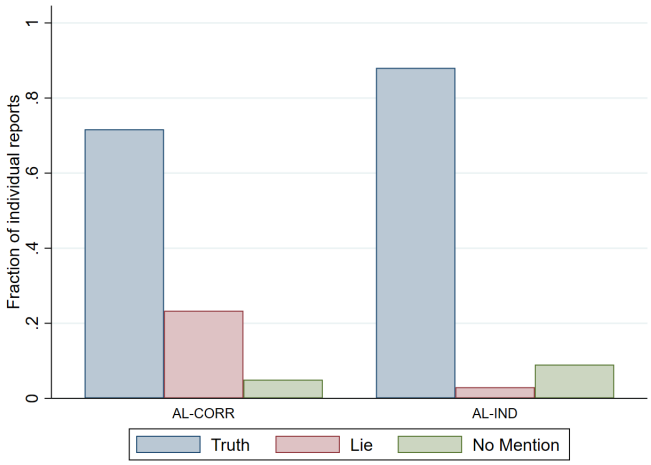


Figure 3.1: Frequency of instances where individuals reveal their signal truthfully ("Truth"), explicitly lie about it ("Lie") or do not mention it at all in treatments with aligned preferences.

Figure 3.1 shows the frequency of instances where individuals reveal their signal truthfully, explicitly lie about it or do not mention it at all. In treatment **AL-CORR** participants reveal their signal truthfully less often (71.6% of cases) than in treatment **AL-IND** (88.0%). This difference is highly statistically significant (t-test $p < 0.001$).

[15]The classification is rarely controversial with participants almost always stating explicitly "My ball was purple" or "I had yellow". Appendix Figure xxx shows a few typical chats....

Participants are somewhat more truthful when committees are diverse (72.3% vs 69.6%) in **AL-CORR** but the difference is not statistically significant at conventional levels (t-test $p = 0.1796$). In **AL-IND** there is virtually no difference between diverse and non-diverse committees in terms of truth-telling (88% vs 87.9%).

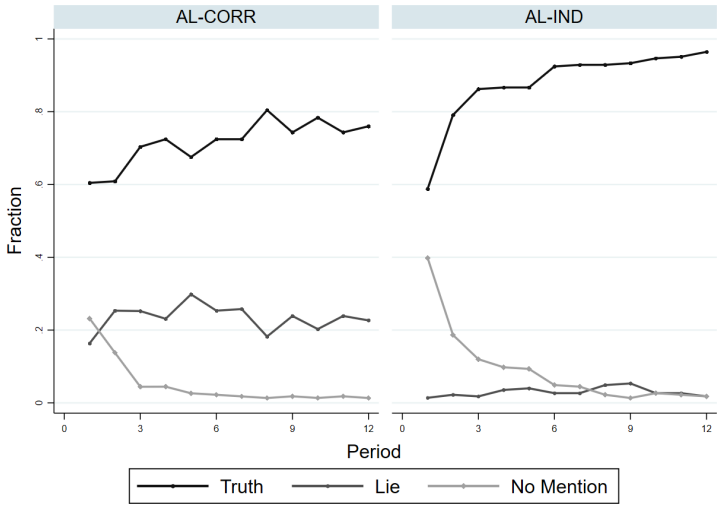


Figure 3.2: Frequency of instances where individuals reveal their signal truthfully ("Truth"), explicitly lie about it ("Lie") or do not mention it at all in treatments with aligned preferences over time.

Figure 3.2 shows truth-telling over time. In both treatments around 60 percent of participants tell the truth initially. But, while in **AL-IND** participants quickly converge to revealing their signal truthfully all the time, in treatment **AL-CORR** truth telling stabilized below 80 percent. One possible reason why we observe less truth-telling in **AL-CORR** is that some participants might try to correct for potential biases of others by misrepresenting their private information.

We also ask whether committees discuss explicitly that signals are correlated. Correlation is discussed 4.06% of the time in **AL-CORR** and 36% of committees discuss correlation at least once in this treatment. This suggests that at least sometimes an explicit effort is made at addressing correlation neglect and convincing others of the need to account for correlation. Correlation is discussed somewhat more often in diverse committees, but the difference is not statistically significant. Non-surprisingly much fewer committees explicitly discuss correlation in **AL-IND**, where signals are independent, with correlation discussed 0.9% of the time.

3.5.2 Results: Misaligned Preferences

This section reports our results for the case of misaligned preferences. We start with regressions testing the predictions of Propositions 3 and 4 and then dig deeper into the chats and the underlying decision-making process.

Table 3.5 shows the percentage of correct committee decisions depending on the type of committee. Proposition 3 predicted that with independent information committees consisting of only moderates will have higher decision accuracy than either type of diverse committees. It is tested directly in columns (1)-(2), where the baseline contains only the moderate non-diverse committees. The table shows that those committees do better than diverse committees with extremists as predicted by Proposition 3. They do also better than non-diverse committees with extremists. Diverse committees with a moderate majority do somewhat worse compared to fully moderate committees but here the difference is not statistically significant. The reason that committees with extremists do worse is that extremists have a stronger preference for voting purple even if the balance of probabilities does not favour that state. Indeed extremists in diverse committees

	Independent				*Correlated*			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
DIVERSE			-0.032	-0.033			0.042	0.041
			(0.033)	(0.033)			(0.039)	(0.040)
DIV-EXT	-0.145**	-0.145**			-0.006	-0.006		
	(0.059)	(0.060)			(0.066)	(0.066)		
DIV-MOD	-0.054	-0.055			0.019	0.020		
	(0.036)	(0.036)			(0.051)	(0.050)		
NOT-EXT	-0.124*	-0.124*			-0.065	-0.064		
	(0.061)	(0.061)			(0.084)	(0.083)		
Constant	0.850***	0.863***	0.764***	0.777***	0.663***	0.674***	0.617***	0.628***
	(0.044)	(0.054)	(0.028)	(0.038)	(0.060)	(0.080)	(0.027)	(0.059)
Observations	876	876	876	876	888	888	888	888
R-squared	0.019	0.020	0.013	0.013	0.037	0.037	0.036	0.036

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 3.5: **Percent of Correct committee decisions. Misaligned Preferences**. LPM regression of binary variable indicating that committee decision was correct on dummy indicating the type of committee. Columns (1)-(4) show independent information and Columns (5)-(8) correlated information. All regressions include session fixed effects. Even columns linearly control for period.

vote purple around 62%, which is slightly more often than moderates in diverse committees (59%), but much more often than moderates in homogeneous committees (50% in line with the actual frequency of a purple state being realized). Appendix Table C.6.2 shows that committees are much better in correctly identifying the purple state than they are at identifying the yellow state. Indeed conditionally on the committee getting it wrong, the state is yellow almost 67% of the time in **MIS-CORR** and 62% in **MIS-IND**. This is an increase of more than 5 percentage points compared to the case with aligned preferences. In columns (3) and (4) we aggregate all diverse (and non-diverse) committees. Those columns show that - on average - these two types of committees do not differ in terms of the quality of decision-making. The reason is that - unlike predicted by theory - non-diverse committees with extremists do as badly as diverse committees with an extremist majority, while moderate-majority committees do somewhat better than predicted. Columns (5)-(8) focus on the case of correlated information, where from Proposition 4 we would have expected diverse committees to do better in terms of decision accuracy. Indeed the decisions of diverse committees are more accurate, but the difference is not statistically significant.

	Independent				*Correlated*			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
DIVERSE			-0.056	-0.085			0.427	0.427
			(0.262)	(0.268)			(0.312)	(0.312)
DIV-EXT	-0.696	-0.727*			0.191	0.190		
	(0.406)	(0.420)			(0.449)	(0.449)		
DIV-MOD	-0.175	-0.228			0.386	0.387		
	(0.339)	(0.358)			(0.358)	(0.357)		
NOT-EXT	-0.697	-0.720			-0.256	-0.255		
	(0.493)	(0.497)			(0.629)	(0.629)		
Constant	7.134***	7.765***	6.646***	7.268***	4.686***	4.707***	4.498***	4.514***
	(0.330)	(0.504)	(0.226)	(0.371)	(0.400)	(0.589)	(0.215)	(0.491)
Observations	876	876	876	876	888	888	888	888
R-squared	0.036	0.041	0.033	0.039	0.052	0.052	0.052	0.052

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 3.6: **Welfare for moderate committee members. Misaligned Preferences**. Regression welfare for moderate committee members on dummies indicating type of committee. Columns (1)-(4) show independent information and Columns (5)-(8) correlated information. All regressions include session fixed effects. Even columns linearly control for period.

Table 3.6 analyzes the implications of committee type on the welfare of moderate committee members. Columns (1)-(4) focus again on independent information, where we would expect (from Proposition 3) that non-diverse moderate committees produce higher welfare for moderate committee members than either type of diverse committee. We do indeed observe this in line with the fact that those committees have higher decision accuracy, even though the welfare differences are mostly not statistically significant. With

correlated information the picture looks somewhat different. Here diverse committees seem to do generally better than non-diverse committees, even though the differences are not statistically significant. It should also be noted that we do not find evidence for the welfare ranking predicted in Proposition 4. In particular moderate-majority committees do better than extremist-majority committees unlike what is predicted by theory.

Chat Data The percentage of truth-telling is 64.6 percent with independent information compared to 72.1 percent with correlated info (Figure 3.3). In line with our expectations from theory, there is hence much less truth telling with independent information when preferences are misaligned compared to the case where preferences are aligned (see Figure 3.1).

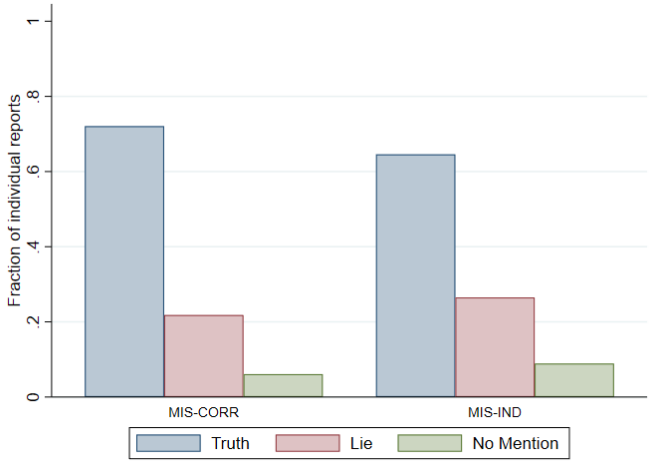


Figure 3.3: Frequency of instances where individuals reveal their signal truthfully ("Truth"), explicitly lie about it ("Lie") or do not mention it at all in treatments with misaligned preferences.

In both **MIS-CORR** and **MIS-IND**, there is more truth-telling in homogeneous committee compared to diverse committees, as predicted by theory, but the differences are small and not statistically significant. Hence one of the main theoretical costs of diversity - that it may lead to less truth-telling - does not seem to economically nor statistically important in our data. We do find that extremists lie more than moderates with independent information (57.4 vs 71.2 percent truth-telling, p-value 0.0025), while there is no difference with correlated information.

Figure 3.4 shows the proportion of truth-telling over time. In treatment **MIS-CORR**, the proportion of truth-telling initially increases sharply and then remains approximately stable at just below 80 percent. There is a small decrease in truth-telling over the last three rounds, mirroring an increase in the proportion of lies. The proportion of participants who do not mention their signal is only significantly above zero in the first few rounds of the experiment. In treatment **MIS-IND** there are no clear trends but both the proportion of participants who lie and the proportion of participants who do not mention their signal remain substantially above zero at all times. Overall lying is clearly more prevalent with misaligned preferences than with aligned preferences (Figure 3.2), especially when information is independent.

We also asked whether participants discuss the fact that information is correlated for participants of the same type. They do so in only 2.4% of cases in **MIS-CORR** and never in **MIS-IND**. While correlation is not discussed frequently 38% of participants are at least once in a committee that discusses correlation.

3.5.3 Results: Questionnaire

At the end of the experiment we elicited three measures from our participants (in addition to demographics): (i) a measure of risk aversion; (ii) a measure of correlation neglect and (iii) a measure of confidence

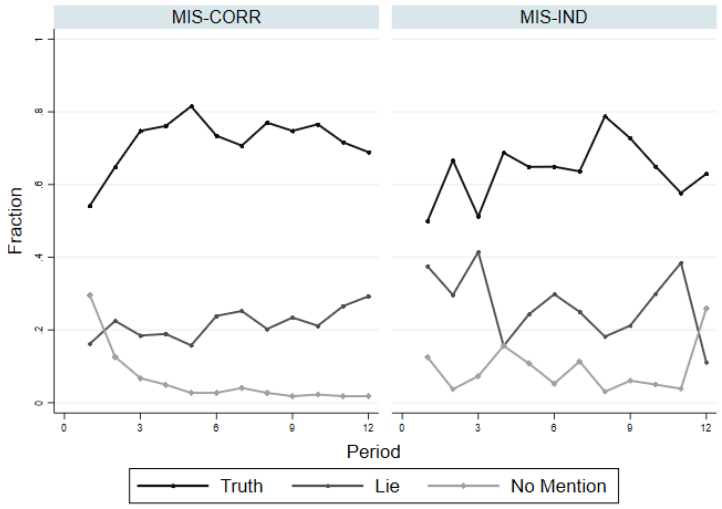


Figure 3.4: Frequency of instances where individuals reveal their signal truthfully ("Truth"), explicitly lie about it ("Lie") or do not mention it at all in treatments with misaligned preferences over time.

in their choice. Risk aversion is a useful control in any situation where participants have to make decisions under uncertainty. We elicit risk aversion by asking participants how willing they are to take risks on a scale from 1,...,5. Correlation neglect is interesting in our context as accounting for correlation is a key ingredient for good decision-making in treatments **CORR**. We elicit correlation neglect using a task from Enke and Zimmermann (2018). Last, confidence is an important measure in all contexts where individuals deliberate to reach a joint decision. The reason is that confidence can impact how people self select into participating in the deliberation process and to which extent they are keen to influence the committee decision. If those with higher quality decisions tend to have higher confidence in their decisions we can expect them to have a higher weight in joint decision. Of course we only observe confidence only after deliberation, so it should be kept in mind that our confidence measure could be impacted by the deliberation itself. We measure confidence by asking participants how often they believe they guessed the state correctly compared to others with options (i) a lot less often, (ii) a bit less often, (iii) about the same, (iv) a bit more often or (v) a lot more often. We call answers (i) and (ii) "low confidence" and answers (iv) and (v) "high confidence".

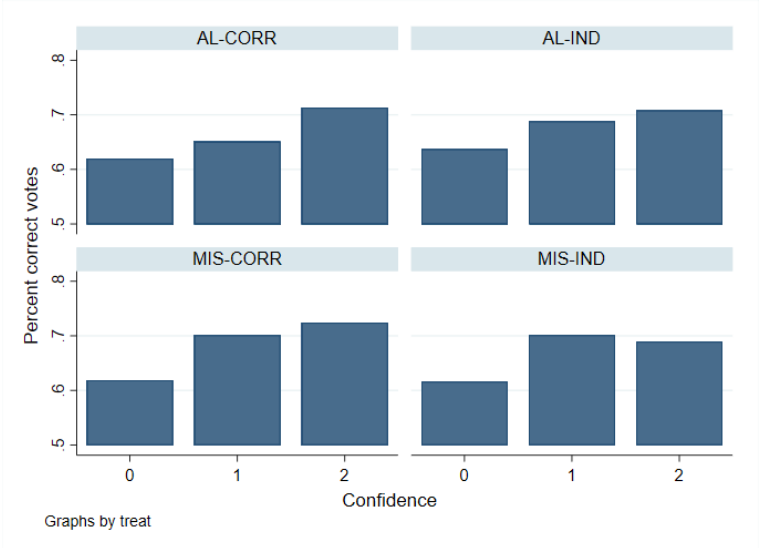


Figure 3.5: Percent of correct votes (votes matching the state) depending on confidence expressed at the end of the experiment.

Figure 3.5 shows the percentage of time a participant voted correctly (matched the state) depending on whether they expressed low, medium or high confidence. Participants with higher confidence have a higher percentage of correct votes. This shows that participants have - on average - a good sense of how well they were doing in the experiment. One exception is maybe treatment **MIS-IND** where the relationship between correct votes and confidence is not monotone. Appendix Figures C.7.1 and C.7.2 show that neither risk attitude nor correlation neglect predict correct votes.

	Aligned				*Misaligned*			
	AL-IND		AL-CORR		MIS-IND		MIS-CORR	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Share DIVERSE	0.250	0.221	0.337	0.307	-0.409*	-0.425*	0.492**	0.470**
	(0.307)	(0.310)	(0.288)	(0.291)	(0.221)	(0.223)	(0.227)	(0.227)
Age		0.000		-0.005		0.007		-0.011
		(0.008)		(0.008)		(0.006)		(0.009)
Gender		-0.109		-0.001		-0.080		-0.131
		(0.094)		(0.097)		(0.087)		(0.092)
Constant	1.155***	1.340***	0.651**	0.988***	1.741***	1.816***	1.245***	1.815***
	(0.255)	(0.355)	(0.269)	(0.362)	(0.244)	(0.301)	(0.240)	(0.340)
Observations	225	225	225	225	219	219	222	222
R-squared	0.023	0.029	0.068	0.083	0.115	0.124	0.059	0.084

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 3.7: Confidence regressed on the share of diverse committees a participant was part of. All regressions contain session fixed effects and even columns have controls for age, gender and field of study.

How does diversity of the committee impact confidence? Table 3.7 shows how confidence is impacted by the share of diverse committees a participant has experienced. While in the treatments with aligned preferences there is no statistically significant impact of diversity on confidence, this is different when preferences are misaligned. In **MIS-IND** confidence is lower for those who experienced many diverse committees. This is in line with diverse committees actually doing somewhat worse in these treatments (though the difference is not statistically significant). It could also be a result of the conflict resulting from misaligned preferences and having more disagreement in the committee. By contrast in **MIS-CORR** those who experienced more diverse committees express higher confidence in line with decisions in diverse committees being somewhat more accurate than in the non-diverse committees in this treatment. Diverse committees also provide participants with a more accurate signal structure in this treatment.

3.6 Discussion

We investigated the potential of diversity to influence committee decision-making. We think of diversity as more nuanced than ethnic or gender-based diversity an. It comes either in the form of informational diversity or preference diversity. We presented a model in which committee members receive private information on a state of the world, deliberate and vote. Committee members belonged to one of two groups which differed along two dimensions: (a) their preferences and (b) their information structures. We showed that diverse committees optimal when information structures are correlated within groups, while homogeneous committees make fewer mistakes when there is no informational correlation within groups but there is heterogeneity of preferences across groups. We tested the model's key predictions in a laboratory experiment. The model's predictions were largely borne out - diverse committees outperform homogeneous committees when information structures are correlated. This occurred despite subjects revealing more information than theory predicts and updating imperfectly.

There are several directions for future research. One obvious extension is to study larger committees. Another interesting avenue could be to study learning more explicitly (Ely and Thomas, 2020). Future research could also consider broader and general equilibrium effects of committee diversity. In the labour market context Niederle et al., (2012), for example, have shown that diversity costs at the individual firm level may be offset, from the social viewpoint, at equilibrium through increased supply.

Importantly, we should note that our research has focused on the "business case" for diversity and whether or not diverse committees make better decisions than others. Of course there are also normative reasons to consider diversity especially in cases where there is heterogeneity in preferences. Those are clearly important to be kept in mind when making recommendations on the optimal degree of diversity of a committee. Similarly, while we have focused on the impact of diversity on information aggregation; increasing or reducing diversity may have effects e.g. the level of discrimination in an organisation (Daskalova, 2021; Aboutalebi, 2020; Coffman et al., 2021).

Bibliography

Aboutalebi, Z. 2019. The Diversity Paradox.

Ahern, K.R. and Dittmar, A.K. 2012. The changing of the boards: The impact on firm valuation of mandated female board representation. The quarterly journal of economics, 127(1), pp.137-197.

Alesina, A. and Ferrara, E.L. 2005. Ethnic diversity and economic performance. Journal of economic literature, 43(3), pp.762-800.

Apesteguia, J., Azmat, G. and Iriberri, N. 2012. The impact of gender composition on team performance and decision making: Evidence from the field. Management Science, 58(1), pp.78-93.

Austen-Smith, D. and Feddersen, T.J. 2006. Deliberation, preference uncertainty, and voting rules. American political science review, 100(2), pp.209-217.

Austen-Smith, D. and Feddersen, T.J. 2009. Information aggregation and communication in committees. Philosophical Transactions of the Royal Society B: Biological Sciences, 364(1518), pp.763-769.

Azmat, G., 2019. Gender diversity in teams. IZA World of Labor.

Bagues, M.F. and Esteve-Volart, B. 2010. Can gender parity break the glass ceiling? Evidence from a repeated randomized experiment. The Review of Economic Studies, 77(4), pp.1301-1328.

Bagues, M., Sylos-Labini, M. and Zinovyeva, N. 2017. Does the gender composition of scientific committees matter?. American Economic Review, 107(4), pp.1207-38.

Bardhi, A. and Bobkova, N. 2021. Local evidence and diversity in minipublics.

Battaglini, M. and Makarov, U. 2014. Cheap talk with multiple audiences: An experimental analysis. Games and Economic Behavior, 83, pp.147-164.

Becker, G.S. 2010. The economics of discrimination. University of Chicago press.

Cai, H. and Wang, J.T.Y. 2006. Overcommunication in strategic information transmission games. Games and Economic Behavior, 56(1), pp.7-36.

Chandrasekhar, A.G., Larreguy, H. and Xandri, J.P. 2020. Testing models of social learning on networks: Evidence from two experiments. Econometrica, 88(1), pp.1-32.

Charness, G., Karni, E. and Levin, D. 2007. Individual and group decision making under risk: An experimental study of Bayesian updating and violations of first-order stochastic dominance. Journal of Risk and uncertainty, 35(2), pp.129-148.

Charness, G. and Levin, D. 2005. When optimal choices feel wrong: A laboratory study of Bayesian updating, complexity, and affect. American Economic Review, 95(4), pp.1300-1309.

Che, Y.K. and Kartik, N. 2009. Opinions as incentives. Journal of Political Economy, 117(5), pp.815-860.

Chen, R. and Chen, Y. 2011. The potential of social identity for equilibrium selection. American Economic Review, 101(6), pp.2562-89.

Chen, Y., Kartik, N. and Sobel, J. 2008. Selecting cheap-talk equilibria. Econometrica, 76(1), pp.117-136.

Coffman, K., Flikkema, C.B. and Shurchkov, O. 2021. Gender Stereotypes in Deliberation and team decisions. Games and Economic Behavior.

Coughlan, P.J. 2000. In defense of unanimous jury verdicts: Mistrials, communication, and strategic voting. American Political science review, 94(2), pp.375-393.

Crawford, V.P. and Sobel, J. 1982. Strategic information transmission. Econometrica: Journal of the Econometric Society, pp.1431-1451.

D'Acunto, F., Fuster, A. and Weber, M. 2021. Diverse policy committees can reach underrepresented groups (No. w29275). National Bureau of Economic Research.

Daskalova, V. 2021. In-group Favouritism in Collective Decisions. Working paper.

De Paola, M. and Scoppa, V. 2015. Gender discrimination and evaluators' gender: evidence from Italian academia. Economica, 82(325), pp.162-188.

Deschamps, P. 2018. Gender quotas in hiring committees: A boon or a bane for women. Sciences Po LIEPP Working Paper, 82.

Ely, R.J. and Thomas, D.A. 2020. Getting Serious About Diversity. Harvard Business Review, 98(6), pp.114-122.

Enke, B. and Zimmermann, F. 2019. Correlation neglect in belief formation. The Review of Economic Studies, 86(1), pp.313-332.

Fehrler, S. and Hughes, N., 2018. How transparency kills information aggregation: theory and experiment. American Economic Journal: Microeconomics, 10(1), pp.181-209.

Gerling, K., Grüner, H.P. Kiel, A. and Schulte, E. 2005. Information acquisition and decision making in committees: A survey. European Journal of Political Economy, 21(3), pp.563-597.

Göhlmann, S. and Vaubel, R. 2007. The educational and occupational background of central bankers and its effect on inflation: An empirical analysis. European Economic Review, 51(4), pp.925-941.

Goeree, J.K. and Yariv, L. 2011. An experimental study of collective deliberation. Econometrica, 79(3), pp.893-921.

Gomez, L.E. and Bernet, P. 2019. Diversity improves performance and outcomes. Journal of the National Medical Association, 111(4), pp.383-392.

Gompers, P. and Kovvali, S. 2018. Diversity dividend. Harvard Business Review, pp.72-77.

Grimm, V. and Mengel, F. 2020. Experiments on belief formation in networks. Journal of the European Economic Association, 18(1), pp.49-82.

Hamilton, B.H., Nickerson, J.A. and Owan, H. 2012. Diversity and productivity in production teams. In Advances in the Economic Analysis of participatory and Labor-managed Firms. Emerald Group Publishing Limited.

Hjort, J. 2014. Ethnic divisions and production in firms. The Quarterly Journal of Economics, 129(4), pp.1899-1946.

Hong, L. and Page, S.E., 2001. Problem solving by heterogeneous agents. Journal of economic theory, 97(1), pp.123-163.

Hoogendoorn, S., Oosterbeek, H. and Van Praag, M. 2013. The impact of gender diversity on the performance of business teams: Evidence from a field experiment. Management Science, 59(7), pp.1514-1528.

Jensen, M.K. 2019. On the Strategic Benefits of Diversity (No. 1419). School of Economics, University of Surrey.

Kandel, E. and Lazear, E.P. 1992. Peer pressure and partnerships. Journal of political Economy, 100(4), pp.801-817.

Kets, W. and Sandroni, A. 2021. A theory of strategic uncertainty and cultural diversity. The Review of Economic Studies, 88(1), pp.287-333.

Lazear, E. 1999. Culture and Language. Journal of Political Economy.

Levin, D., Peck, J. and Ivanov, A. 2016. Separating bayesian updating from non-probabilistic reasoning: An experimental investigation. American Economic Journal: Microeconomics, 8(2), pp.39-60.

Levy, G. and Razin, R. 2015a. Correlation neglect, voting behavior, and information aggregation. American Economic Review, 105(4), pp.1634-45.

Levy, G. and Razin, R. 2015b. (Old) Combining Forecasts. 1–29.

Lyons, E. 2017. Team production in international labor markets: Experimental evidence from the field. American Economic Journal: Applied Economics, 9(3), pp.70-104.

Malmendier, U., Nagel, S. and Yan, Z. 2021. The making of hawks and doves. Journal of Monetary Economics, 117, pp.19-42.

Malmström, M., Voitkane, A. Johansson, J. and Wincent, J. 2018. VC Stereotypes About Men and Women Aren't Supported by Performance Data. Harvard Business Review.

Marx, B., Pons, V. and Suri, T. 2021. Diversity and team performance in a Kenyan organization. Journal of Public Economics, 197, p.104332.

Mengel, F. 2021. Gender Bias in Opinion Aggregation. International Economic Review.

Niederle, M., Segal, C. and Vesterlund, L. 2013. How costly is diversity? Affirmative action in light of gender differences in competitiveness. Management Science, 59(1), pp.1-16.

Ottaviani, M. and Sørensen, P. 2001. Information aggregation in debate: who should speak first?. Journal of Public Economics, 81(3), pp.393-421.

Pogorelskiy, K. and Shum, M. 2019. News we like to share: How news sharing on social networks influences voting outcomes. Available at SSRN 2972231.

Prat, A. 2002. Should a team be homogeneous?. European Economic Review, 46(7), pp.1187-1207.

Le Quement, M.T. and Marcin, I. 2020. Communication and voting in heterogeneous committees: An experimental study. Journal of Economic Behavior Organization, 174, pp.449-468.

Sommers, S.R. 2006. On racial diversity and group decision making: identifying multiple effects of racial composition on jury deliberations. Journal of personality and social psychology, 90(4), p.597.

Thomas, D. and Ely, R.J. 1996. Making Differences Matter: A New Paradigm for Managing Diversity. Harvard Business Review, 71.

Weidmann, B. and Deming, D.J. 2021. Team players: How social skills improve team performance. Econometrica, 89(6), pp.2637-2657.

Appendix A

Online Appendix for "An Experimental Study on the Role of Social Norms About Tax Compliance"

A.1 Appendix: Rule Following Proclivity and Evasion

As subjects in our experiment face a joint decision of evasion and avoidance, we would like to see how the rule-following task fares in case of evasion. Like avoidance, we find that evasion is lower under the social cost treatment (352.1 Lab Sterling) than under no social cost treatment (452.7 Lab Sterling). However, a Wilcoxon rank-sum test shows that this difference is not statistically significant (p-value=0.84). Rule following is much more strongly negatively correlated with evasion than with avoidance. A spearman rank correlation test on evasion and rule-sensitivity gives $\rho = -0.2922$ with $p - value = 0.0021$. Evasion is also found to be significantly lower among rule followers (216.3 Lab Sterling) than among rule breakers (587.06 Lab Sterling) split at mean rule-sensitivity (Wilcoxon rank-sum test gives $p - value = 0.002$). Avoidance is different only among strongly rule following and rule breaking subjects. The results indicate that subjects in our experiment show higher morale costs associated with evasion than with avoidance. This is consistent with previous literature (Blaufus et al., 2016).

We also report OLS regressions for evasion in table A1. The regression output in the first column shows that rule sensitivity is significantly and negatively related to evasion. In the second column, we introduce the treatment dummy which is found to be insignificant. In column 3 and 4, we include demographics, risk preferences, other control variables and session fixed effects etc. Rule-sensitivity continues to be significantly and negatively related to evasion. R^2 improves from 6 percent in column 1 to about 25 percent in column 4.

	Amount Evaded			
rule-sensitivity	-402.4**	-390.1**	-232.7*	-270.1*
	(120.3)	(125.9)	(112.3)	(120.2)
Unequal Opportunity		-63.18	-8.974	130.4***
		(75.42)	(106.7)	(15.61)
Constant	630.3***	657.4***	-10.16	37.90
	(77.48)	(70.49)	(237.1)	(236.2)
Observations	109	109	109	109
R-squared	0.065	0.068	0.218	0.246
Demographics			YES	YES
Risk			YES	YES
Justified			YES	YES
TaxFair			YES	YES
Session FE				YES

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table A1: OLS regression: Dependent variable is the amount evaded. Rule-sensitivity $\in [0, 1]$. *Demographics* include age and gender. *Unequal Opportunity* is the treatment dummy (=1 under unequal opportunity to earn condition). *Risk* is the risk preference measured using the method by Dohmen et al. (2011). *TaxFair* measures whether subjects think the tax system applied in the experiment was justified. *Justified* measures whether subjects think it can always or never be justified to avoid taxes. *Session FE* controls for session fixed effects. Standard errors are clustered around session.

A.2 Appendix: Initial Beliefs

A.2.1 Empirical Expectation Treatment

Figure A.1 shows the distribution of subjects' initial beliefs about the use of tax loophole by other experimental participants. The figure shows that subjects expect other participants to avoid taxes. 2.94

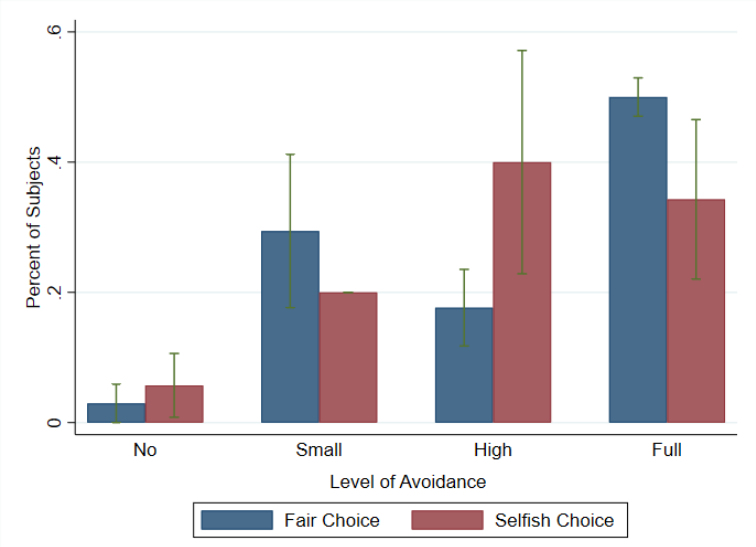


Figure A.1: Subjects' belief of level of tax avoidance by majority

percent of subjects in the *FC* treatment thought that the most common choice in the previous session was to 'not avoid taxes at all'. About 5.71 percent of subjects in the *SC* treatment in had the same opinion. 50 percent of subjects in the *FC* treatment thought that the most common choice in the previous session was to 'avoid taxes maximally'. About 34.29 subjects in the *SC* treatment had the same opinion. Also, in our norm elicitation task, the perceived norm that emerges is such that not avoiding taxes is considered the most appropriate action whereas maximal tax avoidance is considered the least appropriate action. Initial beliefs of subjects may be a signal that they do not find it worthwhile for individuals to conform to such a norm in the absence of any formal sanctions (avoidance was modelled as an innocuous choice).

A.2.2 Normative Expectation Treatment

Figure A.2 shows the distribution of initial beliefs about the level of avoidance undertaken for the treatment fair belief and selfish belief. 27.6 percent of subjects in the *FB* treatment believed that most

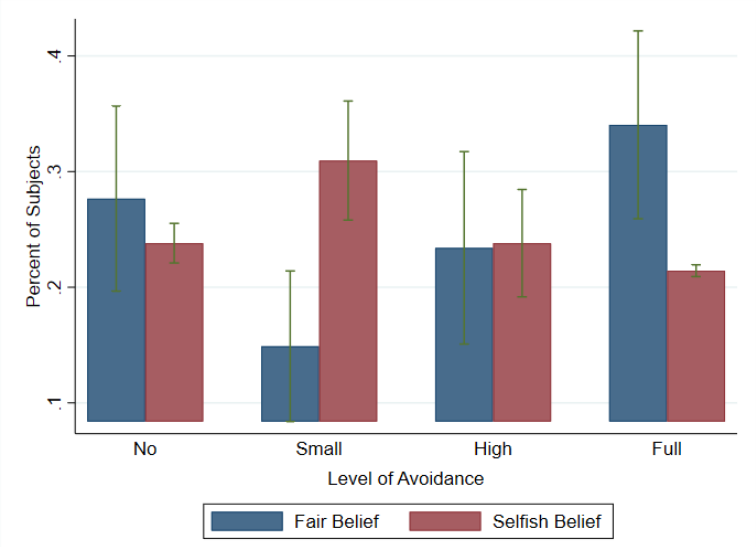


Figure A.2: Subjects' belief of level of tax avoidance by majority

subjects in the previous session thought that the most appropriate action was to 'not avoid taxes at all'. Only 23.8 percent in the *SB* treatment had the same opinion. Similarly, 34 percent in the *FB* treatment believed that the most common choice in the previous session was to 'avoid taxes maximally'. Only 21.4 in the *SB* treatment. On the other hand, in our norm elicitation task the perceived norm that emerges is where not avoiding taxes is the most appropriate action whereas maximal tax avoidance is the least appropriate action. Initial beliefs of subjects may be a signal that they do not find it worthwhile for individuals to conform to such a norm in the absence of any formal sanctions (there were no consequences for avoiding taxes).

A.2.3 Conflicting Information Treatments

Figure A.3 shows results for the treatments involving conflicting information about expectations. In treatment *FB+SC*, none of the subjects thought that 'not avoiding taxes' was the most common choice in the previous session whereas about 22.5 percent subjects believed that most subjects in the previous session thought that 'not avoiding taxes' was the most appropriate action. Likewise, about 51.6 percent of subjects believed that most subjects in the previous session undertook maximum tax avoidance whereas only 29 percent of subjects believed that maximal tax avoidance was the most appropriate action. These differences are significant. Therefore, we see a similar conflict in the empirical and normative expectations of subjects as discovered by the HMRC study.

In treatment *SB+FC* about 3.2 percent of subjects believed that 'not avoiding taxes' was the most common choice in the previous session while about 51.6 percent subjects in the same treatment believed that most subjects in the previous session thought 'not avoiding taxes' was the most appropriate action. Similarly, about 41.5 percent of subjects believed that most subjects in the previous session undertook maximal tax avoidance whereas only 9.6 percent of subjects believed that undertaking maximal tax avoidance was the most appropriate action. Therefore, again we see conflicting empirical and normative expectations.

Initial beliefs of subjects may be a signal that they do not find it in the immediate self-interest of individuals to conform to such a norm in the absence of any formal sanctions (there were no consequences for undertaking tax avoidance). While they expect a large number of people to hold beliefs against tax avoidance, in the

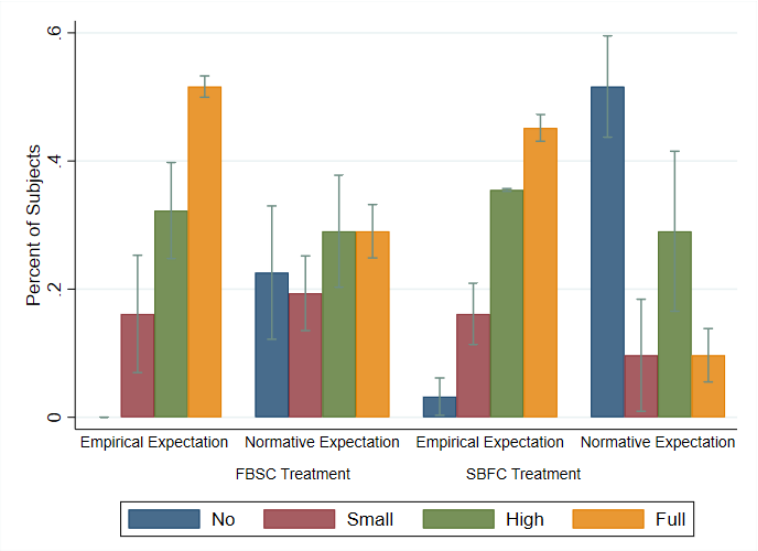


Figure A.3: Subjects' belief of level of tax avoidance by majority

absence of any formal sanctions they also expect violations to be very common.

A.3 Appendix: Tax Avoidance at Different Levels of Loophole Usage

A.3.1 Normative Information Treatments

In Figure A.4 (a), we classify tax avoidance into four categories. As before, choosing to avoid 0 sheets is considered as no tax avoidance and choosing to avoid 5 sheets is considered as undertaking maximal tax avoidance. Small level of tax avoidance arises when someone chooses to avoid either 1 or 2 sheets and high level of tax avoidance arises when someone chooses to avoid 3 or 4 sheets. We now present the results. In the fair belief treatment more subjects did not avoid taxes than those in the control group (42.5 vs. 22 percent; t-test; $p = 0.0312$) and less subjects undertook maximal tax avoidance (34 percent vs. 38 percent; t-test; $p = 0.6886$) though the latter is not significant. In the selfish belief treatment, fewer subjects made no use of the loophole (7 percent vs. 22 percent; t-test; $p = 0.0409$) and more subjects undertook maximal tax avoidance (76.1 percent vs. 38 percent; t-test; $p = 0.0001$).

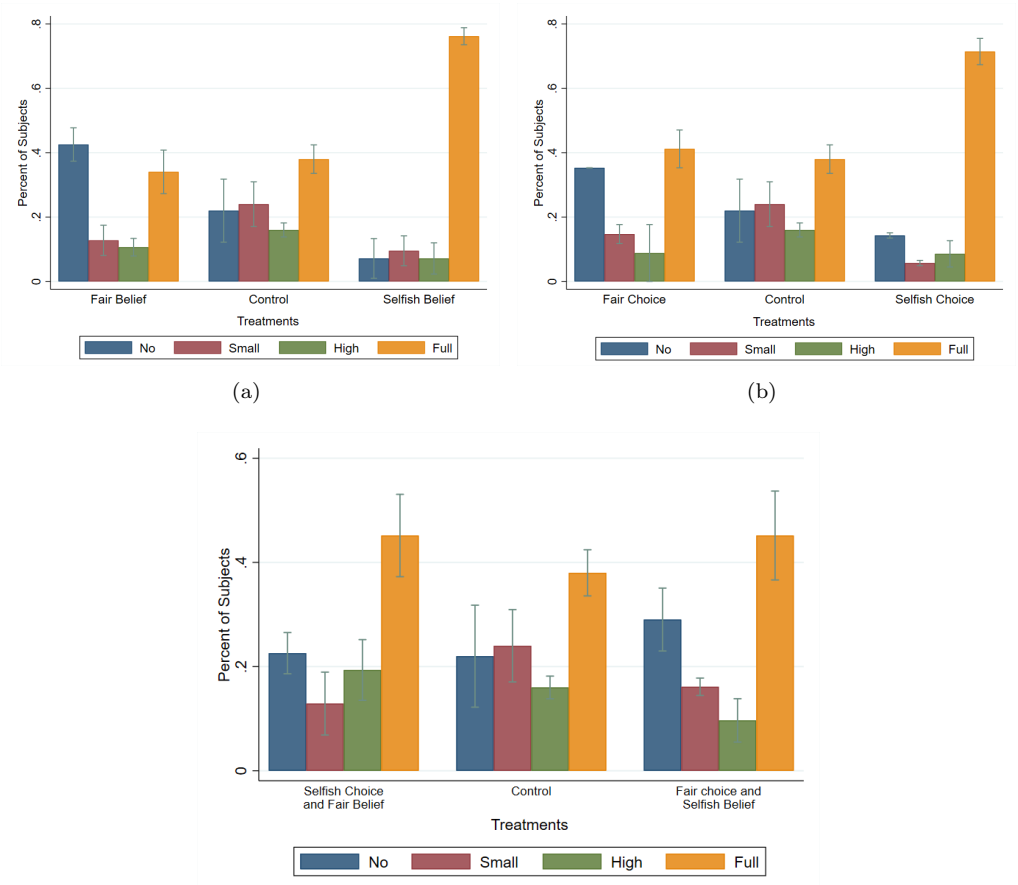


Figure A.4: **(a)** Subjects' belief of level of tax avoidance during their session

A.3.2 Empirical Information Treatments

Fig A.4 (b) shows that compared with the control group, percentage of subjects who do not avoid taxes is higher in the fair choice treatment (35.3 percent vs. 22 percent; t-test gives $p = 0.092$) and lower in the selfish choice treatment (14.3 percent vs. 22 percent; t-test gives $p = 0.8117$). Similarly, maximal tax

avoidance is significantly higher in the selfish belief treatment (71.4 percent vs. 38 percent; t-test gives $p = 0.001$) and not significantly different in the fair belief treatment (41.2 percent vs. 38 percent; t-test gives $p = 0.386$). Within each treatment, small and high levels of tax avoidance are not significantly different from one another. The impact of selfish choice treatment is strong in the sense that going from control to selfish choice treatment maximal tax avoidance increases drastically and this comes together with a decrease in other levels of tax avoidance. When we move from control to fair choice treatment we see a significant increase in no tax avoidance whereas differences at all other levels are insignificant.

A.3.3 Conflicting Information Treatments

We then move on to look at the incidence of tax avoidance for each level of tax avoidance (see Figure A.4 (c)). Higher fraction of people avoid no or small amounts of taxes in fair choice (SB+**FC**) management (29 and 16.1 percent respectively) than in selfish choice (FB+**SC**) management (22.5 and 12.9 percent). Also a higher fraction of subjects undertake high level of tax avoidance in treatment with selfish choice (FB+**SC**) management (19.3 percent) as compared with the treatment with fair choice (SB+**FC**) management (9.6 percent). The results show that subjects, when faced with a choice to avoid taxes, give relatively more weight to empirical part of the information than normative part of the information.

A.4 Appendix: Experimental Instructions

A.4.1 Rule-Following Task Instructions

You are about to start stage 1. In this part of the experiment, you will decide how to allocate 100 balls between two buckets. Your task is to put each ball, one-by-one, into one of the two buckets: the blue bucket or the yellow bucket.

The balls will appear in the center of your screen, and you can allocate each ball by clicking and dragging it to the bucket of your choice.

For each ball you put in the blue bucket, you will receive 5 pence, and for each ball you put in the yellow bucket, you will receive 10 pence.

The rule is to put the balls in the blue bucket.

Once the experiment begins, you will have 8 minutes to put the balls into the buckets. When you are finished, please wait quietly until the end of the 8-minute period.
Your payment from this stage will be based on your decisions. It is the sum of payments from the blue and yellow buckets.

This is the end of the instructions for stage 1. If you have any questions, please raise your hand and an experimenter will answer them privately.

A.4.2 Encryption Task Instructions

You are about to start stage 2 of the experiment. During this stage of the experiment you can earn income by performing a task. The task is described below

Important Note: On the left hand side of your table you see a page with the title '*Encryption Task Example*'. It provides an example of what you will see on the screen when you begin stage 2. You should NOW pick it up and use it to understand the instructions for the task. You should leave it back once you have understood the instructions given below.

Description and Instructions of the Encryption Task

In the '*Encryption Task Example*' you see a table (at the top of the page) which contains letters and numbers assigned to each letter. Beneath this table you see a word. You are required to encrypt this word using the numbers from the given table. You should enter the relevant number in the box beneath each letter. Some of you will receive **5** letter words to encrypt and some of you will receive **10** letter words to encrypt. Everyone has an equal chance of getting either 5 letter words or 10 letter words. You can think of it as a toss of a fair coin such that if it lands HEADS you get 5 letter words and if it lands TAILS you get 10 letter words. This selection is completely random and does not depend on decisions that you have made earlier in the experiment. Whether you get 5 letter words or 10 letter words, you will all have only **5 minutes** to earn as much as you can.

In the '*Encryption Task Example*' the correct encryption should be as given below. Remember, you will be selected to encrypt either 5 letter words or 10 letters words but not both. This selection is purely random.

Word	B	L	A	C	K	S	M	I	T	H
Code	6	25	2	24	8	16	18	19	11	5

Once you have coded the given word click '**NEXT**' to move to the next word. If you make a mistake, computer will inform you to correct the mistake before you can proceed to the next word. For each correct encryption you will earn **100** Lab Sterling. At the end of the experiment, for each **100** Lab Sterling you will get 1 Pound Sterling.

Please make sure you understand what you are required to do. If you have any questions at this point please raise your hand and one of the lab assistants will come and answer you.

A.4.3 Tax Return Instructions

There is a tax rate of **50 percent** on your income i.e. you have to pay half of your income in taxes. This tax is due only on the income that you decide to declare. No matter how much income you decide to declare there is a fixed probability of audit by the computer which is equal to **10 percent** i.e. 1 out of every 10 subjects will be audited and their declared income will be compared with their true income. None of your decisions determine whether you will be audited or not. The audit is completely random.
If you get audited and you declared your true income, there will be no consequences. If you get audited and you declared less than your true income then the entire amount that you did not declare will be taken away (i.e. the fine rate is equal to **100 percent**).

Example: let us say someone earned 4,000 Lab Sterling and declared only 3,000 Lab Sterling. If they get audited then the 1,000 Lab Sterling that they did not declare will be taken away as a fine and they will get 1,500 Lab Sterling=4,000-[(50/100)×3,000]-1,000. If they do not get audited then there will be no fine and they will get

$$2,500 \textit{ Lab Sterling} = 4,000 - [(50/100)3,000]$$

The tax system also involves redistribution. The amount of taxes paid by you and by everyone else in the experiment will be added and will then be redistributed equally among all of you. The amount available for redistribution depends on the amount of taxes collected from all subjects during the experiment. In order to understand this, suppose taxes collected during the experiment are 18,000 Lab Sterling and there are 30 subjects (including you) in this session then you will each get an additional *600 Lab Sterling = (18,000/30).*
Tax Loophole: On the same screen you will also be given an opportunity to exempt some of your encryptions from taxes so that you pay less tax and your income increases. Note that this opportunity has arisen due to a loophole in the tax system which you can work in your favour. You will have to specify how much you want to use the loophole by moving the slider. If you do not wish to use the loophole, you can select 0. Each additional use of the loophole will reduce your taxable income by *100 Lab Sterlings.* Up to 5 encryptions can be made tax free if you make a maximal use of tax avoidance opportunity.
After Tax Income: Before you declare your income and choose the level of the loophole, you can check your expected final after-tax income corresponding to different levels of declared income and use of loophole. You will see three tables related to three different possibilities. First possibility is that in your session the average amount declared by other subjects is less than your earnings from the encryption task. Second possibility is that the average amount declared by other subjects is equal to your earnings from the encryption task. Third possibility is that the average amount declared by other subjects is more than your earnings from the encryption task. These possible levels of income have been calculated for the case when no audit takes place.

In the box at the bottom of the screen you will be asked to declare your income and make use of the loophole.

A.4.4 Scenario Task Instructions

On the following screens, you will read descriptions of **THREE** scenarios. These descriptions correspond to three different situations in which one person, 'Individual A', must make a decision.

For each of these situations, you will be given a description of the situation and the decision faced by 'Individual A'. This description will include several possible choices available to this individual.

After you have read the description of the situation, you will be asked to evaluate the different possible choices available to 'Individual A' and to decide, for each of the possible actions, whether taking that action would be '**socially appropriate**' and 'consistent with moral or proper social behaviour' or '**socially inappropriate**' and 'inconsistent with moral or proper social behaviour'. By socially appropriate, we mean behaviour that most people agree is the 'correct' or 'ethical' thing to do in the given situation. Another way to think about what we mean is that if 'Individual A' were to select a socially inappropriate choice, then someone else might be angry at 'Individual A' for doing so.

In each of your responses, we would like you to answer as truthfully as possible, based on your opinions of what constitutes socially appropriate or socially inappropriate behaviour.

To give you an idea of how the experiment will proceed, we will go through an example and show you how you will indicate your response. On the next screen you will see an example of a situation.

A.4.5 Further Instructions for Scenario Task

For each situation, you will read a description of the situation. You will then indicate your appropriateness rating by placing a check mark in the corresponding box.

At the end of the experiment today, we will randomly select one of the situations. For this situation, we will also randomly select one of the possible choices that Individual A could make. Thus, we will select both a situation and one possible choice at random. For the choice selected, we will determine which response was selected by the most people here today. If you give the same response as that most frequently given by other people, then you will receive an additional £5. This amount will be paid to you, in cash, at the conclusion of the experiment. For instance, if we were to select the example situation above and the possible choice 'Leave the wallet where it is', and if your response had been 'somewhat socially inappropriate', then you would receive £5, in addition to your earnings from previous tasks, if this was the response selected by most other people in today's session. Otherwise you would receive £0 from this task.

If you have any questions from this point on, please raise your hand and wait for the experimenter to come to you.

A.4.6 Scenario Task Description

Individual A has been invited to an experiment identical to yours along with other participants. The identity of each participant is anonymous and unknown to other participants and the experimenter.

In the experiment, Individual A and all other subjects perform an 'Encryption Task' to earn an income.

They are then asked to declare the income for tax collection and make use of the tax loophole. Individual A as well as all other subjects read the following instructions.

There is a tax rate of **50 percent** on your income i.e. you have to pay half of your income in taxes. This tax is due only on the income that you decide to declare. No matter how much income you decide to declare there is a fixed probability of audit by the computer which is equal to **10 percent** i.e. 1 out of every 10 subjects will be audited and their declared income will be compared with their true income. None of your decisions determine whether you will be audited or not. The audit is completely random.
If you get audited and you declared your true income, there will be no consequences. If you get audited and you declared less than your true income then the entire amount that you did not declare will be taken away (i.e. the fine rate is equal to **100 percent**).

Example: let us say someone earned 4,000 Lab Sterling and declared only 3,000 Lab Sterling. If they get audited then the 1,000 Lab Sterling that they did not declare will be taken away as a fine and they will get 1,500 Lab Sterling=4,000-[(50/100)×3,000]-1,000. If they do not get audited then there will be no fine and they will get

$$2,500 \textit{ Lab Sterling} = 4,000 - [(50/100)3,000]$$

The tax system also involves redistribution. The amount of taxes paid by you and by everyone else in the experiment will be added and will then be redistributed equally among all of you. The amount available for redistribution depends on the amount of taxes collected from all subjects during the experiment. In order to understand this, suppose taxes collected during the experiment are 18,000 Lab Sterling and there are 30 subjects (including you) in this session then you will each get an additional *600 Lab Sterling = (18,000/30).*
Tax Loophole: On the same screen you will also be given an opportunity to exempt some of your encryptions from taxes so that you pay less tax and your income increases. Note that this opportunity has arisen due to a loophole in the tax system which you can work in your favour. You will have to specify how much you want to use the loophole by moving the slider. If you do not wish to use the loophole, you can select 0. Each additional use of the loophole will reduce your taxable income by *100 Lab Sterlings*. Up to 5 encryptions can be made tax free if you make a full use of the loophole.
After Tax Income: Before you declare your income and choose the level of the loophole, you can check your expected final after-tax income corresponding to different levels of declared income and use of loophole. You will see three tables related to three different possibilities. First possibility is that in your session the average amount declared by other subjects is less than your earnings from the encryption task. Second possibility is that the average amount declared by other subjects is equal to your earnings from the encryption task. Third possibility is that the average amount declared by other subjects is more than your earnings from the encryption task. These possible levels of income have been calculated for the case when no audit takes place.

In the box at the bottom of the screen you will be asked to declare your income and make use of the loophole.

A.4.7 Eliciting Empirical Expectations

In today's experiment you and all the other participants made a decision to use the tax loophole. In this task we ask you to guess the percentage of participants in your session who made a certain use of the tax loophole. Remember, there were six possible actions to choose from i.e. 0, 1, 2, 3, 4, and 5.

0 meant no use of the loophole whereas 5 meant the maximal use of the loophole.
In the table below, write your estimate of the percentage of people you believe used that specific action. For each action, you can enter any number from 0 to 100. However, the sum of all your estimates should be exactly equal to 100. At the end of the experiment we will randomly select one of the possible actions. You will earn an additional £2 if your estimate is within 4 percentage points of the actual use of the loophole.

For example, let us say the randomly chosen action was 3 and the actual use of this action during this session was 21 percent i.e. 21 percent of the participants decided to exempt 3 encryptions from their taxable income by using the loophole. If your estimate for this action was anywhere between 17 and 25 then you will get the additional £2. However, if your estimate was either less than 17 or greater than 25 then you will earn £0.

Action	**Percentage of People Who Chose This Action**
0	
1	
2	
3	
4	
5	

A.4.8 Eliciting Personal Normative Beliefs and Normative Expectations

Use of Loophole Task

Note: The total number of subjects in today's experiment is 30.

Q1. Do you think one should make no use of a tax loophole (i.e. choose 0 Encryptions) to reduce the amount of taxes they have to pay? YES/NO

Q2. How many subjects in today's experiment do you think said yes to the previous question? .....

Q3. Do you think one should make a small use of a tax loophole (i.e. choose 1 or 2 Encryptions) to reduce the amount of taxes they have to pay? YES/NO

Q4. How many subjects in today's experiment do you think said yes to the previous question? .....

Q5. Do you think one should make a high use of a tax loophole (i.e. choose 3 or 4 Encryptions) to reduce the amount of taxes they have to pay? YES/NO

Q6. How many subjects in today's experiment do you think said yes to the previous question? .....

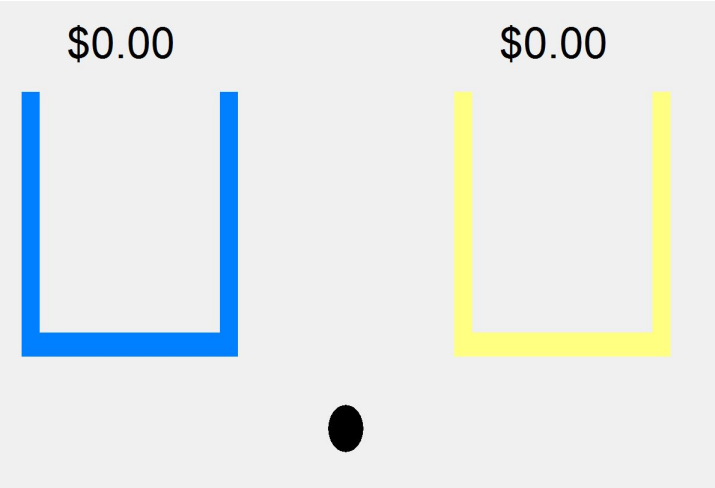
Q7. Do you think one should make a full use of a tax loophole (i.e. choose 5 Encryptions) to reduce the amount of taxes they have to pay? YES/NO

Q8. How many subjects in today's experiment do you think said yes to the previous question? .....

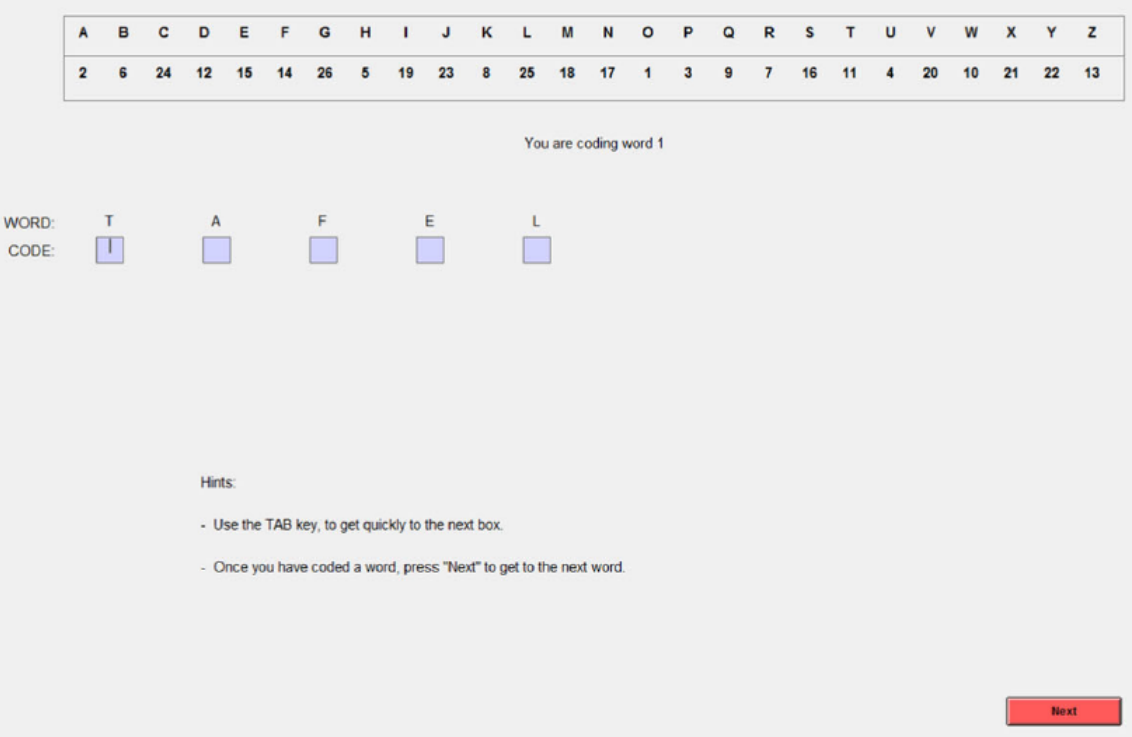
We will randomly select either Question 2, 4, 6 or 8 and you will earn an additional £2 if your answer to the randomly selected question is the same as the actual number of people who gave that response.

A.5 Appendix: Screenshots

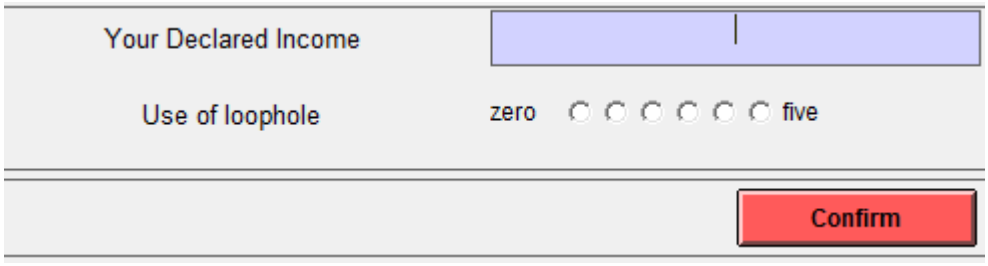
A.5.1 Stage 1: Rule Following



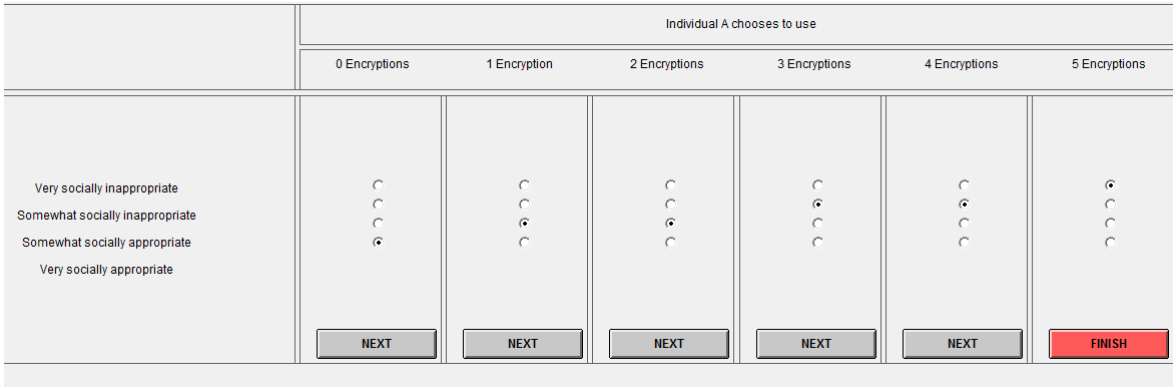
A.5.2 Stage 2: Income Generation Stage



A.5.3 Stage 3: Tax Return



A.5.4 Stage4: Norm Elicitation Task



Appendix B

Online Appendix for "Incentives for whistle-blowing: An Experiment"

B.1 Appendix: Questionnaire

B.1.1 Survey Questions

1. Your gender is: (a) Male) (b) Female
2. Your Age [in years]: ……..
3. Country of Origin (birth): ……
4. What is your ethnic origin? ……
5. What is your religion? ……
6. Your current marital status is

   (a) Single (b) Married (c) Widowed (d) Separated
7. How many siblings do you have [including you]? ……
8. Currently, I am

   (a) A Student (b) Part-time Employed (c) Full-time Employed (d) Self-Employed (e) Unemployed (f) Retired
9. Which qualification are you currently pursuing?

   (a) Bachelors (b) Masters (c) Doctoral degree (d) Other (e) None
10. Which program are you currently pursuing or in which program did you finish your last degree in?

    (a) Economics (b) Politics / International Relations (c) Psychology (d) Mathematics / Statistics (e) Language / Linguistics (f) Accounting / Finance (g) Law (h) Social Sciences (i) Engineering (j) Other
11. Which is the highest qualification that you have completed?

    (a) High school or below (b) Bachelors (c) Masters (d) Doctoral degree
12. What is your monthly disposable income (after rent, approximately in pounds)?

    (a) Less than £300 (b) £301-£600 (d) £601-£900 (e) £901-£1200 (f) More than £1200
13. Generally, how willing are you to take risks?

    (1) VERY UNWILLING …….. (10) VERY WILLING
14. Have you ever been asked to pay a bribe?

    (a) YES (b) NO
15. How widespread you think bribe taking is in your country?

    (a) Almost no public officials are engaged in it (b) A few public officials are engaged in it (c) Most public officials are engaged in it (d) Almost all public officials are engaged in it
16. In what context do you mostly hear about corrupt behavior? (Choose one)

    (a) Involving politicians and bureaucrats (b) Involving companies and rich people (c) Harassment of ordinary people for basic services (d) Other
17. In general, you can trust people.

    (a) Strongly Disagree (b) Somewhat Disagree(c) Somewhat Agree (d) Strongly Agree

18. Nowadays, you cannot rely on anybody.

    (a) Strongly Disagree (b) Somewhat Disagree(c) Somewhat Agree (d) Strongly Agree

19. Could you tell me how much confidence do you have in civil service in your country

    (a) A great deal of confidence (b) Quite a lot of confidence (c) Not very much confidence (d) No confidence at all

20. Could you tell me how much confidence do you have in the Parliament in your country

    (a) A great deal of confidence (b) Quite a lot of confidence (c) Not very much confidence (d) No confidence at all

21. Could you tell me how much confidence do you have in the judicial system in your country

    (a) A great deal of confidence (b) Quite a lot of confidence (c) Not very much confidence (d) No confidence at all

    Please tell me for the following action whether you think it can always be justified, never be justified, or something in between.

22. A government employee asking a citizen to pay a bribe in order to approve his/her passport application.

    (1) NEVER JUSTIFIED ..... (5) ALWAYS JUSTIFIED

23. Citizen offers a bribe to a government employee to accept his passport application without proper documents.

    (1) NEVER JUSTIFIED ..... (5) ALWAYS JUSTIFIED

24. Which of the following law against corruption is implemented in your country? (Choose one)

    (a) Bribe giver and bribe taker are both fined (b) Bribe taker is fined and nothing happens to the bribe giver (c) Bribe giver is rewarded and bribe taker is fined (d) Nothing happens to the bribe giver or the bribe taker (e) I don't know about such laws

25. Consider the political situation in the country where you are a citizen. In this country, where would you place yourself ideologically: to the left or to the right of the political center? Choose among the following five possible answers:

    (a) LEFT (b) CENTRE-LEFT (c) CENTRE (d) CENTRE-RIGHT (e) RIGHT

B.1.2 Summary Statistics of Survey Responses

Variable Name	**Value Range**	**Mean**	**(SD)**	**Min. Value**	**Max. Value**
Ever paid a bribe	0-1	0.19	0.39	0	1
Official asking bribe	1-5	1.43	0.79	1	5
Citizen offering bribe	1-5	1.69	0.98	1	5
Trust in Judiciary	1-4	2.68	0.85	1	4
Trust in Parliament	1-4	3.14	0.78	1	4
Trust in Civil Service	1-4	2.60	0.75	1	4
Trust Others	1-4	2.59	0.74	1	4
Rely on others	1-4	2.58	0.83	1	4

Table A1: Questionnaire Summary Statistics.

B.2 Appendix: Experimental Instructions

B.2.1 Encryption Task Instructions

You are about to start stage 1 of the experiment. During this stage you can earn a prize of 200 lab Sterlings by performing a task. At the end of the experiment, for every 100 Lab Sterlings you will be paid £2. The task is described below

Important Note: On the left hand side of your table you see a page with the title '*Encryption Task Example*'. It provides an example of what you will see on the screen when you begin stage 2. You should NOW pick it up and use it to understand the instructions for the task. You should leave it back once you have understood the instructions given below.

Description and Instructions of the Encryption Task

In the '*Encryption Task Example*' you see a table (at the top of the page) which contains letters and numbers assigned to each letter. Beneath this table you see a word. You are required to encrypt this word using the numbers from the given table. You should enter the relevant number in the box beneath each letter. All of you will receive **5** letter words to encrypt. You will have to encrypt **5 words** in **5 minutes** to win the prize.

In the '*Encryption Task Example*' the correct encryption should be as given below. Remember, you will be selected to encrypt either 5 letter words or 10 letters words but not both. This selection is purely random.

Word	P	A	N	D	A
Code	3	2	17	12	2

Once you have coded the given word click '**NEXT**' to move to the next word. If you make a mistake, computer will inform you to correct the mistake before you can proceed to the next word.

Please make sure you understand what you are required to do. If you have any questions at this point please raise your hand and one of the lab assistants will come and answer you.

B.2.2 Bribery Game Instructions

Treatment With Higher Evidence Threshold and Low Judicial Accuracy

You are now going to take part in stage 2 of the experiment. In this stage, you will be paired with another subject. Through a random draw one of you will be assigned the role of **Player A** and the other will be assigned the role of **Player B**.

Short Summary of Stage 2

- If you have successfully completed the encryption task then you have earned a prize worth 600 Lab Sterling. For every 100 Lab Sterling, you will be paid £2.
- Each group member will be informed who won the prize.
- Player A is asked to hand over the prizes in the group. First Player A decides whether to ask Player B to pay a charge to receive their prize. Player A can ask up to 200 lab Sterling.
- Player B decides whether to report Player A or not. Reporting is costly and it costs **150** Lab Sterling to Player B.
- A random draw then determines whether Player A is guilty or not.
- The likelihood of being found guilty depends on whether player A did ask for a charge and whether they were reported by player B (see below).
- Please read on for a more detailed explanation of Stage 2.

Scenario 1: (Player A does not ask for a charge)

If Player A decides not to ask for a charge to hand over the prize then Player B has to choose between two options,

1. Report Player A OR
2. Do not report Player A

Reporting costs **150** lab Sterlings to Player B. If Player B reports Player A then there is a **40%** chance that Player A is declared guilty.

In the following table you can see a summary of the possible payoffs if Player A does not ask for a charge.

If Player A does not ask for a charge	**Player A gets**	**Player B gets**
and if Player B does not report Player A	600	600
and if Player B reports Player A and report decides Player A is guilty (40% chance)	600-400	600+200-150
and if Player B reports Player A and report decides Player A is not guilty (60% chance)	600	600-150

Table A2: Payoff Table.

Scenario 2: (Player A asks for a Charge)

If Player A decides to ask for a charge to hand over the prize then he/she has to specify the amount of charge for Player B. Player A can choose any integer value from 1 to 200 (Lab Sterling) as the charge. Player B is then informed of the charge he/she needs to pay and can choose one of the four options,

1. Refuse to pay the charge and not report Player A OR
2. Refuse to pay the charge and report Player A OR
3. Pay the charge and not report Player A OR
4. Pay the charge and report Player A.

Reporting costs **150** Lab Sterling to Player B. If Player B reports Player A then there is a **60%** chance that Player A is declared guilty.

In the following table you can see a summary of the possible payoffs if Player A asks for a charge where **'C'** is the amount of charge.

If Player A does not ask for a charge	**Player A gets**	**Player B gets**
and if Player B refuses to pay and does not report player A	550	150
and if Player B refuses to pay and reports player A and if the report decides Player A is guilty (60% chance)	550-400	150+200-150
and if Player B refuses to pay and reports player A and if the report decides Player A is not guilty (40% chance)	550	150-150
and if Player B pays the charge and reports player A and if the report decides Player A is guilty (60% chance)	550+C-400	550-C+200-150
and if Player B pays the charge and reports player A and if the report decides Player A is not guilty (40% chance)	550+C	550-C-150
and if Player B pays the charge and does not report player A	550+C	550-C

Table A3: Payoff Table.

Treatment With Low Evidence Threshold and High Judicial Accuracy

You are now going to take part in stage 2 of the experiment. In this stage, you will be paired with another subject. Through a random draw one of you will be assigned the role of **Player A** and the other will be assigned the role of **Player B**.

Short Summary of Stage 2

- If you have successfully completed the encryption task then you have earned a prize worth 600 Lab Sterling. For every 100 Lab Sterling, you will be paid £2.
- Each group member will be informed who won the prize.
- Player A is asked to hand over the prizes in the group. First Player A decides whether to ask Player B to pay a charge to receive their prize. Player A can ask up to 200 lab Sterling.
- Player B decides whether to report Player A or not. Reporting is costly and it costs **10** Lab Sterling to Player B.
- A random draw then determines whether Player A is guilty or not.
- The likelihood of being found guilty depends on whether player A did ask for a charge and whether they were reported by player B (see below).
- Please read on for a more detailed explanation of Stage 2.

Scenario 1: (Player A does not ask for a charge)

If Player A decides not to ask for a charge to hand over the prize then Player B has to choose between two options,

1. Report Player A OR
2. Do not report Player A

Reporting costs **10** lab Sterlings to Player B. If Player B reports Player A then there is a **10%** chance that Player A is declared guilty.
In the following table you can see a summary of the possible payoffs if Player A does not ask for a charge.

If Player A does not ask for a charge	**Player A gets**	**Player B gets**
and if Player B does not report Player A	600	600
and if Player B reports Player A and report decides Player A is guilty (10% chance)	600-400	600+200-10
and if Player B reports Player A and report decides Player A is not guilty (90% chance)	600	600-10

Table A4: Payoff Table.

Scenario 2: (Player A asks for a Charge)

If Player A decides to ask for a charge to hand over the prize then he/she has to specify the amount of charge for Player B. Player A can choose any integer value from 1 to 200 (Lab Sterling) as the charge. Player B is then informed of the charge he/she needs to pay and can choose one of the four options,

1. Refuse to pay the charge and not report Player A OR
2. Refuse to pay the charge and report Player A OR
3. Pay the charge and not report Player A OR
4. Pay the charge and report Player A.

Reporting costs **10** Lab Sterling to Player B. If Player B reports Player A then there is a **90%** chance that Player A is declared guilty.

In the following table you can see a summary of the possible payoffs if Player A asks for a charge where **'C'** is the amount of charge.

If Player A does not ask for a charge	**Player A gets**	**Player B gets**
and if Player B refuses to pay and does not report player A	550	150
and if Player B refuses to pay and reports player A and if the report decides Player A is guilty (90% chance)	550-400	150+200-10
and if Player B refuses to pay and reports player A and if the report decides Player A is not guilty (10% chance)	550	150-10
and if Player B pays the charge and reports player A and if the report decides Player A is guilty (90% chance)	550+C-400	550-C+200-10
and if Player B pays the charge and reports player A and if the report decides Player A is not guilty (10% chance)	550+C	550-C-10
and if Player B pays the charge and does not report player A	550+C	550-C

Table A5: Payoff Table.

Treatment With Higher Evidence Threshold and High Judicial Accuracy

You are now going to take part in stage 2 of the experiment. In this stage, you will be paired with another subject. Through a random draw one of you will be assigned the role of **Player A** and the other will be assigned the role of **Player B**.

Short Summary of Stage 2

- If you have successfully completed the encryption task then you have earned a prize worth 600 Lab Sterling. For every 100 Lab Sterling, you will be paid £2.
- Each group member will be informed who won the prize.
- Player A is asked to hand over the prizes in the group. First Player A decides whether to ask Player B to pay a charge to receive their prize. Player A can ask up to 200 lab Sterling.
- Player B decides whether to report Player A or not. Reporting is costly and it costs **150** Lab Sterling to Player B.
- A random draw then determines whether Player A is guilty or not.
- The likelihood of being found guilty depends on whether player A did ask for a charge and whether they were reported by player B (see below).
- Please read on for a more detailed explanation of Stage 2.

Scenario 1: (Player A does not ask for a charge)

If Player A decides not to ask for a charge to hand over the prize then Player B has to choose between two options,

1. Report Player A OR
2. Do not report Player A

Reporting costs **150** lab Sterlings to Player B. If Player B reports Player A then there is a **10%** chance that Player A is declared guilty.
In the following table you can see a summary of the possible payoffs if Player A does not ask for a charge.

If Player A does not ask for a charge	**Player A gets**	**Player B gets**
and if Player B does not report Player A	600	600
and if Player B reports Player A and report decides Player A is guilty (10% chance)	600-400	600+200-150
and if Player B reports Player A and report decides Player A is not guilty (90% chance)	600	600-150

Table A6: Payoff Table.

Scenario 2: (Player A asks for a Charge)

If Player A decides to ask for a charge to hand over the prize then he/she has to specify the amount of charge for Player B. Player A can choose any integer value from 1 to 200 (Lab Sterling) as the charge. Player B is then informed of the charge he/she needs to pay and can choose one of the four options,

1. Refuse to pay the charge and not report Player A OR
2. Refuse to pay the charge and report Player A OR
3. Pay the charge and not report Player A OR
4. Pay the charge and report Player A.

Reporting costs **150** Lab Sterling to Player B. If Player B reports Player A then there is a **90%** chance that Player A is declared guilty.

In the following table you can see a summary of the possible payoffs if Player A asks for a charge where **'C'** is the amount of charge.

If Player A does not ask for a charge	**Player A gets**	**Player B gets**
and if Player B refuses to pay and does not report player A	550	150
and if Player B refuses to pay and reports player A and if the report decides Player A is guilty (90% chance)	550-400	150+200-150
and if Player B refuses to pay and reports player A and if the report decides Player A is not guilty (10% chance)	550	150-150
and if Player B pays the charge and reports player A and if the report decides Player A is guilty (90% chance)	550+C-400	550-C+200-150
and if Player B pays the charge and reports player A and if the report decides Player A is not guilty (10% chance)	550+C	550-C-150
and if Player B pays the charge and does not report player A	550+C	550-C

Table A7: Payoff Table.

Treatment With Low Evidence Threshold and Low Judicial Accuracy

You are now going to take part in stage 2 of the experiment. In this stage, you will be paired with another subject. Through a random draw one of you will be assigned the role of **Player A** and the other will be assigned the role of **Player B**.

Short Summary of Stage 2

- If you have successfully completed the encryption task then you have earned a prize worth 600 Lab Sterling. For every 100 Lab Sterling, you will be paid £2.
- Each group member will be informed who won the prize.
- Player A is asked to hand over the prizes in the group. First Player A decides whether to ask Player B to pay a charge to receive their prize. Player A can ask up to 200 lab Sterling.
- Player B decides whether to report Player A or not. Reporting is costly and it costs **10** Lab Sterling to Player B.
- A random draw then determines whether Player A is guilty or not.
- The likelihood of being found guilty depends on whether player A did ask for a charge and whether they were reported by player B (see below).
- Please read on for a more detailed explanation of Stage 2.

Scenario 1: (Player A does not ask for a charge)

If Player A decides not to ask for a charge to hand over the prize then Player B has to choose between two options,

1. Report Player A OR
2. Do not report Player A

Reporting costs **10** lab Sterlings to Player B. If Player B reports Player A then there is a **40%** chance that Player A is declared guilty.
In the following table you can see a summary of the possible payoffs if Player A does not ask for a charge.

If Player A does not ask for a charge	**Player A gets**	**Player B gets**
and if Player B does not report Player A	600	600
and if Player B reports Player A and report decides Player A is guilty (40% chance)	600-400	600+200-10
and if Player B reports Player A and report decides Player A is not guilty (60% chance)	600	600-10

Table A8: Payoff Table.

Scenario 2: (Player A asks for a Charge)

If Player A decides to ask for a charge to hand over the prize then he/she has to specify the amount of charge for Player B. Player A can choose any integer value from 1 to 200 (Lab Sterling) as the charge. Player B is then informed of the charge he/she needs to pay and can choose one of the four options,

1. Refuse to pay the charge and not report Player A OR
2. Refuse to pay the charge and report Player A OR
3. Pay the charge and not report Player A OR
4. Pay the charge and report Player A.

Reporting costs **10** Lab Sterling to Player B. If Player B reports Player A then there is a **60%** chance that Player A is declared guilty.

In the following table you can see a summary of the possible payoffs if Player A asks for a charge where **'C'** is the amount of charge.

If Player A does not ask for a charge	**Player A gets**	**Player B gets**
and if Player B refuses to pay and does not report player A	550	150
and if Player B refuses to pay and reports player A and if the report decides Player A is guilty (60% chance)	550-400	150+200-10
and if Player B refuses to pay and reports player A and if the report decides Player A is not guilty (40% chance)	550	150-10
and if Player B pays the charge and reports player A and if the report decides Player A is guilty (60% chance)	550+C-400	550-C+200-10
and if Player B pays the charge and reports player A and if the report decides Player A is not guilty (40% chance)	550+C	550-C-10
and if Player B pays the charge and does not report player A	550+C	550-C

Table A9: Payoff Table.

B.2.3 (Scenario) Norm Elicitation Task

Instructions

On the following screens, you will read descriptions of **THREE** scenarios. These descriptions correspond to three different situations in which one person, 'Player A' or 'Player B', must make a decision.

For each of these situations, you will be given a description of the situation and the decision faced by 'Player A' or 'Player B'. This description will include several possible choices available to this individual.

After you have read the description of the situation, you will be asked to evaluate the different possible choices available to 'Player A' or 'Player B' and to decide, for each of the possible actions, whether taking that action would be '**socially appropriate**' and consistent with moral or proper social behaviour or '**socially inappropriate**' and inconsistent with moral or proper social behaviour. By socially appropriate, we mean behaviour that most people agree is the 'correct' or 'ethical' thing to do in the given situation. Another way to think about what we mean is that if 'Player A' or 'Player B' were to select a socially inappropriate choice, then someone else might be angry at 'Player A' or 'Player B' for doing so.

In each of your responses, we would like you to answer as truthfully as possible, based on your opinions of what constitutes socially appropriate or socially inappropriate behaviour.

To give you an idea of how the experiment will proceed, we will go through an example and show you how you will indicate your response. On the next screen you will see an example of a situation.

Further Instructions for Scenario Task

Are there any questions about this example situation or about how to indicate your responses? On the following screens, you will see three different situations, all dealing with decisions that 'Player A' or 'Player B', a participant in an experiment, has to make.
For each situation, you will read a description of the situation. You will then indicate your appropriateness rating by placing a check mark in the corresponding box.

At the end of the experiment today, we will randomly select one of the situations. For this situation, we will also randomly select one of the possible choices that 'Player A' or 'Player B' could make. Thus, we will select both a situation and one possible choice at random. For the choice selected, we will determine which response was selected by the most people here today. If you give the same response as that most frequently given by other people, then you will receive an additional £2. This amount will be paid to you, in cash, at the conclusion of the experiment. For instance, if we were to select the example situation above and the possible choice 'Leave the wallet where it is', and if your response had been 'somewhat socially inappropriate', then you would receive £2, in addition to your earnings from previous tasks, if this was the response selected by most other people in today's session. Otherwise you would receive £0 from this task.

If you have any questions from this point on, please raise your hand and wait for the experimenter to come to you.

Situation 1: Player A's Choice

Player A and *Player B* participated in a another session of today's experiment. They have both earned a prize of **600** Lab Sterling. Player A has been asked to hand over the prize to Player B. Player A can choose whether to ask Player B to pay him/her a charge to get their prize. If Player A chooses to ask for a charge then he/she needs to specify the amount of charge. They can ask an amount from 1 to 200 Lab Sterling.
On the next screen you will see a table with the possible actions that 'Player A' could take. For each action, please indicate whether you believe choosing that action is 'very socially inappropriate', 'somewhat socially inappropriate', 'somewhat socially appropriate', or 'very socially appropriate'. By 'socially appropriate' we mean the behaviour most people agree is the 'correct' or 'ethical' thing to do in the given situation. You can indicate your appropriateness rating by placing a check mark in the relevant box.
Remember: When we select a situation and an action for payment, you will receive £2 only if your rating is the same as the rating chosen by most other people in this session.

Situation 2: Player B Decides When Player A Did Not Ask for a Charge

Player A and *Player B* participated in a another session of today's experiment. They have both earned a prize of **600** Lab Sterling. Player A has been asked to hand over the prize to Player B. Player A can choose whether to ask Player B to pay him/her a charge to get their prize. If Player A chooses to ask for a charge then he/she needs to specify the amount of charge. They can ask an amount from 1 to 200 Lab Sterling.
Player B can choose one of the following actions.

1. Report Player A OR
2. Do not report Player A

On the next screen you will see a table with the possible actions that 'Player A' could take. For each action, please indicate whether you believe choosing that action is 'very socially inappropriate', 'somewhat socially inappropriate', 'somewhat socially appropriate', or 'very socially appropriate'. By 'socially appropriate' we mean the behaviour most people agree is the 'correct' or 'ethical' thing to do in the given situation. You can indicate your appropriateness rating by placing a check mark in the relevant box.
Remember: When we select a situation and an action for payment, you will receive £2 only if your rating is the same as the rating chosen by most other people in this session.

Situation 3: Player B Decides When Player A Asked for a Charge

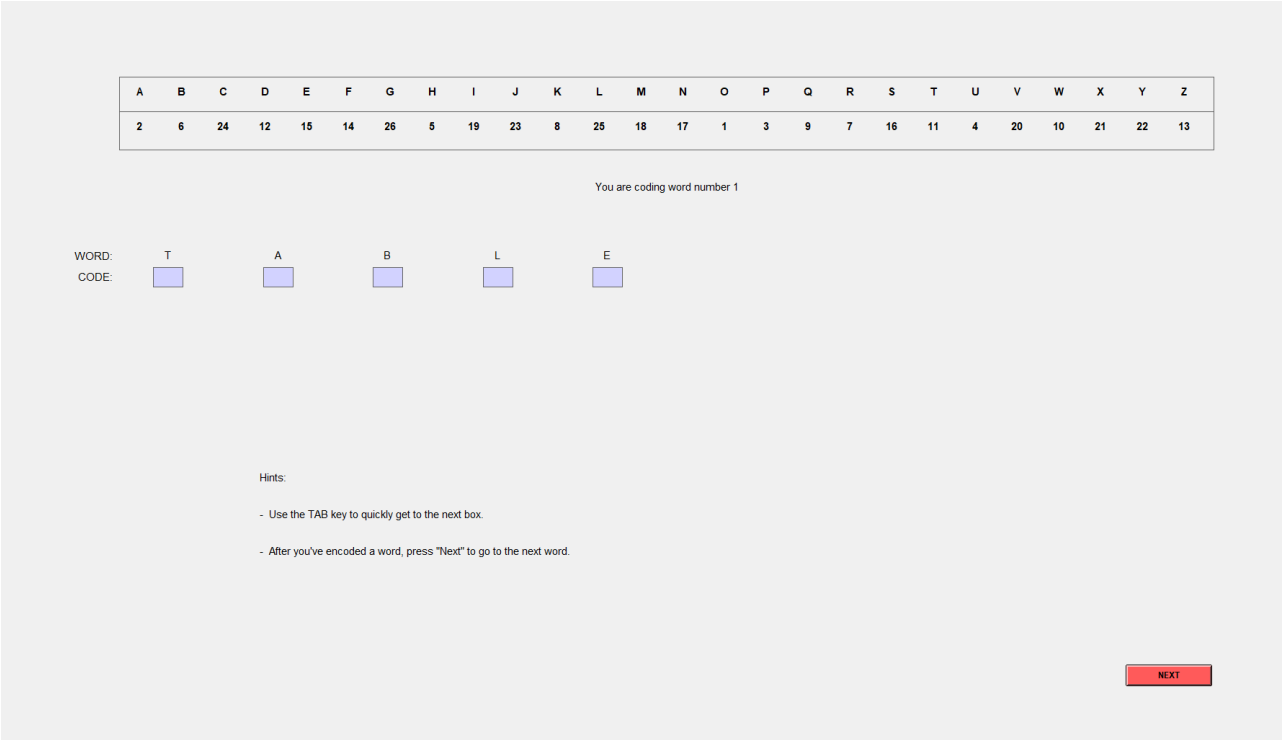
Player A and *Player B* participated in a another session of today's experiment. They have both earned a prize of **600** Lab Sterling. Player A has been asked to hand over the prize to Player B. Player A can choose whether to ask Player B to pay him/her a charge to get their prize. If Player A chooses to ask for a charge then he/she needs to specify the amount of charge. They can ask an amount from 1 to 200 Lab Sterling.
Player B can choose to one of the following actions.

1. Refuse to pay the charge and not report Player A OR
2. Refuse to pay the charge and report Player A OR
3. Pay the charge and not report Player A OR
4. Pay the charge and report Player A.

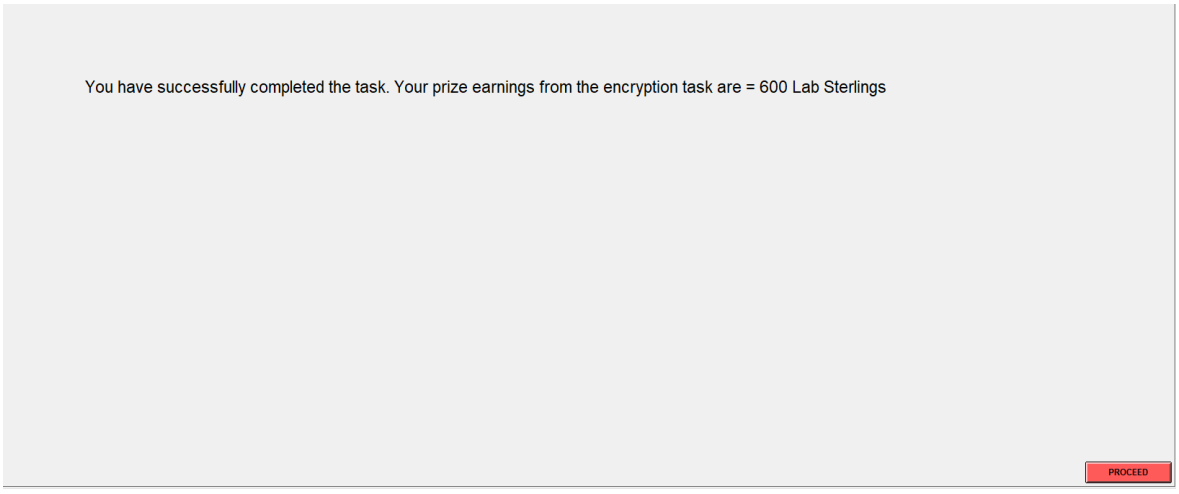
On the next screen you will see a table with the possible actions that 'Player A' could take. For each action, please indicate whether you believe choosing that action is 'very socially inappropriate', 'somewhat socially inappropriate', 'somewhat socially appropriate', or 'very socially appropriate'. By 'socially appropriate' we mean the behaviour most people agree is the 'correct' or 'ethical' thing to do in the given situation. You can indicate your appropriateness rating by placing a check mark in the relevant box.
Remember: When we select a situation and an action for payment, you will receive £2 only if your rating is the same as the rating chosen by most other people in this session.

B.3 Screenshots

B.3.1 Encryption Task - Encryption Table

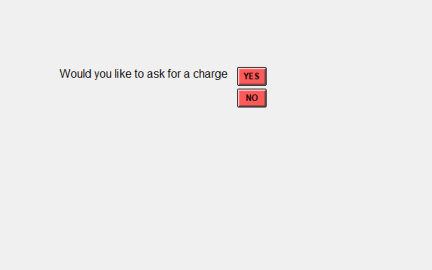


B.3.2 Encryption Task - Prize Announced

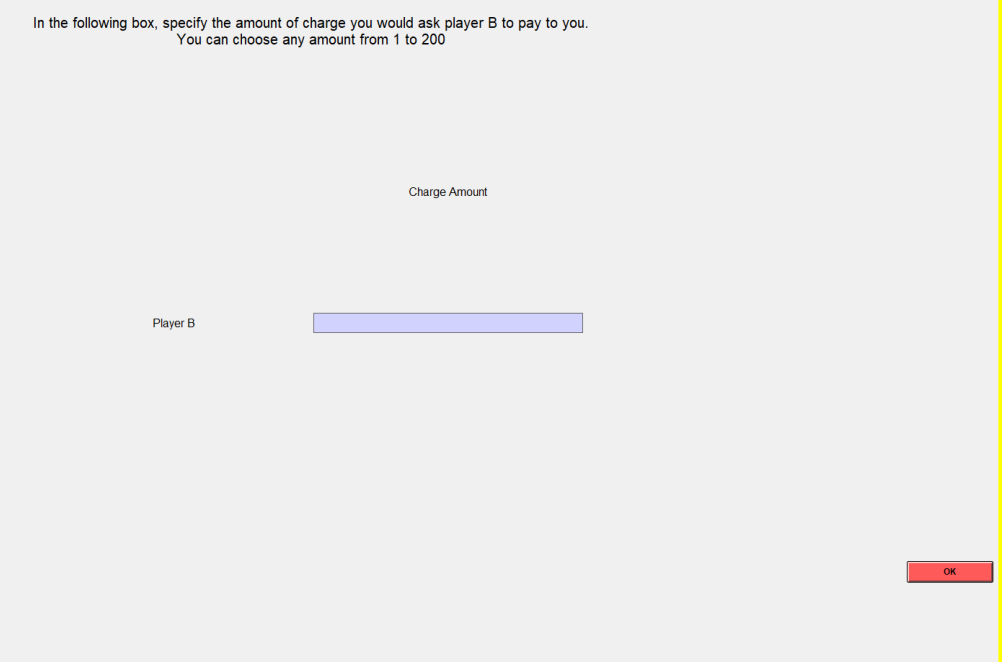


B.3.3 Bribery Game - Roles Revealed

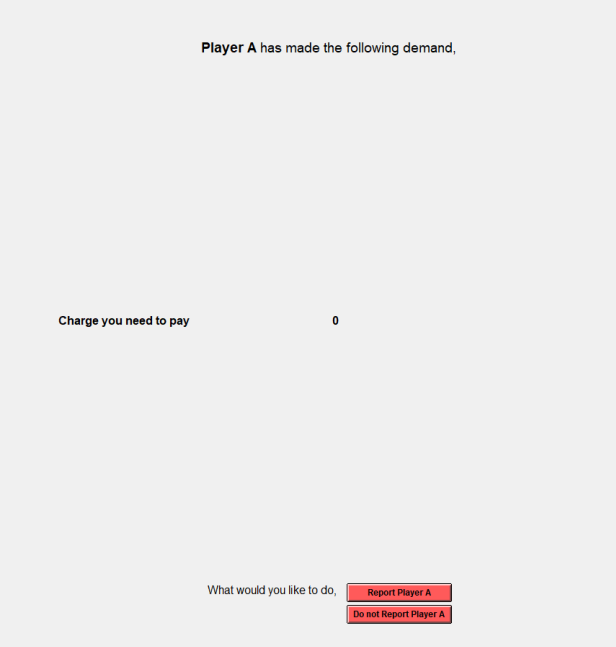
B.3.4 Bribery Game - Player A Decides to Charge or Not



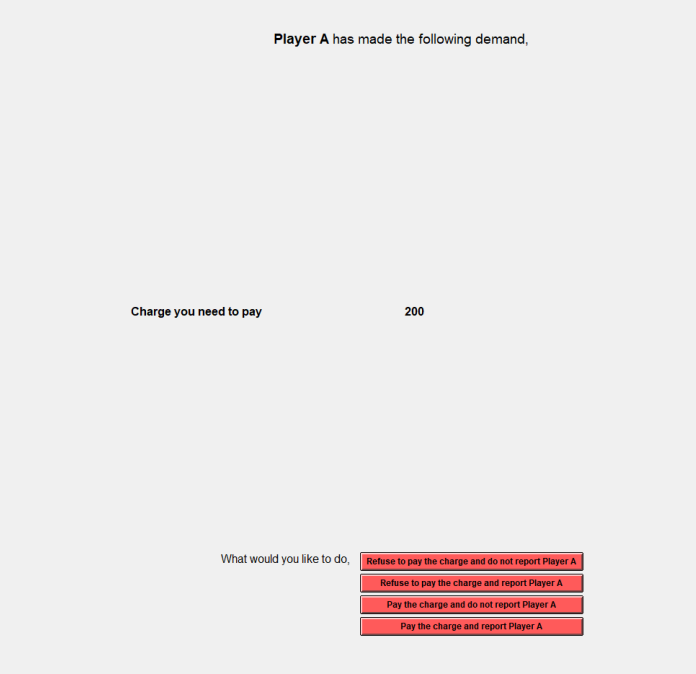
B.3.5 Bribery Game - Player A Chooses the Amount to Charge



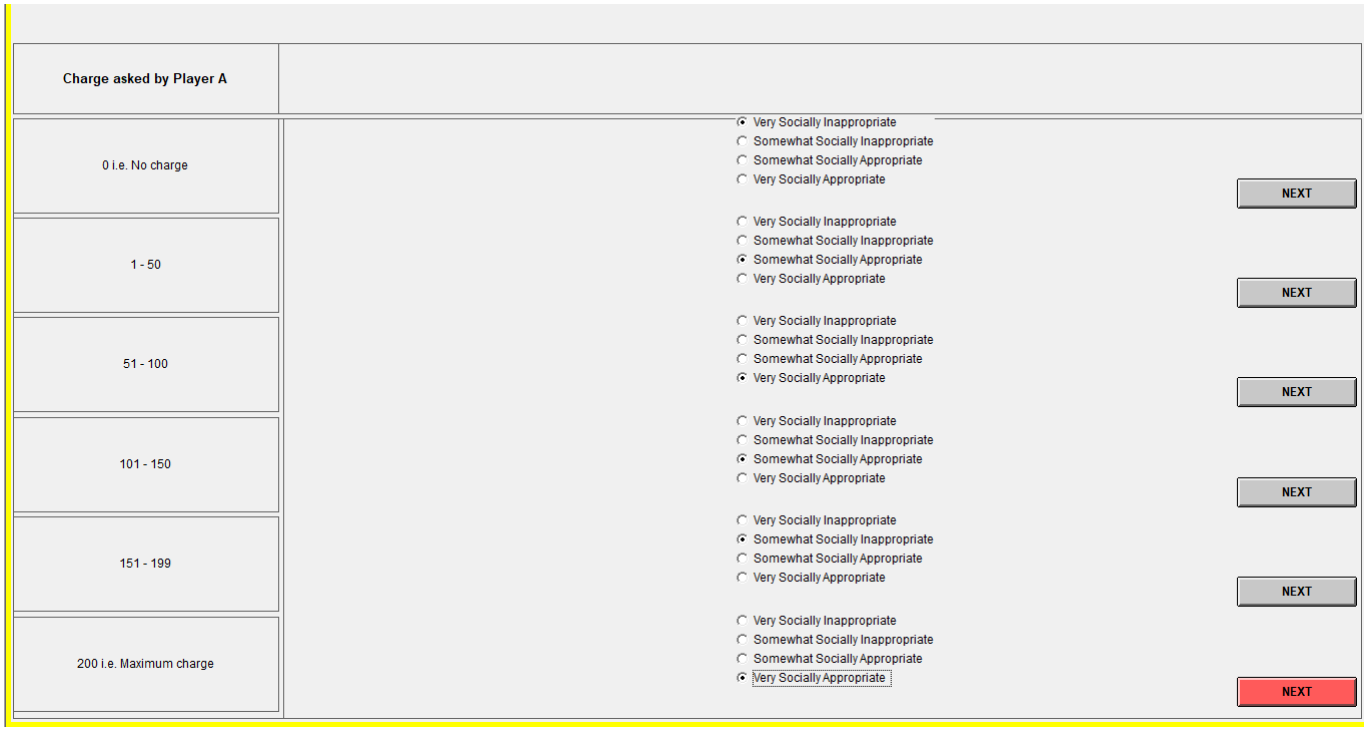
B.3.6 Bribery Game - Player B's Choice When Not Asked for a Charge



B.3.7 Bribery Game - Player B's Choice When Asked for a Charge



B.3.8 Norm Elicitation - Player A's Actions



B.3.9 Norm Elicitation - Player B's Actions When Asked for a Charge

Appendix C

Online Appendix for "Diverse Committees"

C.1 Appendix : Correlation Neglect

In this section, we calculate the decision accuracy and welfare of committees under a correlated information structure but where individuals wrongly believe that information is independent.

Aligned preferences Both homogenous and diverse committees truthfully reveal their signals and then implement the decision with a majority of signals. The probability of a mistake in a homogenous committee is

$$Pr(D \neq S|hcorrneg) = Pr(ppp|Y, hcorr) + 3Pr(ppy|Y, hcorr) \tag{C.1}$$

$$= (1-\sigma)(1 + 3q^2(1-2\sigma)\sigma + 2q^3\sigma(2\sigma - 1)) \tag{C.2}$$

This gives the expected utility of a homogenous committee as

$$EU_{hcorrneg} = (0.5)(1 - Pr(D \neq S|hcorr)) \tag{C.3}$$

The probability of a mistake in a diverse committee is

$$Pr(D \neq S|dcorrneg) = Pr(ppp|Y, dcorr) + Pr(ppy|Y, dcorr) + 2Pr(pyp|Y, dcorr) \tag{C.4}$$

$$= (1-\sigma)(1 + 2q(1-2\sigma)\sigma + q^2\sigma(2\sigma - 1)) \tag{C.5}$$

This gives the expected utility of a homogenous committee as

$$EU_{dcorrneg} = (0.5)(1 - Pr(D \neq S|hcorr)) \tag{C.6}$$

It is straightforward to show that $Pr(D \neq S|dcorrneg) < Pr(D \neq S|hcorrneg)$ and therefore $EU_{dcorrneg} > EU_{hcorrneg}$. That is, even if committee members suffer from correlation neglect, diverse committees still outperform homogenous committees when preferences are aligned. For both homogenous and diverse committees, overall decision accuracy is higher when voters suffer from correlation neglect, but overall welfare is lower. Voters are more likely to match the decision to the state of the world when they ignore the correlated information structure; because $x_1 > 0.5$, this comes at a cost.

Misaligned preferences With misaligned preferences, we have four different committees to consider. In a moderate homogenous committee, decision accuracy and expected utility are as above in the case of aligned preferences. For an extremist homogenous committee, we have already seen that optimal behaviour does not depend on whether information is independent or correlated. Therefore, correlation neglect will not affect these committees.

In a moderate-majority diverse committee, we have seen that committee decisions for each signal realisation are the same under independent or correlated information. Therefore, correlation neglect not affect error rates or expected utility. These are given by Equation C.30 and Equation C.31.

In an extremist-majority diverse committee, we saw in Proposition 4 that correlated information made truth-telling possible, which in turn made these committees the optimal choice in this setting. If voters neglect this correlation, truth-telling is no longer possible - the moderate member babbles. With correlation neglect the overall error rates and expected utility are the same as those for a moderate-majority diverse committee given by Equation C.30 and Equation C.31.

The expected utility for a moderate from both an extremist-majority diverse committee and from a

moderate homogenous committee is lower that would be the case if members fully took account of the correlation structure. Nonetheless, by comparing and Equation C.31, we see it is still the case under correlation neglect that moderates gain a higher expected utility from extremist-majority diverse committees than moderate homogenous committees. In fact, with correlation neglect, any diverse committee will generate higher expected utility for moderates than a moderate majority committee.

C.2 Appendix : Misaligned Preferences with $x_2 < 0.5 < x_1$

C.3 Appendix: Proofs

Proof of Proposition 1

Proof. The expected utility of a committees from choosing $D = P$ or $D = Y$ following a generic collection of signals sss are given by

$$EU(D = P|sss) = 0.5x_1(1 - Pr(Y|sss))$$
$$EU(D = Y|sss) = 0.5(1 - x_1)Pr(Y|sss)$$

A committee will, therefore, choose $D = P$ whenever $Pr(Y|sss) < x_1$, and $D = Y$ otherwise. With aligned preferences and independent information a committee chooses whichever option commands a majority of signals because $0.5 < x_1 < \sigma$. The probability of making a mistake in each state of the world is:

$$Pr(D = P|S = Y, ind) = Pr(D = Y|S = P, ind) = Pr(ppp|Y, ind) + 3Pr(ppy|Y, ind) \quad (C.7)$$
$$= (1 - \sigma)^3 + 3\sigma(1 - \sigma)^2 \quad (C.8)$$

As signals are not correlated within groups, the probability of making a mistake is the same for homogenous and diverse committees. The expected utility of a homogenous or diverse committee is

$$\begin{aligned} EU_{ind} &= (0.5)(1 - x_1)[1 - Pr(D = P|S = Y, ind)] + (0.5)x_1[1 - Pr(D = Y|S = P, ind)] \\ &= (0.5)(1 - (1 - \sigma)^3 - 3\sigma(1 - \sigma)^2) \quad (C.9) \end{aligned}$$

□

Proof of Proposition 2

Proof. In a homogenous committee with correlated information structure, the probability of each committee signal realisation is as follows;

$$\begin{aligned} Pr(yyy|Y, hcorr) = Pr(ppp|P, hcorr) &= [q^3 + 3q^2(1 - q)]\sigma^3 + 3q(1 - q)^2\sigma^2 + (1 - q)^3\sigma \\ Pr(yyy|P, hcorr) = Pr(ppp|Y, hcorr) &= [q^3 + 3q^2(1 - q)](1 - \sigma)^3 + 3q(1 - q)^2(1 - \sigma)^2 + (1 - q)^3(1 - \sigma) \\ Pr(yyp|Y, hcorr) = Pr(ppy|P, hcorr) &= [q^3 + 3q^2(1 - q)]\sigma^2(1 - \sigma) + [(1 - q)^2q]\sigma(1 - \sigma) \\ Pr(yyp|P, hcorr) = Pr(ppy|Y, hcorr) &= [q^3 + 3q^2(1 - q)]\sigma(1 - \sigma)^2 + [(1 - q)^2q]\sigma(1 - \sigma) \end{aligned}$$

From this we can work out the posterior probability of the state being Y given the signal profile.

$$Pr(Y|yyy,hcorr) = Pr(P|ppp,hcorr) = \frac{Pr(yyy|Y,hcorr)Pr(Y)}{Pr(yyy|Y,hcorr)Pr(Y) + Pr(yyy|P,hcorr)Pr(P)}$$
$$= \frac{[q^3 + 3q^2(1-q)]\sigma^3 + 3q(1-q)^2\sigma^2 + (1-q)^3\sigma}{[q^3 + 3q^2(1-q)](\sigma^3 + (1-\sigma)^3) + 3q(1-q)^2(\sigma^2 + (1-\sigma)^2) + (1-q)^3} > \sigma$$

$$Pr(Y|yyp,hcorr) = Pr(P|ppy,hcorr) = \frac{Pr(yyp|Y,hcorr)Pr(Y)}{Pr(yyp|Y,hcorr)Pr(Y) + Pr(yyp|P,hcorr)Pr(P)}$$
$$= \frac{[q^3 + 3q^2(1-q)]\sigma^2(1-\sigma) + [(1-q)^2q]\sigma(1-\sigma)}{[q^3 + 3q^2(1-q) + 2(1-q)^2q]\sigma(1-\sigma)}$$
$$= \frac{1-(2-q)q-(2q-3)q\sigma}{2-q} < \sigma$$

Using the restriction on x_1 in Assumption 1 and the fact that signals are informative, we have $Pr(Y|ppp,hcorr) < Pr(Y|ppy,hcorr) < Pr(Y|yyp,hcorr) < x_1 < \sigma < Pr(Y|yyy,hcorr)$. A homogenous committee chooses $D = Y$ only is all three signals are y, otherwise it implements $D = P$.

The probability of an incorrect group decision in each state is

$$Pr(D = P|S = Y,hcorr) = Pr(ppp|Y,hcorr) + 3[Pr(ppy|Y,hcorr) + Pr(yyp|Y,hcorr)] \tag{C.10}$$
$$= [q^3 + 3q^2(1-q)](1-\sigma)^3 + 3q(1-q)^2(1-\sigma)^2 + (1-q)^3(1-\sigma)$$
$$+[q^3 + 3q^2(1-q) + 2(1-q)^2q]3\sigma(1-\sigma)$$

$$Pr(D = Y|S = P,hcorr) = Pr(yyy|P,hcorr) = [q^3 + 3q^2(1-q)](1-\sigma)^3 + 3q(1-q)^2(1-\sigma)^2 + (1-q)^3(1-\sigma) \tag{C.11}$$

As states are equally likely ex ante, we have

$$Pr(D \neq S|hcorr) = (0.5)Pr(D = P|S = Y,hcorr) + (0.5)Pr(D = Y|S = P,hcorr) \tag{C.12}$$

This gives the expected utility of a homogenous committee as

$$EU_{hcorr} = (0.5)(1-x_1)[1 - Pr(D = P|S = Y,hcorr)] + (0.5)x_1[1 - Pr(D = Y|S = P,hcorr)] \tag{C.13}$$

Next, we analyse behaviour in diverse committees.

Suppose we have a committee with two members from group 1, and one member from group 2.[1] In what follows, the third signal is that of the minority committee member - in this case, the member from group 2. If we see yyp it means that both members of group 1 received y signals while the member of group 2 got a p signal. If, instead, the signal realisation was pyy it would mean that the two members of group 1 received different signals. Recall that preferences are identical for both groups, so any difference between homogenous and diverse groups will stem from the information structure.

In a diverse committee with correlated information structure, the probability of each committee signal realisation is as follows;

$$Pr(yyy|Y,dcorr) = Pr(ppp|P,dcorr) = [q^3 + 3q^2(1-q) + 2q(1-q)^2]\sigma^3 + (q(1-q)^2 + (1-q)^3)\sigma^2$$

$$Pr(ppp|Y,dcorr) = Pr(yyy|P,dcorr) = [q^3 + 3q^2(1-q) + 2q(1-q)^2](1-\sigma)^3 + (q(1-q)^2 + (1-q)^3)(1-\sigma)^2$$

$$Pr(yyp|Y,dcorr) = Pr(ppy|P,dcorr) = [q^3 + 3q^2(1-q) + 2q(1-q)^2](1-\sigma)\sigma^2 + (q(1-q)^2 + (1-q)^3)(1-\sigma)\sigma$$

[1] It is easy to see that two members from group 2 and one from group 1 is equivalent.

$$Pr(ppy|Y,dcorr) = Pr(yyp|P,dcorr) = [q^3+3q^2(1-q)+2q(1-q)^2](1-\sigma)^2\sigma + (q(1-q)^2+(1-q)^3)(1-\sigma)\sigma$$

$$\begin{aligned} Pr(ypy|Y,dcorr) = Pr(pyy|Y,dcorr) &= Pr(pyp|P,dcorr) = Pr(ypp|P,dcorr) \\ &= [q^3+3q^2(1-q)+2q(1-q)^2](1-\sigma)\sigma^2 \end{aligned}$$

$$\begin{aligned} Pr(pyp|Y,dcorr) = Pr(ypp|Y,dcorr) &= Pr(ypy|P,dcorr) = Pr(pyy|P,dcorr) \\ &= [q^3+3q^2(1-q)+2q(1-q)^2](1-\sigma)^2\sigma \end{aligned}$$

As in the case of a homogenous committee, we can work out the probability of the state conditional on a given profile of signal realisations.

$$\begin{aligned} Pr(Y|yyy,dcorr) &= Pr(P|ppp,dcorr) \\ &= \frac{Pr(yyy|Y,dcorr)Pr(Y)}{Pr(yyy|Y,dcorr)Pr(Y)+Pr(yyy|P,dcorr)Pr(P)} \\ &< Pr(Y|yyy,ind) \end{aligned}$$

$$\begin{aligned} Pr(Y|ypy,dcorr) &= Pr(Y|pyy,dcorr) = Pr(P|pyp,dcorr) = Pr(P|ypp,dcorr) \\ &= \frac{Pr(ypy|Y,dcorr)Pr(Y)}{Pr(ypy|Y,dcorr)Pr(Y)+Pr(ypy|Y,dcorr)Pr(P)} \\ &= \sigma \end{aligned}$$

Which simply says that in a committee where the two members from group 1 get different signals, all signals are independent.

$$\begin{aligned} Pr(Y|yyp,dcorr) = Pr(P|ppy,dcorr) &= \frac{Pr(yyp|Y,dcorr)Pr(Y)}{Pr(yyp|Y,dcorr)Pr(Y)+Pr(yyp|P,dcorr)Pr(P)} \\ &= \frac{1-(q-2)q(\sigma-1)}{2-q(2-q)} < Pr(Y|yyp,hcorr) \end{aligned}$$

This says that if a diverse committee gets yyp, they are less likely to think the state is Y than a homogenous committee with the same signals.

Conditional on having two yellow signals and one purple signal, a diverse committee will have a higher posterior on the state being Y than a homogenous committee if the realisation is pyy or ypy, but will have a lower posterior if the realisation is yyp.

Using the restriction on x_1 in Assumption 1 and the fact that signals are informative, we have $Pr(Y|ppp,dcorr) < Pr(Y|pyp,dcorr) = Pr(Y|ypp,dcorr) < Pr(Y|ppy,dcorr) < Pr(Y|yyp,dcorr) < x_1 < \sigma = Pr(Y|ypy,dcorr) = Pr(Y|pyy,dcorr) < Pr(Y|yyy,dcorr)$. A diverse committee chooses $D = Y$ if all three signals are y, or if it has two y signals from members of different groups; otherwise it implements $D = P$.

The probability of an incorrect group decision in each state is:

$$\begin{aligned} Pr(D = P|S = Y, dcorr) &= Pr(ppp|Y, hcorr) + 3Pr(ppy|Y, hcorr) + Pr(yyp|Y, hcorr) \\ &= (1-\sigma)(1+\sigma-(2-q)q\sigma^2) \end{aligned} \tag{C.14}$$

$$\begin{aligned} Pr(D = Y|S = P, dcorr) &= Pr(yyy|P, dcorr) + 2Pr(ypy|P, dcorr) \\ &= (1-\sigma)^2(1+2q\sigma-q^2\sigma) \end{aligned} \tag{C.15}$$

As states are equally likely ex ante, we have

$$Pr(D \neq S|dcorr) = (0.5)Pr(D = P|S = Y, dcorr) + (0.5)Pr(D = Y|S = P, dcorr) \tag{C.16}$$

This gives the expected utility of a diverse committee as

$$EU_{dcorr} = (0.5)(1-x_1)[1 - Pr(D = P|S = Y, dcorr)] + (0.5)x_1[1 - Pr(D = Y|S = P, dcorr)] \tag{C.17}$$

Next, we show that $Pr(D \neq S|dcorr) < Pr(D \neq S|hcorr)$. Subbing Equation C.10 and Equation C.11 into Equation C.12 and simplifying, we get

$$Pr(D \neq S|hcorr) = (2q-3)q^2\sigma^3 + (\frac{9}{2} - 3q)q^2\sigma^2 + (q^3 - \frac{3}{2}q^2 - 1)\sigma + 1 \tag{C.18}$$

Subbing Equation C.14 and Equation C.15 into Equation C.16 and simplifying, we get

$$Pr(D \neq S|dcorr) = (2-q)q\sigma^3 + (\frac{3}{2}q - 3)q\sigma^2 + (q^3 - \frac{3}{2}q^2 - 1)\sigma + 1 \tag{C.19}$$

Equation C.19 is always less than Equation C.18 for values $0 < q < 1$ and $0.5 < \sigma < 1$.

Finally, we show $EU_{dcorr} > EU_{hcorr}$. Subbing Equation C.10 and Equation C.11 into Equation C.13 and simplifying, we get:

$$EU_{hcorr} = \big(\frac{1}{2} - \frac{1}{2}q^3 + \frac{3}{2}q^2(1-x_1) + q(3x_1 - \frac{3}{2})\big)\sigma + \big(\frac{3}{2} + \frac{3}{2}q^2 - 3x_1 + q(\frac{3}{2}x_1 - 3)\big)q\sigma^2 + (\frac{3}{2} - q)q^2\sigma^3 \tag{C.20}$$

Subbing Equation C.14 and Equation C.15 into Equation C.17 and simplifying, we get:

$$EU_{dcorr} = \big(1 - q + \frac{1}{2}q^2\big)x_1\sigma + \big(x_1(q - 1 - \frac{1}{2}q^2) + \frac{1}{2} - q - \frac{1}{2}q^2\big)\sigma^2 + (\frac{1}{2}q - 1)q\sigma^3 \tag{C.21}$$

Equation C.21 is always greater than Equation C.20 for values $0 < q < 1$ and $0.5 < \sigma < 1$. □

Proof of Proposition 3

Proof. A homogenous committee of moderates has the same preferences as any committee in Proposition 1. The probability of a wrong group decision and expected utility are identical to the case of aligned preferences with independent information, given in Equation C.7 and Equation C.9.

A diverse committee can either have two moderate members (from group 1) and one extremist (from group 2); or two extremists and one moderate. We examine each in turn. If signals were all truthfully revealed, the optimal group decision for moderate committee members is to follow the majority of signals. Extremists only favour implementing $D = Y$ when all three signals point in favour of it. Therefore, their preferences are not aligned when the committee gets two y signals and one p signal.

We first examine committees with two moderates and one extremist. As the committee decision is made by majority rule, the decision preferred by moderates prevails. If all signals are truthfully revealed, they implement the decision with a majority of signals.

However, we now show that truth-telling is not an equilibrium in a moderate-majority diverse committee. If messages are believed, an extremist member has an incentive to deviate from truth-telling, instead always reporting p. If the true realisation of signals is two y signals and one p signal where the extremist received an y signal, the moderate members would then believe there are two p signals and one y signal. The committee would implement P, rather than Y - a utility gain for the extremist member. In equilibrium, moderate committee members are not fooled. They ignore the babbling message of the extremist. The most informative equilibrium is where moderates truthfully reveal their signals, while the extremist babbles. With only two reliable signals, the moderate committee members implement Y when both of their signals are y, but otherwise implement P.

The probability of an incorrect group decision in each state is given below where $dmis$ denotes a diverse committee under independent information and misaligned preferences.

$$\begin{aligned} Pr(D = P|S = Y, dmis) &= Pr(ppp|Y, ind) + 3Pr(ppy|Y, ind) + 2Pr(pyy|Y, ind) \\ &= 1 - \sigma^2 \end{aligned} \tag{C.22}$$

$$\begin{aligned} Pr(D = Y|S = P, dmis) &= Pr(yyy|P, ind) + Pr(yyp|P, ind) \\ &= (1 - \sigma)^2 \end{aligned} \tag{C.23}$$

As states are equally likely ex ante, we have

$$Pr(D \neq S|dmis) = 1 - \sigma \tag{C.24}$$

The expected utility of a moderate-majority diverse committee is

$$EU_{dmis} = (0.5)(1 - x_1)[\sigma^2] + (0.5)x_1[1 - (1 - \sigma)^2] = (0.5 - x_1)\sigma^2 + x_1\sigma \tag{C.25}$$

Next, we turn to a committee with two extremists and one moderate. As the committee decision is made by majority rule, the decision preferred by extremists prevails. If all signals are truthfully revealed, they implement P unless all three signals favour Y.

However, we now show that truth-telling is not an equilibrium in an extremist-majority diverse committee. If messages are believed, a moderate member has an incentive to deviate from truth-telling, instead always reporting y. If the true realisation of signals is two y signals and one p signal where the moderate receives an p signal, the extremist members would then believe there are three y signals and no p signals. The committee would implement Y, rather than P - a utility gain for the moderate member. In equilibrium, extremist committee members are not fooled. They ignore the babbling message of the moderate.

The most informative equilibrium is where extremists truthfully reveal their signals, while the moderate babbles. With only two reliable signals, the extremist committee members implement Y when both of their signals are y, but otherwise they implement P. Given this behaviour, both the probability of an incorrect group decision and expected utility are the same as in a moderate-majority diverse committee.

Decision accuracy is higher in moderate homogenous committees than diverse committees whenever

$$(1 - \sigma)^3 + 3\sigma(1 - \sigma)^2 < 1 - \sigma \tag{C.26}$$

which is true for $0.5 < \sigma < 1$

Expected utility for moderate types is higher in moderate homogenous committees than diverse committees whenever

$$(0.5)(1-(1-\sigma)^3-3\sigma(1-\sigma)^2) > (0.5-x_1)\sigma^2 + x_1\sigma \tag{C.27}$$

which is true for $x_1 < \sigma < 1$, guaranteed by Assumption 1.

□

Proof of Proposition 4

Proof. First, notice that a moderate homogenous committee behaves as described in the proof of Proposition 2 - they will share signals truthfully and implement P unless all three signals are y.

By Assumption 2, behaviour in extremist homogenous committees is not affected by the information structure. These committees always implement P unless all three signals are y.

We now turn to diverse committees. If signals were all truthfully revealed, the optimal group decision for moderate committee members is to implement P if it has the majority of signals or if the committee has two y signals and one p signal but the y signals are from members of the same group. If the committee has two or more y signals not from the same group, they implement Y. Extremists only favour implementing Y when all three signals point in favour of it. Therefore, their preferences are not aligned when the committee gets two y signals and one p signal but the y signals are from members of the same group.

We first examine committees with two moderates and one extremist. As the committee decision is made by majority rule, the decision preferred by moderates prevails. If all signals are truthfully revealed, they implement Y whenever there are two of more y signals not from the same group.

However, we now show that truth-telling is not an equilibrium in a moderate-majority diverse committee. If messages are believed, an extremist member has an incentive to deviate from truth-telling, instead always reporting p. If the true realisation of signals is two y signals and one p signal where the extremist received an y signal, the moderate members would then believe there are two p signals and one y signals. The committee would implement P, rather than Y - a utility gain for the extremist member. In equilibrium, moderate committee members are not fooled. They ignore the babbling message of the extremist.

The most informative equilibrium is where moderates truthfully reveal their signals, while the extremist babbles. With only two reliable signals, the moderate committee members implement Y when both of their signals are y, but otherwise, they implement P.

Letting *mdcorrmis* represent a moderate-majority diverse committee when information is correlated and preferences are misaligned, the probability of an incorrect group decision in each state is:

$$\begin{aligned} Pr(D=P|S=Y,mdcorrmis) &= Pr(ppp|Y,dcorr) + 2Pr(ypp|Y,dcorr) \\ &+ Pr(ppy|Y,dcorr) + 2Pr(pyy|Y,dcorr) \\ &= (1-\sigma)(1+(2q-q^2)\sigma) \end{aligned} \tag{C.28}$$

$$\begin{aligned} Pr(D=Y|S=P,mdcorrmis) &= Pr(yyy|P,dcorr) + Pr(yyp|P,dcorr) \\ &= (1-\sigma)(1-2q\sigma+\sigma q^2) \end{aligned} \tag{C.29}$$

As states are equally likely ex ante, we have

$$Pr(D \neq S|mdcorrmis) = 1-\sigma \tag{C.30}$$

The expected utility of a moderate in a moderate-majority diverse committee with correlated information structure is

$$EU_{mdcorrmis} = (0.5)(1-x_1)[1-(1-\sigma)(1+(2q-q^2)\sigma)] + (0.5)x_1[1-(1-\sigma)(1-2q\sigma+\sigma q^2)] \quad \text{(C.31)}$$

We next examine committees with two extremists and one moderate. As the committee decision is made by majority rule, the decision preferred by extremists prevails. If all signals are truthfully revealed, they implement P unless all three signals are y.

We now show that truth-telling is an equilibrium in an extremist-majority diverse committee. As extremists are in the majority, it is clear that they have no incentive to deviate from truth-telling. If he was believed, a moderate could only alter the group decision when the extremists have y signals and the moderate has a p signal. If he instead claims to have an y signal and is believed, the committee will implement Y. However, such a deviation leads to a lower expected utility for the moderate as a moderate prefers to implement Y only if the committee has two or more signals not from the same group. The most informative equilibrium is where all types truthfully reveal their signal and the committee implements Y only after three y signals.

Letting *edcorrmis* represent an extremist-majority diverse committee when information is correlated and preferences are misaligned, the probability of an incorrect group decision in each state is: The probability of an incorrect group decision in each state is:

$$\begin{aligned} Pr(D=P|S=Y,edcorrmis) &= Pr(ppp|Y,dcorr) + 2Pr(ypp|Y,dcorr) \\ &+ Pr(ppy|Y,dcorr) + 2Pr(pyy|Y,dcorr) + Pr(yyp|Y,dcorr) \\ &= (1-\sigma)(1+(1+2q-q^2)\sigma-(2-q)q\sigma^2) \end{aligned} \quad \text{(C.32)}$$

$$\begin{aligned} Pr(D=Y|S=P,edcorrmis) &= Pr(yyy|P,dcorr) \\ &= (1-\sigma)^2(1-2q\sigma+q^2\sigma) \end{aligned} \quad \text{(C.33)}$$

As states are equally likely ex ante, we have

$$Pr(D \neq S|edcorrmis) = 1-\sigma \quad \text{(C.34)}$$

The expected utility of a moderate in a extremist-majority diverse committee with correlated information structure is

$$EU_{edcorrmis} = (0.5)(1-x_1)[1-(1-\sigma)(1+(1+2q-q^2)\sigma-(2-q)q\sigma^2)] + (0.5)x_1[1-(1-\sigma)^2(1-2q\sigma+q^2\sigma)] \quad \text{(C.35)}$$

We see that the overall error rate of $1-\sigma$ is the same in moderate-majority and extremist-majority diverse committees. However, it is higher in state P for moderate-majority diverse committees and higher in state Y for extremist-majority diverse committees. The overall error rate is lower in diverse committees than homogenous committees if Equation C.18 is greater than $1-\sigma$, which is always true when signals $0.5 < \sigma$.

Now we show that the expected utility for moderate types is highest in extremist-majority diverse committees, then in homogenous committees, then in moderate-majority diverse committees. Expected utility for moderate types is lower in moderate homogenous committees than extremist-majority diverse committees whenever Equation C.20 is less than Equation C.35, which is always true for our parameter

ranges.

Expected utility for moderate types is higher in moderate homogenous committees than moderate-majority diverse committees whenever Equation C.20 is greater than Equation C.31. This holds under Assumption 1.

Expected utility for moderate types is higher in extremist-majority committees than moderate-majority diverse committees whenever Equation C.31 is less than Equation C.35. This holds whenever $x_1 > 0.5$. □

C.4 Experimental Instructions

Sample Instructions AL-CORR Welcome and thanks for participating at this experiment. Please read these instructions carefully. They are identical for all the participants with whom you will interact during this experiment.

If you have any questions please raise your hand. One of the experimenters will come to you and answer your questions. From now on communication with other participants is not allowed. Please do also switch off your mobile phone at this moment. If you do not conform to these rules we are sorry to have to exclude you from the experiment

For your participation you will receive 5 pounds. During the experiment you can earn more. How much depends on your choices as well as those of others and is explained below. All your decisions will be treated confidentially.

THE EXPERIMENT At the beginning of the experiment, you will be asked to answer a few short questions. Afterwards you will be randomly assigned to one of two groups either group either group S or group T. Your group will be displayed to you in the upper right corner of the screen. The experiment consists of 12 rounds. Each round has the following stages.

Stages of the Experiment

1. At the beginning of each round an urn will be chosen. The urn will be either YELLOW or PURPLE. The purple urn has 7 purple and 3 yellow balls and the yellow urn has 7 yellow and 3 purple balls. Both urns are equally likely to be chosen. You will not know which urn was chosen. The picture below illustrates the two urns.

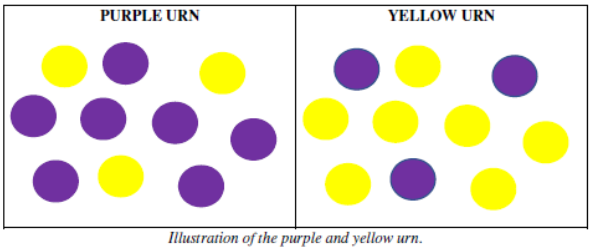


Illustration of the purple and yellow urn.

2. Afterwards balls are drawn from the chosen urn.
   - First one ball is drawn for group S, which we call the S-ball. Then the ball is returned to the urn.
   - Second one ball is drawn for group T, which we call the T-ball. Then the ball is returned to the urn.
3. You will see one ball from the chosen urn.
   - If you are in group S you will see with
     - 50% chance the S-ball and with
     - 50% chance a ball drawn for you from the urn
   - If you are in group T you will see with
     - 50% chance the T-ball and with
     - 50% chance a ball drawn for you from the urn
4. Committees will be randomly formed. Committees will consist of three committee members. You will know the group (S or T) of your committee members but otherwise committee members will be anonymous.

5. You will then be able to chat for 45 seconds with your committee members about which urn you believe was chosen.

6. Afterwards all committee members simultaneously submit a guess on whether the urn is yellow or purple.

7. Each committee member then gets the following payoff

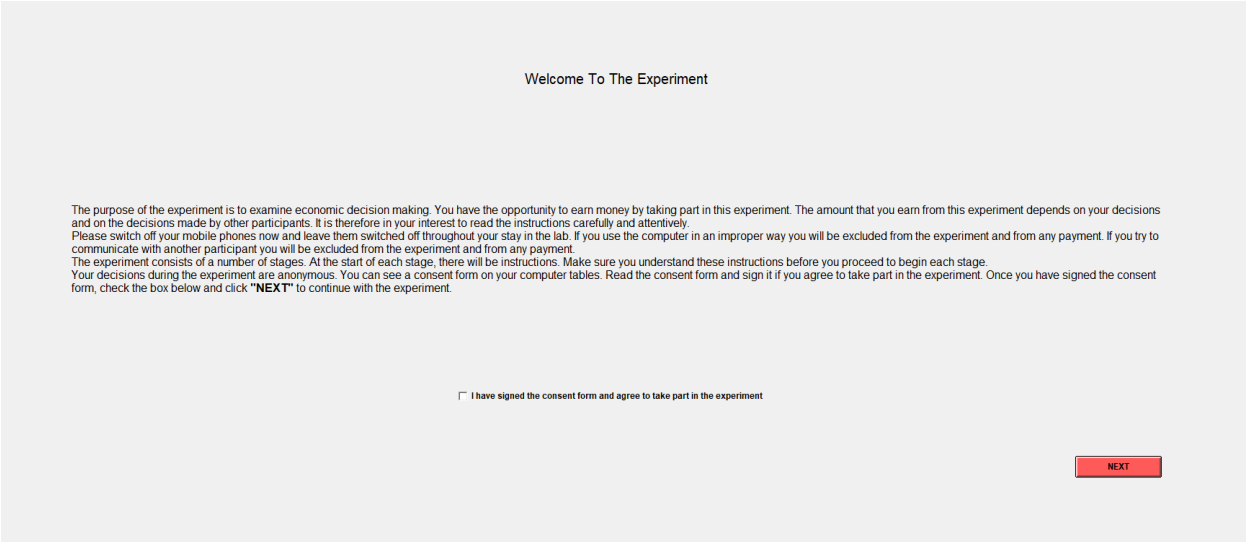
   - If the majority of committee members, i.e. at least two out of the three committee members guess correctly that the urn is purple then a committee member receives 10 GBP if they are a member of group S and 12 GBP if they are a member of group T.
   - If the majority of committee members, i.e. at least two out of the three committee members guess correctly that the urn is yellow then a committee member receives 5 GBP if they are a member of group S and 3 GBP if they are a member of group T.
   - If the majority of committee members, i.e. at least two out of the three committee members do NOT guess correctly, then each committee member receives 0 GBP.

This process repeats for 12 rounds.

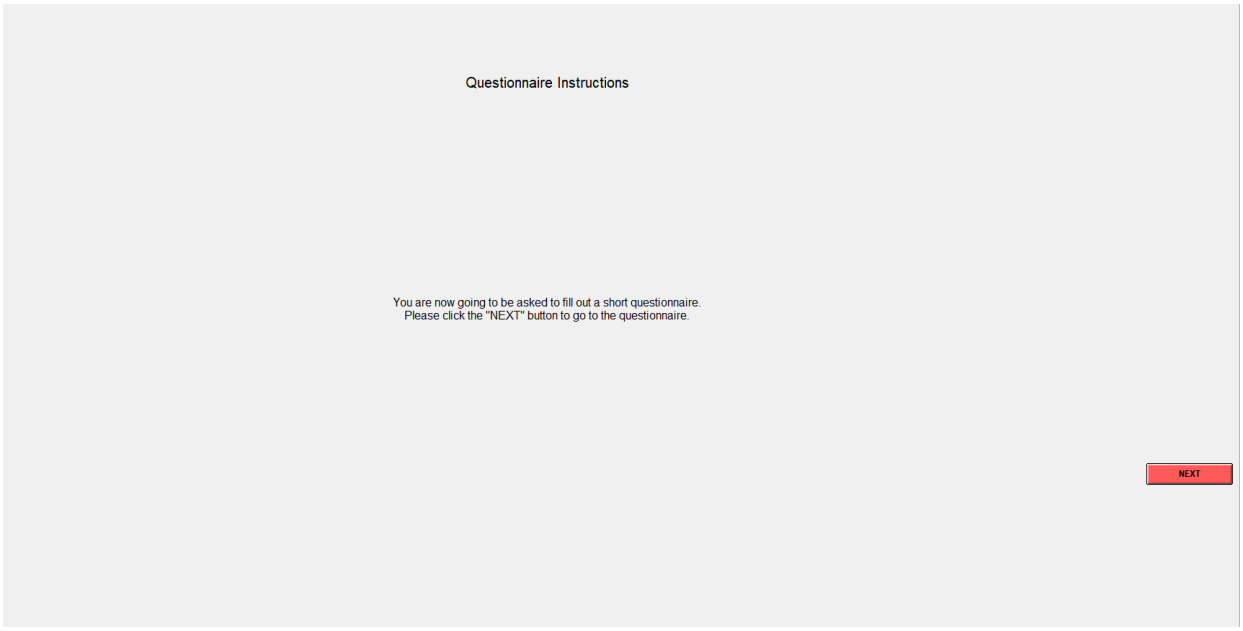
Payment At the end of the experiment we will draw one round at random. You will receive the payment from this round. In addition, all participants receive 5 GBP as a show up fee and 3 GBP for answering the questions in a post-experimental questionnaire.
Questions? If you have any questions please raise your hand.
Enjoy the Experiment!

C.5 Screenshots

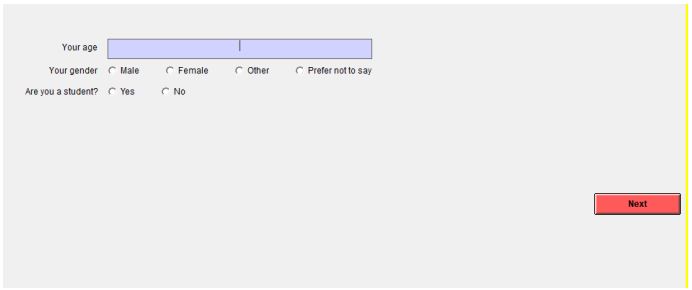
C.5.1 Welcome



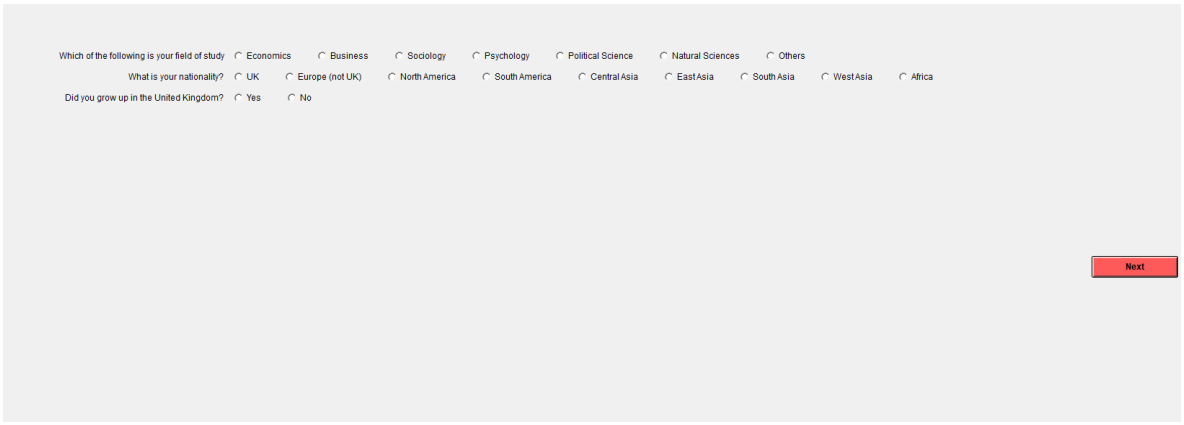
C.5.2 Questionnaire Instructions



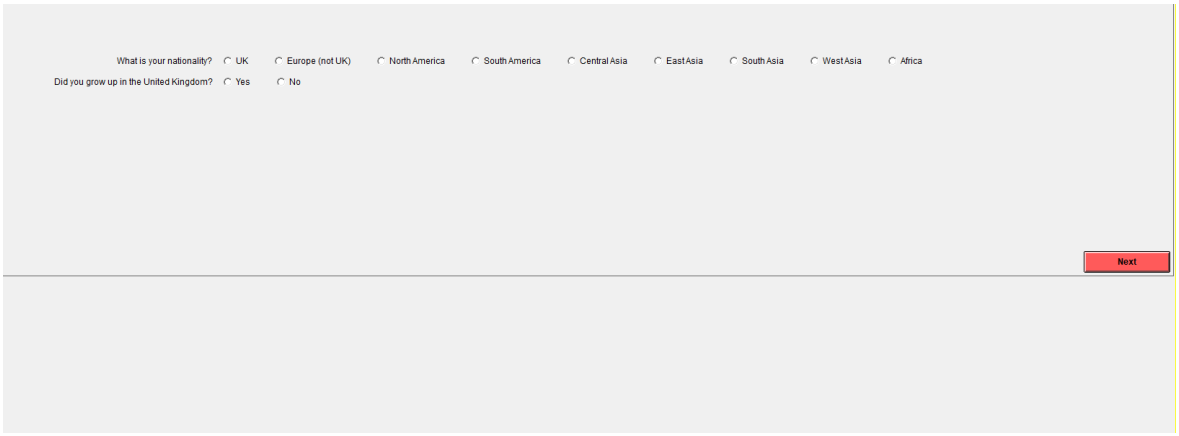
C.5.3 Questions



C.5.4 Questions - continued

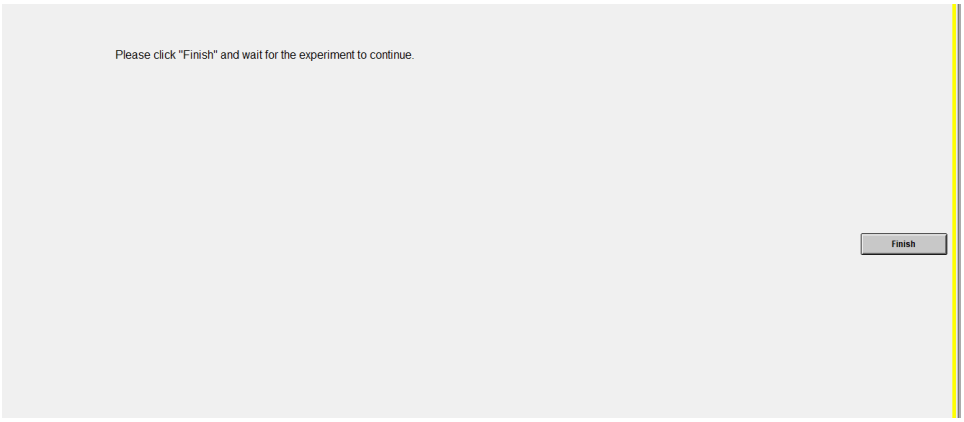


C.5.5 Questions - continued

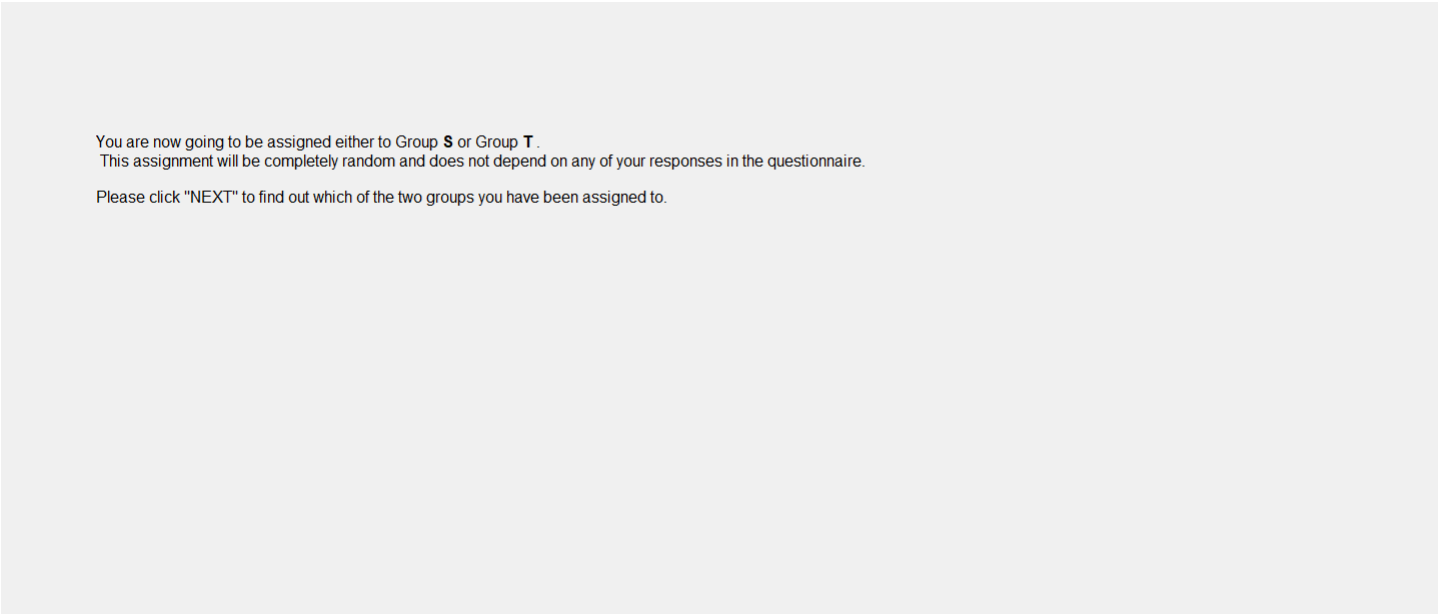


C.5.6 Questions - continued

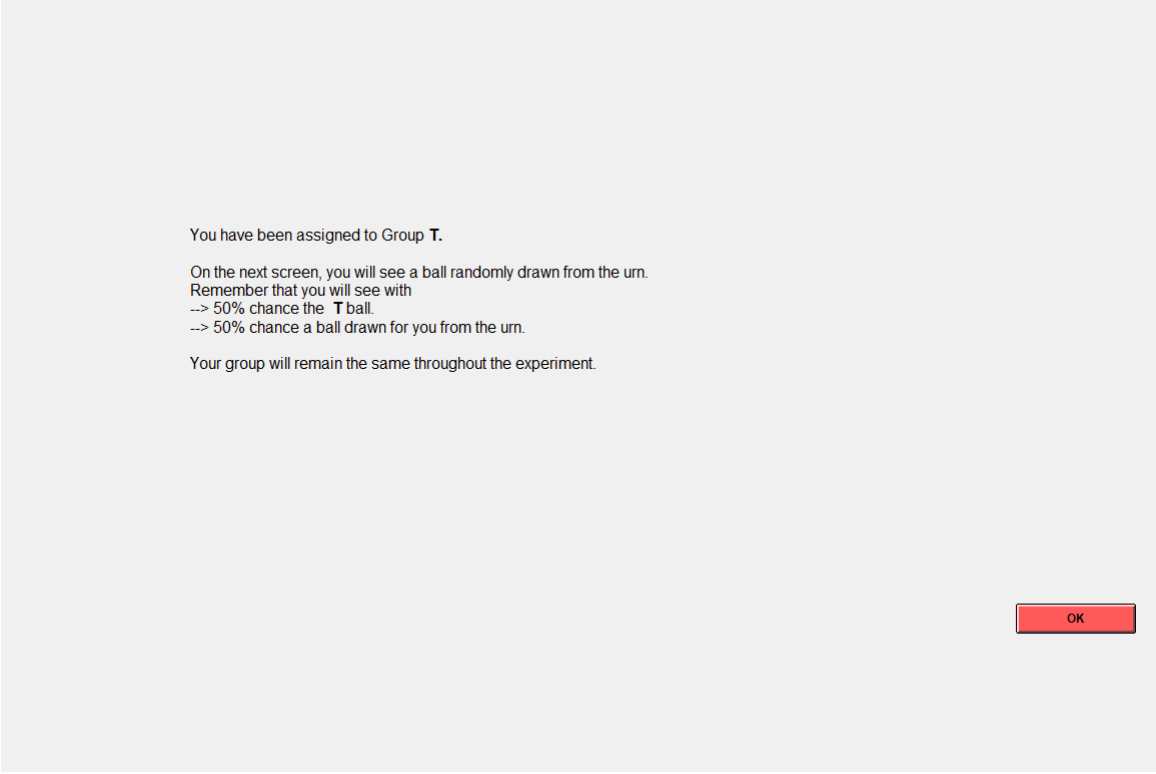
C.5.7 Questions - continued



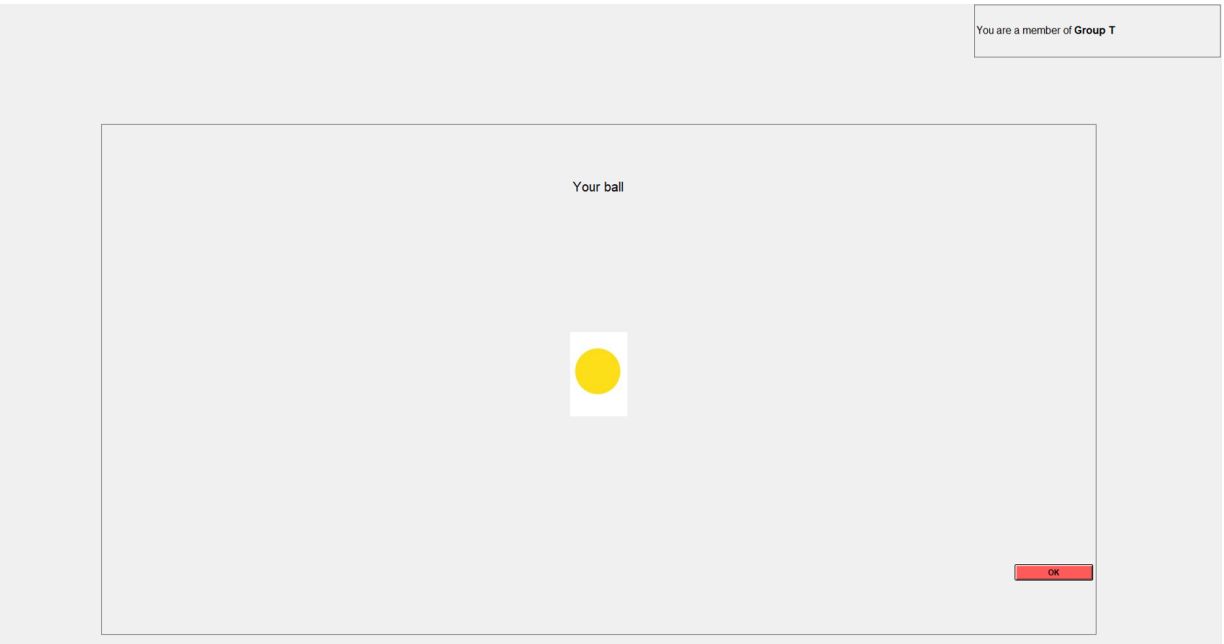
C.5.8 Group Selection



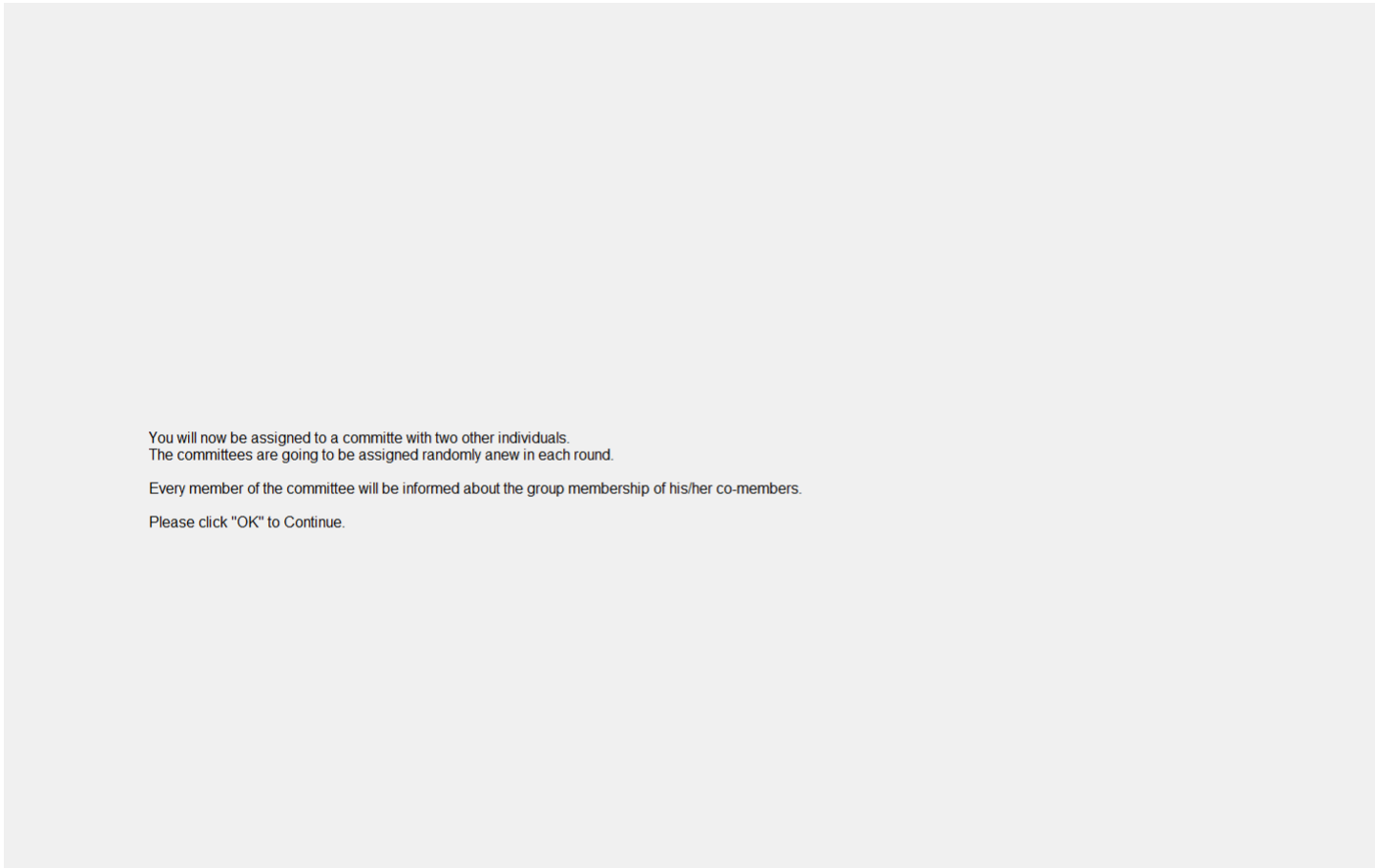
C.5.9 Group announcement



C.5.10 Reveal Signal



C.5.11 Committee Selection



C.5.12 Committee Announcement

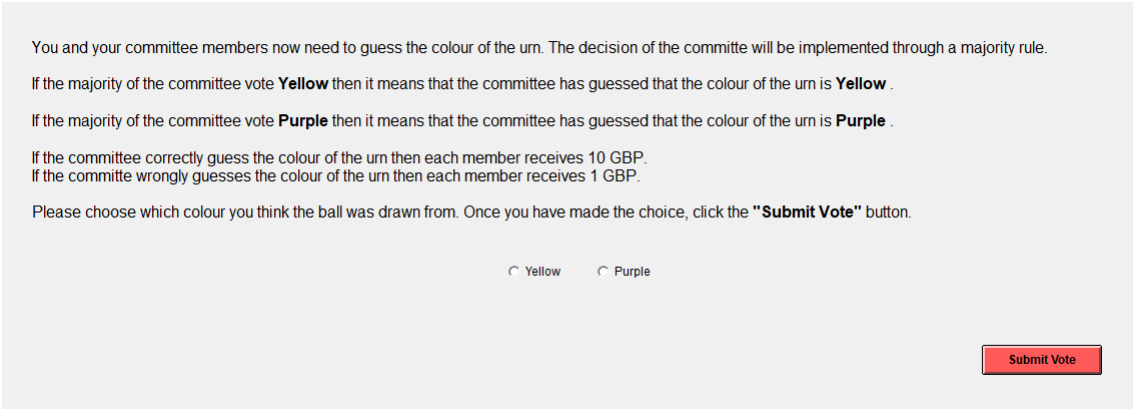
You have been assigned to a committee of 3 individuals. You can find below your group membership and that of the two other committee members.

Your Group	**T**
Group of Committe member 1	**T**
Group of Committee member 2	**S**

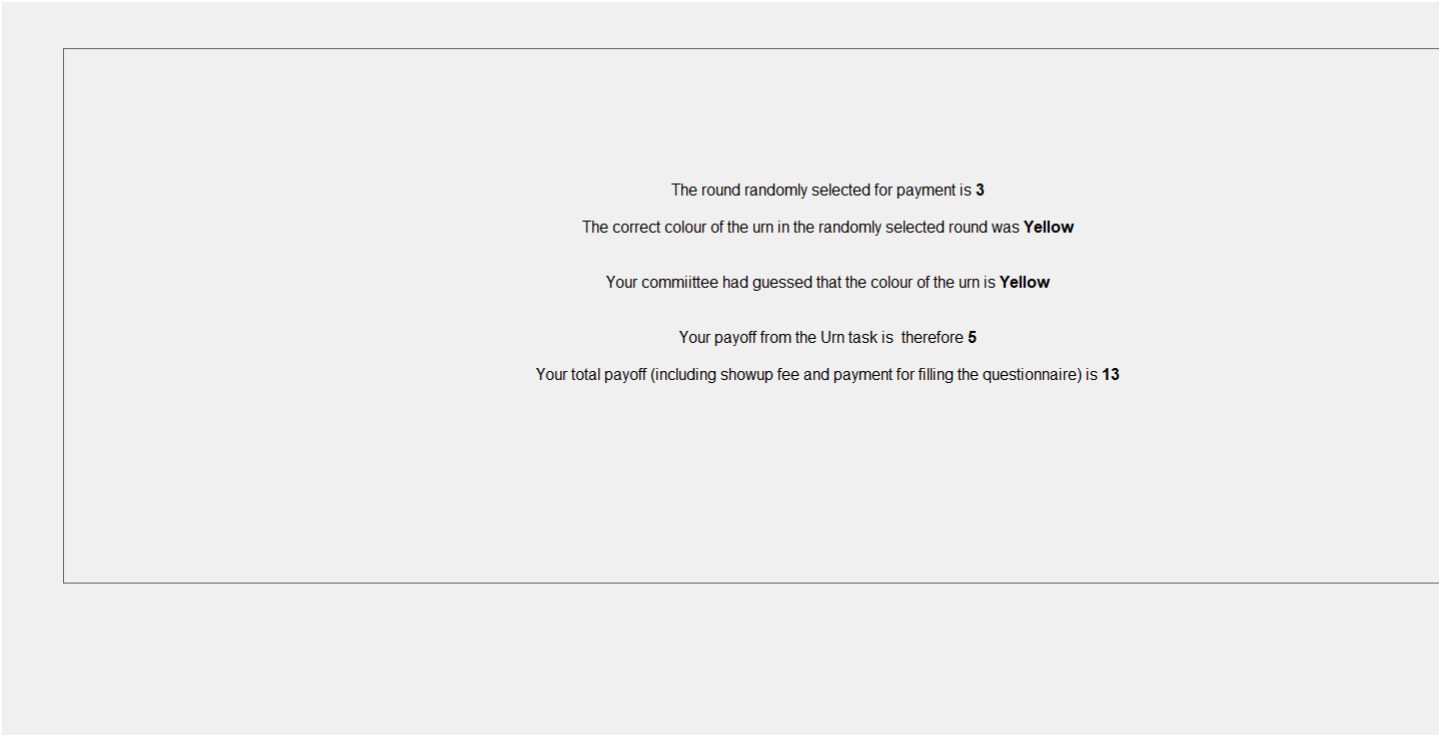
You will now be able to chat with your committee members about which colour of the urn the committee members should guess.

Please click "OK" to continue.

C.5.13 Voting Stage



C.5.14 Results Announced



C.6 Additional Tables

	Independent				*Correlated*			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
HET	0.0170	0.0158	0.0173	0.0184	0.0530**	0.0522**	0.0513**	0.0525**
	(0.0377)	(0.0363)	(0.0358)	(0.0357)	(0.0243)	(0.0256)	(0.0262)	(0.0263)
Constant	0.749***	0.713***	0.574***	0.526***	0.537***	0.516***	0.494***	0.410***
	(0.0263)	(0.0506)	(0.0639)	(0.104)	(0.0198)	(0.0599)	(0.0602)	(0.0724)
Observations	2,700	2,700	2,700	2,580	2,700	2,700	2,700	2,616
Number of id	225	225	225	215	225	225	225	218

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table C.6.1: **Individually correct decisions. Aligned Preferences.** session fixed effects; columns 2-4 also committee fixed effects; column 3 controls for age gender student dummy; column 4 additionally for nationality, risk aversion, confidence and individual measure of correlation neglect.

	Independent				*Correlated*			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	yellow		purple		yellow		purple	
HET_EXT	-0.232**	-0.231**	-0.0728	-0.0732	-0.0313	-0.0311	0.0102	0.00771
	(0.0945)	(0.0957)	(0.0728)	(0.0724)	(0.0913)	(0.0914)	(0.0694)	(0.0716)
HET_MOD	-0.108***	-0.108**	0.00452	0.00669	0.0155	0.0143	0.0264	0.0281
	(0.0377)	(0.0384)	(0.0650)	(0.0632)	(0.0699)	(0.0696)	(0.0619)	(0.0617)
HOM_EXT	-0.213**	-0.213**	-0.0204	-0.0183	-0.120	-0.121	0.00831	0.0115
	(0.0875)	(0.0877)	(0.0888)	(0.0882)	(0.123)	(0.123)	(0.0933)	(0.0925)
Constant	0.798***	0.793***	0.821***	0.791***	0.617***	0.597***	0.736***	0.794***
	(0.0624)	(0.0925)	(0.0677)	(0.0628)	(0.0821)	(0.109)	(0.0702)	(0.0985)
Observations	414	414	462	462	472	472	416	416
R-squared	0.047	0.047	0.026	0.028	0.056	0.056	0.030	0.036

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table C.6.2: Percentage of correct decisions by state.

C.7 Additional Figures

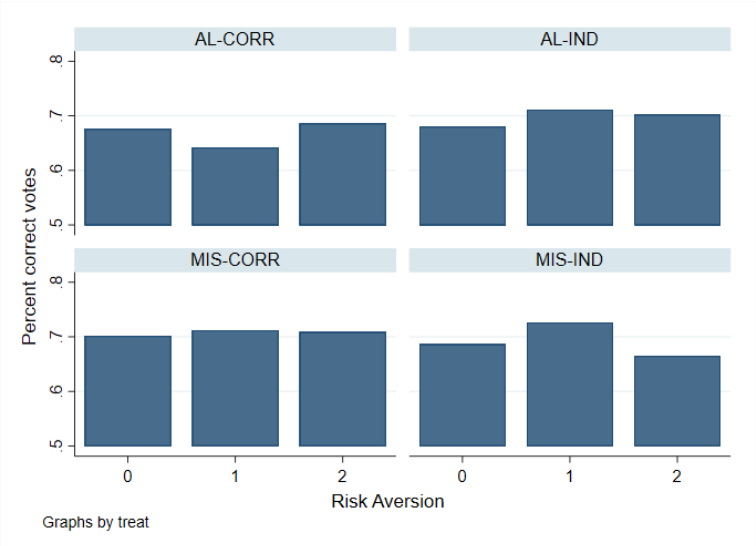


Figure C.7.1: Percent of correct votes (votes matching the state) depending on degree of risk aversion expressed at the end of the experiment.

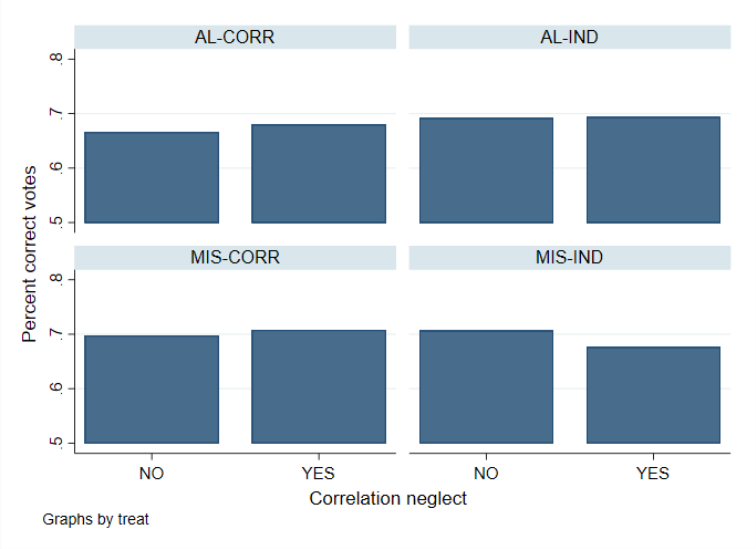


Figure C.7.2: Percent of correct votes (votes matching the state) depending on measure of correlation neglect.