Experiments in norms, cooperation and culture

Shruti Surachita

Thesis submitted to the University of Nottingham for the Degree of Doctor of Philosophy

March 2022

ABSTRACT

Norms are informal rules shaped and transmitted by social learning that determine aggregate behavioural patterns in society. This could range from simple rules such as those associated with social etiquette to the currently relevant ones related to the Covid-19 pandemic. This thesis contributes to the literature exploring social norms, particularly the role of such consideration in social dilemma problems, issues with its measurement and investigating the underlying mechanisms driving these patterns. The thesis includes three independent research studies that use a combination of online and laboratory experiments investigating these issues. Chapter 1 briefly introduces the research objectives, discusses the methods used and summarises the overall contributions from all the 3 studies. Chapter 2 investigates the influence of varying degrees of saliency of a cooperative norm on contribution choices as well as the effect of (induced) cooperation as a function of initial saliency on the perceptions of a norm of cooperation. We find that saliency improves cooperation only in round 1 in a repeated public goods game, and cooperation still decays with time. Additionally, norm of cooperation remains stable, irrespective of the local experience in the game. Chapter 3 tests the effectiveness of two preference elicitation methods in a simple rule-following experiment. There is evidence of preference conditionality with people's rule following behavioural decisions, but limited experimental methods exist to measure such preferences. We find that a variant of the strategy method (Gächter et al., 2021) captures such preferences accurately and is also useful in eliciting preference heterogeneity in individuals with regards to their adherence towards rule compliance. Chapter 4 studies the social beliefs and peoples' conformity attitudes in two countries namely Sweden and Turkey that vary in their degree of individualism. We find that the individualistic cultural orientation with efficient formal institution in Sweden helps shape clear normative message of rule-compliance, which is not the case in Turkey. Chapter 5 concludes with a brief overview of the contribution and scope for future research.

ACKNOWLEDGEMENTS

I would first like to thank my supervisors, Professor Simon Gächter and Professor Daniele Nosenzo for their helpful guidance and invaluable feedback throughout my time of study.

Additionally, I am also very grateful to the members of CeDEx for their useful comments, criticisms, and suggestions at various stages of this research project, which helped me improve the manuscript from time to time. Moreover, I am also very thankful to Dr. Jose Guinot Saporta, the Lab Manager at CeDEx, who helped me undertake the experimental sessions needed for the various studies of this project.

I would also like to express my gratitude to the School of Economics for their financial support which provided a full PhD studentship for the length of my research.

Finally, I would like to thank my friends and colleagues who have provided me support during my research training. Last but not the least, I would like to thank my family for their love and encouragement which helped me throughout the study.

Table of Contents

List of Figures

List of Tables

CHAPTER 1 Introduction

Norms are the informal unwritten rules of society. They have been studied extensively in the social sciences, particularly in philosophy, sociology, anthropology, law, political science, and economics. According to the Oxford dictionary, a norm is a "standard or pattern, especially of social behaviour". However, the term "norm" has more than one meaning (Shaffer 1983, cited in Cialdini et al., 1990, p.1015). It comprises of an *"is"* component that reflects what most people do; and an *"ought"* component that prescribes what most people approve of being done. Elster (1989) distinguished social norms from instrumental rationality, suggesting that such simple informal rules were insensitive to changing times, and not necessarily future-oriented unlike outcomes dictated by pure rationality. These were based on a shared understanding and approval, and non-compliance triggered strong emotions such as shame, embarrassment, and guilt. He also differentiated social norms from moral and legal norms, in the sense that the former was imposed by the greater community and were not always a result of selfish self-interest e.g., cooperation and fairness norms. In fact, certain norms can be very useful in society such as norms against smoking in public spaces and anti-littering norms. When norms get internalised, people follow them even if violations are unobserved by others.

One of the most prominent contributions to norm literature are by Cialdini et al. (1990) who conducted a series of field experiments catered towards anti-littering behaviour. Their "focus theory" suggested that although people's actions are governed by norms, these considerations are not always active in people's minds. By activating these normative considerations, one can guide behaviour in a norm-congruent manner. Normative explanations have also been considered in explaining cooperation outcomes in public good experiments in economics. Ostrom (2000) enlisted a few observations from social dilemma experiments where rational

theory had previously suggested free riding to be the equilibrium solution if all people are self-interested. First, studies showed that most subjects contribute 40-60% of their endowments in one-shot and first round in finitely repeated games. Second, there was a gradual contribution decay however, due to the presence of 'conditional cooperators', and 'willing punishers' along with rational egoists. Third, cooperation sustained with face-to-face communication and other contextual factors. All these findings in conjunction with evolutionary theories indicated that one of the possible rationales for collective action could be explained through people's preference for conformity with social norms.

Norm driven behaviour is inter-dependent on others' actions and opinions. They are driven by an interplay of social beliefs and people's preference for conformity. We briefly summarize some important characteristic features of norm compliance as well as their measurement techniques in experimental literature which will be consistently investigated and discussed in the 3 different studies in Chapters 2-4.

First, norm compliance relies on people's conformity attitudes. Conformity refers to the "act of changing one's behaviour to match the responses of others" (Cialdini & Goldstein, 2004, p. 606). People tend to follow the crowd rather than taking independent actions, especially in a world of uncertainty. Keynes (1930) pointed out that the motivation to follow others stems from one's own ignorance, assuming that other people are better informed. Additionally, he had mentioned that it seemed that people valued being conventionally wrong than unconventionally right, and thus there is safety in numbers. This rationale provides a possible motivation for descriptive norms where aggregate behavioural patterns are based on what others around one do, i.e., actions led by descriptive (or empirical) information. For example,

if people notice a littered environment implying that most others had littered, people tend to litter more themselves (e.g., Cialdini et al., 1990; Keizer et al., 2008).

Additionally, people also are motivated to follow some rules that are guided by beliefs regarding what constitutes socially appropriate and approved conduct. In other words, they care about others' approval, are guided by a desire to "fit in" their social groups and avoid "punishment" by group members or formal institutions for rule violations. It is important to point out here that others or group members refer to the relevant reference groups that individuals belong to and associate the norm with. For instance, stealing is prohibited within the group by Favela dwellers in Brazil but carries no moral repercussions when carried outside the group (Bicchieri, 2016). When decisions are conditional on both what others do (empirical/ descriptive expectations), and approve of being done (normative/injunctive expectations), they are considered as social norms[1].

A follow-up feature is that norm-existence does not promise compliance. A norm exists in the presence of normative expectations i.e., what others think is the socially right thing to do, and people's preference for conformity. However, if empirical and normative expectations are not met, norms will not be obeyed. A typical example used to illustrate this idea is the case of corruption. The act of corruption is always considered as socially inappropriate. However, in some countries people widely act according to these normative considerations, in which case the normative standard of corruption becomes internalised and easy to follow even without regulation. On the other hand, in some other countries, although corruption is perceived as socially inappropriate, people still participate in corrupt activities and so there is a contradiction

[1] Throughout the thesis, we use the terms empirical/descriptive and normative/injunctive as per the context. For instance, in Chapter 2, we write empirical and normative beliefs (which seems more intuitive than injunctive beliefs) and descriptive and injunctive norms.

in what people do and think is the right thing to do. In such cases with inconsistency between empirical and normative expectations, people get easily motivated by the actions of others which further spreads corruption and erodes the appropriate standard of behaviour. This reflects that's norm compliance requires both empirical and normative expectations to be consistent with each other. In case of incongruent beliefs, it is the empirical actions of others that drives behavioural decisions.

Throughout the thesis we focus predominantly on Bicchieri's (2005; 2016) social norm model. She states "a norm is a behavioural rule that 1. Is known to exist and apply to a class of situations, 2. Is followed by individuals in a population on condition that: a) it is believed that sufficiently many others follow it (empirical expectations), and b) it is believed that sufficiently many others believe the rule should be followed, and/or may be willing to sanction deviations from it (normative expectations)" (cited from Bicchieri, 2016, pp, 65-66). This is a very useful framework since it captures the necessary elements of social beliefs and people's conformity preferences for compliance outcome. Additionally, these factors can be measured using experimental tools with belief and preference elicitation methods, and thus offer much potential to experimentalists interested in measuring norms in the lab.

We now briefly review developments in norm-measurement techniques in experimental economics in recent years. One area of experimental research has looked into the normative belief-elicitation protocol. Krupka and Weber (2013) norm elicitation is one of the most used experimental methods to measure second-order normative beliefs of appropriate or inappropriate social behaviour. Here belief-elicitation requires subjects to respond to a decision environment based on what they think "most others would think is the socially right thing to do". Hence, responses are based on a socially coordinated focal point, just as an equilibrium

solution of a coordination game. In a different approach, Bicchieri and Chavez (2010) measure normative beliefs using a two-step protocol. First, they elicit responses regarding a privately held normative belief to determine how individuals feel about a particular action personally. They also elicit a normative expectation regarding what individuals think is the socially appropriate action.

A different set of research investigates individuals' norm-following sensitivities. Norms are simple informal rules, and individuals and societies hold heterogeneous attitudes towards compliance with such rules (e.g., Bicchieri, 2016; Fehr & Schurtenberger, 2018; Kimbrough & Vostroknutov, 2016; Kimbrough & Vostroknutov, 2018). In their study, Kimbrough and Vostroknutov (2016) devised a simple rule-following task (RF task), following the concept of waiting at a traffic light. In this RF task, subjects either wait or move at five separate traffic stops. Each stop would be red at first but would turn green after 5 seconds of the subject reaching the stop. To avoid the context problem that could arise in this case, in another recent study, they used a different rule-following task where the subjects were given 100 balls, which they could drop in either a blue or a yellow bucket. Dropping the ball in the yellow bucket would give more points than dropping the ball in the blue bucket. However, the rule was to "put the ball in the blue bucket." Thus, following the rule would yield fewer points. These tasks are devised to measure individuals' propensity to follow or violate rules.

This brief introduction leads to the revelation of some important unanswered questions that demand more research and deeper understanding of norms and their underlying mechanisms as well as their measurement methods. We enlist three research questions and then discuss how our studies investigate these issues and then briefly summarize our findings.

1.1 Research Problem 1: The role of norm salience for cooperation

Behavioural regularities in public good experiments imply the role of cooperative norms influencing collective outcomes. Experimental evidence shows that people are willing to "punish" (even at a cost to oneself) those who do not actively cooperate and free ride on others. Such punishment motivations are usually driven through emotional triggers of anger, disapproval, and resentment towards such non-cooperators. Thus, such evidence indicates that normative considerations play a role in shaping cooperation outcomes. This leads to the first research objective where we are interested to investigate the role of such considerations in collective group decisions, particularly in social dilemma problems.

As pointed by Cialdini et al. (1990), these normative considerations may not always be active in people's minds. Even if they are, they might vary in their degree of saliency. Sometimes no such thoughts are considered, while on other occasions, people think about them, but they may be limited to their private thoughts only. For instance, some environmentally conscious people might think about the negative repercussions of single use plastics on the environment before buying a cup of coffee but may not necessarily be aware of others' thoughts of such actions. In other cases, there is the saliency of a shared collective normative thought that actively influences decision-making. For example, in the current Covid-19 pandemic, people discuss rules of wearing face masks and maintaining social distance so pro-actively, that these considerations likely stay very prominent in peoples' minds, and also inform people of a shared consensus of what constitutes socially appropriate behaviour.

Norm-saliency refers to the degree to which such normative thoughts are activated prior to making a decision. Although there are many studies that look into the effects of norms on cooperation, there is limited understanding as to how varying degrees of saliency of a

cooperative norm can influence collective action. Moreover, there is sufficient evidence in public good experiments to state that cooperation decays over a period of time. A follow-up question then arises: does this experience of norm erosion influence the perception of a normative standard of cooperation?

Chapter 2 is "Norm saliency in cooperation". In our study, we exogenously manipulate the saliency of a cooperation norm from none to private thoughts, to shared norms prior to deciding in a social dilemma problem. We elicit cooperative norms both prior to a three-person public good (PGG) experiment played over 20 rounds, with belief elicitation and feedback, and also post-play to evaluate normative evaluations after the local experience in the game. We use Krupka and Weber (2013) norm elicitation method in the context of a one-shot PGG as our decision-environment. We carry out two sets of experiments, Study 1 (laboratory experiment) and Study 2 (online replication), both with student samples, recruited from the University of Nottingham.

Our main results are as follows: first, making a norm salient has only a short-term effect on cooperation. This effect is mediated through an improved expectation of group members' contribution (empirical beliefs), that raises cooperation in round 1. However, saliency of norms does not affect cooperation outcomes in the long run, whereby all groups experience cooperation decay over the 20 rounds. Finally, the experience with declining cooperation is associated with group members' perceptions of cooperation norms. We find that those groups that perform better in the 20 rounds compared to others, elicit a harsher normative rating for low and no contribution and a stronger positive rating for full cooperation. Thus, saliency of a shared social norm has a transient effect on cooperation outcomes and normative perceptions differ across subjects with varying experience of cooperation.

1.2 Research Problem 2: Methods to measure conditional rule compliance

Although some methodological advancements have been made in norm-measurement in experimental economics, these remain limited. Norm compliance is based on conformity. People reflect a preference to follow such informal rules, conditional on what others do and approve of being done. These considerations have been studied in many experimental games that have validated the causal effect of such social influences on people's decision-making. Recently, Gächter et al. (2021) carried out experiments to test for such preference conditionality with peer effects and found that compliance is indeed driven by what peers do and what peers think is the socially appropriate action. They find that when peers violate, subjects are more likely to violate. Moreover, if peers disapprove of a violation, subjects are also less likely to violate. Both results reflect people's conformity attitudes conditional on what others do (empirical information) and what other disprove of being done (normative information).

However, this experiment incorporated a between-subject design, with 28 treatment groups that varied in group size and social information. Additionally, they used a "live treatment" where subjects observed the actions of their peer(s), in real time in the decision-making environment. Both of these factors make the design cumbersome to carry out in experimental setups and call for the need to elicit such conditional preferences in a more efficient manner.

Chapter 3 is on "Methods in Rule Following". This is a methodological study where the aim is to test different methods to elicit people's conformity preferences with others' actions. We use a simple rule following task adapted from Gächter et al. (2021) as our decision environment for evaluating rule-compliance outcomes. We test the effectiveness of two different

experimental methods in the measurement of preference elicitation. The two methods are a) a variant of the strategy method (S-method) adopted from Gächter et al. (2021) and b) a variant of a direct elicitation (D-method) adapted from Abeler et al. (2019). Both methods are well-organized to be implemented in experimental studies and survey applications compared to the real-time task used in Gächter et al (2021) with peer effects. The S-method is economical and has been used extensively in preference elicitation methods. The D-method circumvents the issues of the experimental demand effects that may be an issue with the S-method.

We conduct online experiments, by recruiting subjects using Amazon MTurk with US samples. For both S and D methods, we ask subjects to elicit a response between following and violating a rule in four scenarios that differ in the fraction of others' (previous MTurkers) who had violated the rule. The aim is to check if varying proportion of others' violation outcomes, an important factor driving conformity, affect subjects' decisions to violate.

We indeed find evidence of such conditionality in the S-method: there is an increased proportion of (elicited) violation preference with the (empirical) information of an increased rate in others' rule violation outcome. This reflects that people tend to violate rules, by imitating others, especially when they observe a majority of others (more than 50%) violating the rule. Thus, the S-method is useful in eliciting subjects' conformity preferences. Additionally, this method is also beneficial in suggesting "types" of decision-makers namely, the unconditional rule followers and rule violators and the conditional rule-followers (i.e., the majority type) as well as their degree of norm-sensitivity with compliance outcomes.

However, the findings from the D-method indicate that although there is an influence of varying proportion of others' violation, subjects elicit a lowered preference to violate

conditional on such information. We speculate normative considerations to play a role in driving this result: this method incorporates graphical representation of varying fractions of rule followers and violators, following which subjects are asked to think of the possible rationale for following or breaking the rule. This thinking may have activated their normative thoughts associated with rule following that might have played a role in dissociating themselves from violation decisions.

1.3 Research Problem 3: Cultural differences in rule following

It is established by now that preferences are conditional on social beliefs, and that the interplay of such social influences and conditional preferences drives aggregate norm or rule compliance outcomes. Studies in cross-cultural research investigating rule violation have found evidence of a systematic pattern to this outcome (e.g., Licht et al., 2007; Gächter & Schulz, 2016). More specifically, countries that have an individualistic and egalitarian social structure with efficient formal institutions display higher rule compliance than countries with a collectivist and hierarchical structure and weak formal setups. Additionally, studies in cross-cultural research have indicated the relevance of in-group relationships that affect the need for approval and avoidance of disapproval. However, the existing literature lacks a comprehensive understanding of how social structures inform social influences that shape aggregate compliance outcomes.

The title of Chapter 4 is "Cultural orientation informs rule compliance". This involves a three-part experiment, which tests elements from Bicchieri's social norm framework in two culturally distinct countries that differ in their country's level of individualism (IDV) scale. We choose IDV, since this is the basis of social structures and defines the role of social groups

(informal) and formal institutions, both of which have a role in shaping people's conformity attitudes, social expectations, and the degree of rule compliance in countries.

We conduct a series of cross-cultural online experiments using subjects recruited through Qualtrics, from a low IDV country, Turkey and a high IDV country Sweden. We first evaluate the degree of rule compliance using a simple rule-following task, the abstract task, adopted from Gächter et al. (2021). Our results fall in line with the hypothesis that rule-following is lower in low IDV Turkey compared to Sweden, a high IDV. After establishing this result, we run different set of experiments eliciting subjects' empirical and normative expectations regarding rule compliance using direct belief elicitation and Krupka Weber (2013) norm elicitation methods. We also elicit subjects' conformity attitudes with regards to empirical and normative information in both countries using a variant of the strategy method adapted from Gächter et al (2021).[2]

We find that Swedish subjects have more accurate beliefs than their Turkish counterparts regarding the degree of rule-compliance in their country. Additionally, we find that subjects from both countries elicit a similar normative standard of rule-compliance but differ in their perceptions of the social (in) appropriateness of rule-violation. Turkish subjects display no shared consensus on the appropriateness or inappropriateness of a rule violation, whereas Swedish subjects elicit a strong shared perception of the social inappropriateness of violation. We also find that subjects display a preference to violate conditional on an increasing rate of rule-violation (empirical information) in both countries, with a slightly stronger conditionality in Turkey. On the other hand, we do not find evidence of any such preference conditionality

[2] This decision was based on our results from Chapter 3, where we find that the strategy method is useful in eliciting conditional preferences.

with normative information. In other words, subjects from both countries display conformity with others' actions, but not with others' disapproval towards violation.

All these observations lead us to conclude that the relevance of group thinking over individualistic thinking, the importance of informal social groups and the presence of weak formal institutions in low IDV(s), lead to ambiguity in framing clear normative messages regarding the social appropriateness of rule following behaviour. As a result, people are easily motivated to violate rules, by imitating others and this further erodes aggregate compliance outcome. Thus, social structures with their underlying belief systems shape people's (conditional) preferences towards rule or norm compliance.

To summarize, this thesis contains three independent research studies incorporated in Chapters 2-4. Each study contains an appendix with additional analyses. All our experiments are programmed using LIONESS Lab (Giamattei et al., 2020), a free web-based platform where experimenters can create, share, and run their experiments. The software has been developed at the Centre for Decision Research and Experimental Economics (University of Nottingham, UK) and the Chair of Economic Theory (University of Passau, Germany). LIONESS provides an experimental URL link for each experiment. This is primarily useful for running experiments online. We use the software for running laboratory-based experiments too.

We use a combination of both online and laboratory experiments for data-collection for these studies. In the last couple of years, online experimentation has become an attractive option for behavioural and cross-cultural experiments since data-collection is cheaper and faster than the traditional laboratory experiments. In addition, it provides the potential for collecting data from a representative sample of the general population. Recent studies show no significant

difference in the behavioural outcomes of subjects when recruited using online samples (e.g., Arechar et al., 2018; Dandurand et al., 2008; Horton et al., 2011). We use online samples recruited through Amazon MTurk and Qualtrics. We conduct both laboratory and online experiments with student samples recruited through ORSEE, from the University of Nottingham, UK.

Chapter 5 briefly provides the conclusion and contribution of each chapter along with scope of future research.

CHAPTER 2 Norm Saliency in Cooperation

2.1 Abstract

Normative messaging happens all the time in reality, but the saliency of such messages differs across situations. Norm-saliency refers to the extent to which normative considerations are activated before making a decision: from none at all, to private thoughts, to shared social norms. However, not much evidence exists to show how varying levels of norm-saliency causally influence cooperation in a social dilemma problem. In this study, we manipulate the degree of saliency of a cooperation norm using the Krupka-Weber norm-elicitation method and observe the subsequent behaviour in a three-person public good game with belief elicitation and feedback. We also measure the effect of this (induced) behaviour on norms of cooperation. We find that saliency of norms, through a shared normative belief, influences the initial empirical belief of group members' contribution thereby leading to an immediate improved cooperation. However, this effect dissipates over time. Moreover, we find no differences in the elicited norm of cooperation, post play with varying exposure of saliency manipulation. However, groups that performed better in terms of average cooperation in the 20 rounds, irrespective of the initial saliency, elicited a stronger norm of cooperation.

2.2 Introduction

Normative considerations are relevant for behaviour in social-dilemma problems (Fehr & Fischbacher, 2004). Thoughts of what *"ought to be done"* or what *"is the right thing to do"* may influence behaviour depending on the extent to which they are activated prior to making decisions. They might not be activated at all, or they may be activated but limited to one's own thoughts or they might be salient through active discussions of acceptable standards of behaviour in people's social circles or in society at large. For instance, the Covid-19 pandemic illustrates the situation where rules of behaviour such as wearing a facemask, maintaining social distance, staying at home, and protecting lives are made very salient such that people are not only aware of the negative repercussions of non-compliance themselves, but also know what most people around them agree is the socially acceptable thing to do. Saliency of norms differs across situations and time. Analysing the role of norm saliency for cooperation is the main goal of this chapter.

Descriptive (empirical) and injunctive (normative) information to increase saliency of normative considerations has been used extensively in various studies across the social sciences to make people act in a norm-congruent manner. For instance, various forms of norm-nudges across a wide range of laboratory experiments have shown that saliency of normative information, as well as norm-enforcement mechanisms, from expressing mild disapproval to ostracism of group members often improve pro-social behaviour, at least to some extent (e.g., Capraro et al., 2019; Cinyabuguma et al., 2005; Dal Bó & Dal Bó, 2014; Ellingsen & Johannesson, 2008; Krupka & Weber, 2009; Maier-Rigaud et al., 2010; Masclet et al., 2003; Noussair & Tucker, 2005; Peeters & Vorsatz, 2013; Rege & Telle, 2004; Xiao & Houser, 2009). Norm-nudges have also been used in many field experiments by incorporating either descriptive or injunctive norms to improve behaviour such as in the case of tax evasion (Bott

et al., 2020; Hallsworth et al., 2017), energy demand (Allcott & Rogers, 2014; Fornara et al., 2016; Ito et al., 2018; Pellerano et al., 2017), environmental concerns (Demarque et al., 2015; Goldstein et al., 2008; Hamann et al., 2015; Reno et al., 1993) and credit card repayment (Bursztyn et al., 2019).

Reminding people of the existing norms can help in improving behavioural outcomes. Looking at the vast majority of the relevant literature, we find multiple evidence of the usage of norm saliency through personal, moral messaging with varying levels of non-monetary sanctions from slight disapproval to group expulsion to induce and increase pro-sociality, both in the lab (e.g., Masclet et al., 2003) and in the field (e.g., Allcott & Rogers, 2014). However, in a social dilemma problem, making a norm salient at the individual level may not be strong enough a nudge, since behavioural outcomes are inter-dependent among group members involved in the dilemma, and thus norm saliency may need to be heightened to increase common knowledge of the norm at the group level to lead to a collective change. But there is not much known on how varying salience of a cooperative norm, and a shared normative belief introduced prior to cooperation choices can influence cooperation.

Moreover, as Bicchieri and Chavez (2010) suggested, norm saliency and the presence of the right empirical and normative beliefs lead to conformity. However, limited evidence exists to see causality between the two kinds of beliefs and its effect on compliance. Additionally, research on social dilemma problems have established that cooperation among group members declines over repeated interactions due to preference heterogeneity within members. Although some literature show how varying degree of norm-adherence affects its normative evaluation in other settings, especially dictator and one-shot prisoner dilemma games (e.g., López-Pérez

and Vorsatz, 2010), there is not much evidence in the context of a repeated public goods game. This chapter provides novel evidence on this question.

Our three main research objectives are as follows. First, we make use of varying degree of norm saliency to find the necessary requirement to improve inter-dependent cooperation decisions in a 3-person public good game (PGG), played over 20 rounds. Next, we investigate how saliency of an injunctive norm of cooperation influences the subsequent empirical belief of cooperation, and if such beliefs improve cooperation. Specifically, we look into both short and long-term dynamics of the inter-dependencies of such beliefs and the subsequent cooperation outcomes. Finally, we aim to find how varying experience of declining cooperation, as a function of the initial saliency may be associated with the subsequent normative standard of cooperation.

To investigate these questions, we exogenously manipulate the level of saliency of a cooperative norm, using the Krupka-Weber norm elicitation (KW henceforth) in a PGG. We incorporate three main treatments: NO, which acts as our baseline treatment, where the cooperative norm is not salient; LOW, where a cooperative norm is made salient to the individual group members but kept private and not communicated to anyone; and HIGH, where not only is the norm made salient to each member, but it is also shared with group-members to induce a shared normative belief. This is followed by a 20 round PGG in fixed groups of three with belief elicitation and feedback after every round. This allows us to observe both transient and long-term effects of varying degree of norm saliency on group members' beliefs and contribution choices. At the end of the 20 rounds, we again elicit the cooperation norm to see if the experience of the game was associated with the groups' evaluations of normative standard

of cooperation. We run the study both in the lab (Study 1) and online (Study 2), to test for the replicability of the experiment and robustness of the results.

We find that HIGH improved both round 1 beliefs and contribution choices in both studies, while LOW improved these outcomes only in Study 1. HIGH improved average group beliefs across the 20 rounds in Study 1 only and had no effect on average group contribution in both studies. On the other hand, LOW had no effect on average group beliefs and contribution decisions in both Study 1 and 2. Moreover, neither treatment affected cooperation decay. Additionally, varying degree of norm-saliency had no influence on post-play norms. However, groups experiencing high(er) cooperation irrespective of the initial saliency, elicit a stronger norm of cooperation after their experience in the repeated interaction.

We make two observations about cooperation outcomes: first, norm saliency of an injunctive cooperative norm through a shared normative belief (HIGH), led to a transient effect on cooperation, specifically through improved empirical cooperative beliefs of group members. In both studies, we find that it is the presence of a shared normative belief which is necessary and sufficient to improve initial group cooperation, and not just the saliency of the cooperation norm at the individual level. Secondly, this effect nonetheless is short-lived. People have heterogeneous cooperative preferences, and so the effect of saliency of such an injunctive norm dissipated over time due to the presence of possible heterogeneous cooperative preferences (e.g., Fischbacher & Gächter, 2010). The empirical information of what such group members actually contributed guided behavioural outcomes that led to gradual cooperation decay across the 20 rounds.

We also find the following results regarding post-play cooperation norms. The experience of declining cooperation was associated with the subsequent evaluation of the strength of a normative standard of cooperation. Although we find a consensus on the content of the norm post-play indicating the presence of an entrenched meta norm of cooperation, there is a weakened (strengthened) norm of cooperation in groups experiencing low (high) level of cooperation in the 20 rounds. Specifically, we find that this shift is in the strength of norm with regards to normative evaluation of no and low contribution choices, i.e., groups performing better in the 20 rounds, although experience a gradual cooperation decay, rate the contribution choices of 0, 1 and 2 tokens (0-40% contribution of the total endowment value) more harshly than groups who witnessed low contribution outcomes over the repeated rounds. Additionally, we find evidence of positive correlation between the high performing groups and their normative evaluation for full cooperation (contributing 5 tokens).

On the other hand, groups that fail to cooperate well, also eventually evaluate no and low contribution outcomes as slightly more "socially appropriate". This could be explained either by ways of self-serving bias or updated pessimistic belief of norm of cooperation due the frequent display of such "bad behaviour" by fellow group members, while groups that exhibit better cooperation outcomes, rate such actions as slightly more "socially inappropriate". This shows how norm-erosion by ways of declining cooperation widens the gap in the perception of the normative standard of cooperation among subjects with heterogeneous cooperation preferences, thereby affecting the strength of a norm of cooperation. To summarise post-play norms, groups experiencing higher cooperation in the 20 rounds, irrespective of the initial degree of norm-saliency, provided a harsher judgement for zero or low contribution choices, and improved their evaluation of full cooperation, thereby eliciting a stronger norm of cooperation.

We make a few important contributions. First, we show how norm-saliency through a shared normative belief acts as an important "norm-nudge" for inducing an immediate collective change in a social-dilemma problem. Secondly, we provide a deeper understanding of how saliency of a normative standard leads to belief shift of other's action, consequently improving compliance with a cooperation outcome. Third, we reveal that varying degree of saliency of normative consideration has no long-term effect on cooperative outcomes. Finally, we show how normative strength of cooperation is shaped by subjects' experiences within their group.

The organization of this chapter is as follows: Section 2.3 discusses the related literature, followed by Section 2.4 which describes the experimental design. Section 2.5 presents the results; Section 2.6 discusses the results and concludes.

2.3 Related literature and our contribution

Our study investigates the effect of varying degree of saliency on behavioural outcomes of cooperation in 20 rounds of PGG and also the influence of cooperative experiences on subsequent norm evaluations. In subsection 2.3.1, we discuss some of the relevant literature on how saliency improves pro-sociality, followed by some key papers related to our study in particular. In 2.3.2, we then briefly review the norm-literature that discuss how normative and empirical beliefs interact, and other studies that have elicited norms post-play in various other experimental games, followed by section 2.3.3 which discusses our contribution.

2.3.1 Norm-saliency and Pro-social Behaviour

Although normative considerations guide behaviour, they are not always activated in people's minds. One of the earliest contributions was Cialdini et al. (1991), who suggested that making

such considerations salient, i.e., activating the empirical and normative information of what most others do or approve of doing based on moral standards of acceptable behaviour, may influence pro-sociality. They conducted a series of field experiments whereby a confederate littered in a clean environment and also in an already littered environment, which acted as norm-saliency cues. One of the most interesting findings was that subjects littered less in a clean environment and even less in an environment with only 1 piece of litter, since both these environments activated anti-littering norms. On the other hand, subjects' littering progressively increased as they witnessed more litter in the environment, highlighting pro-littering norms. Reno et al. (1993) also found that saliency of an injunctive norm of littering reduced subjects littering behaviour irrespective of the state of the environment.

Such norm nudges have been used with other framing effects and similar observations have been shown in various other laboratory experiments (e.g., Cinyabuguma et al., 2005; Ellingsen & Johannesson, 2008; Maier-Rigaud et al., 2010; Peeters & Vorsatz, 2013; Rege & Telle, 2004; Xiao & Houser, 2009). Direct and indirect elicitation of moral and social expectations have influenced subjects' actions in one-shot as well as repeated games. Moral norms also express the "right thing to do" based on moral values and personal beliefs but are unconditioned from societal expectations. For instance, Dal Bó and Dal Bó (2014) found that moral nudges introduced in the middle of a repeated two-person PGG had a positive but temporary effect on cooperation where cooperation increased right after the moral message but declined thereafter. Shank et al. (2019) made subjects play a 31-period PGG with 6 rounds of communication involving gossip and injunctive norm manipulations and found that subjects' contribution to the group account spiked after each communication round and declined thereafter.[3]Bicchieri

[3] Shank et al. (2019) used concrete injunction statements such as: "We should keep a lot to ourselves/ we should contribute a lot to the group" and abstract injunction statement: "We should focus completely on being individuals/ we should focus completely on being one group". This primarily reflects the tension between individual

and Chavez (2010) used similar mechanisms in the context of ultimatum games in the lab. Peysakhovich and Rand (2016) created cultures of cooperation and defection in the lab and found that subjects from cooperative cultures, implying a higher moral standard, took part in more pro-social choices, had higher beliefs and were willing to punish violators more, compared to subjects from defective cultures[4].

Sometimes, norm-saliency is introduced through approval (disapproval) by ways of public observability, which is also shown to improve pro-sociality. For instance, Rege and Telle (2004) found that display of social approval (or disapproval) through direct observability of action led to increased contribution in a one-shot PGG. Additionally, communication of direct approval using written messages to extreme forms of disapproval through ostracism has also helped improve behaviour in lab experiments. There is established evidence of the positive effect of communication on cooperation (see Balliet, 2009 for a review).

For instance, early evidence was provided by Masclet et al. (2003) who used non-monetary sanctions in the form of disapproval points after each round in a repeated PGG and found that such display of disapproval for non-cooperative actions, improved cooperation in the first few rounds. Simpson et al. (2017) found that inter-personal moral judgements after every round in a repeated PGG, helped sustain higher cooperation levels and improved trust between group members compared to the baseline treatment. In other studies, Xiao and Houser (2009) as well as Ellingsen and Johannesson (2008) found that written messages expressing disapproval, induced dictators to offer equal spilt offers in dictator games. On the other hand, Cinyabuguma

selfishness and group cooperation, via an implied norm of cooperation as perceived by the group members and not like the KW elicitation which helps capture a societal normative standard of cooperation.

[4] They created these environments in the lab using infinitely repeated prisoner dilemma games with differences in continuation probability and payoffs.

et al. (2005) and Maier-Rigaud et al. (2010) found that stronger informal sanctions such as expulsion also helped increase and retain cooperation in public good experiments.

There are a few other papers that are more closely related to our work. Krupka and Weber (2009) studied the focusing and informational effects of norms in a non-strategic one-shot binary dictator game. More specifically, their focusing treatment made subjects think of what others do, or think ought to be done, while the informational treatment provided the information of what four other subjects actually did in the dictator game. All treatments led to improved pro-social behaviour even when sometimes the empirical belief and information encompassed selfish behaviour. Norm saliency was introduced through either an empirical or a normative belief in the focusing treatment, and an actual empirical information in the informational treatment. Our study is similar with the idea of the elicitation of a normative belief before decision making (in LOW), and feedback regarding empirical information, as is with the feedback of group members' contribution after each round. LOW also helps with round 1 beliefs and contribution in Study 1, consistent with the results from this study. Our design, however, also helps us to study group-dynamics over 20 rounds in PGG, while theirs' used a one-shot DG game.

In a different study, Schram and Charness (2015) invoked injunctive norms of fair giving in the lab with advice and observed the effects on dictator shares. They found that with advice and observability, dictator-giving was less selfish. It should be noted that since this advice was created through a detailed discussion among three group-members, they suggested that such advice or norms reflected a shared understanding among participants and reinforced behaviour through the fear of being disapproved or shamed. The advice reflected an "ought to do" component of a norm of giving, based on a shared consensus of group members, similar to

HIGH in our experiment. However, our study differs from Schram and Charness (2015) in that ours involves a within-subject treatment, where the same groups take part in both norm-elicitation and behavioural outcomes. This is useful to investigate how the group members' own normative thoughts of cooperation displayed before the game, influenced their group's cooperation, as opposed to spectators' advice from their study.

Finally, in a recent study, Capraro et al. (2019) found that nudging subjects by asking questions related to personal and societal normative beliefs, before making choices, increased people's pro-social behaviour: saliency effects improved cooperation in the one-shot prisoner's dilemma (PD) game. Our experiment differs from this study in some aspects: first, we elicit a norm of cooperation using a coordination device following the Krupka-Weber (2013) method, where belief elicitation is incentivised to match the true standard of a societal norm of cooperation; second, we observe behaviour in a 20 round PGG to observe saliency effects over time; third, in our choice set of the PGG, subjects had the option to make 6 contribution choices, from 0-5 tokens, as opposed to a binary outcome of a PD. Most importantly, we were interested to see whether the effect of our treatment manipulation affected cooperation by itself, or if it was moderated through empirical beliefs of group members' contribution.

2.3.2 Normative evaluations of behavioural outcomes

Compliance with norms differs across people based on their intrinsic norm sensitivity (Bicchieri, 2016; Fehr & Schurtenberger, 2018; Kimbrough & Vostroknutov, 2016) and this may affect their perception of the social appropriateness of the norm. For instance, Kimbrough and Vostroknutov (2018) employed a rule-following task to measure rule-compliance and followed it up with dictator giving and elicitation of normative standard of such giving. They found that although the content of the norm remained stable between groups of rule followers

and rule breakers (equal spilt very appropriate action), the strength of the norm varied between these groups. Likewise, López-Pérez and Vorsatz (2010) observed that cooperators provided stronger feedback in terms of approval/disapproval of co-player's prior decision through informal message than defectors in a prisoner's dilemma game. Disapproval of free riding in their case was a reflection of subjects own personal beliefs. Our design differs from this study in the sense that we elicit norms using the Krupka-Weber (2013) methodology where the subjects' responses indicate their perception of the "social appropriateness" or "social inappropriateness" of the possible contribution choices, i.e., what they think "most others think is the socially right thing to do".

A different channel of literature proposes that people's exposure to varying degrees of norm compliance may also affect the strength of the norm in context. In a recent study, Bicchieri et al. (2020) informed subjects about a high incidence of honest behaviour from a die-under-cup task, and subsequently elicited the normative beliefs in one condition, while provided the normative standard of lying and elicited the empirical belief of lying behaviour. They found that honest behaviour also reflects a strong disapproval towards lying, but not vice-versa. In another study, Bicchieri et al. (2022) elicited personal normative beliefs and expectations before and after feedback from a different Take or Give dictator game with information of varying degrees of violations and found that information with frequent violations led to belief updating of normative expectations and weakened the injunctive strength of the norm. Such belief updating due to frequent violations lowers the normative standard of "compliance", thereby making violations more acceptable with time.

In a recent study, Kölle and Quercia (2021), investigated subjects' perceptions of cooperation norms in social dilemma with and without strategic uncertainty. More specifically, they

implemented two sets of experiments: in the first set, they ran a one-shot PGG (with simultaneous contribution choices) that involved strategic uncertainty, and then also made subjects elicit their cooperative preferences using a variant of the strategy method. They then asked a different set of subjects (who played the role of spectators) for their empirical and normative elicitation regarding the choices made in the one-shot PGG with strategic uncertainty and also for the elicited cooperative preferences in the sequential version of the game. Subjects' empirical expectation was always (rightly) short of full cooperation irrespective of the context. However, there were some differences in the normative judgements: in the presence of strategic uncertainty, subjects judged that the most appropriate action would be to contribute everything to the public account. In the absence of such uncertainty, subjects displayed a strong normative belief of conditional cooperation.

2.3.3 Our contribution

Our results make three important contributions. Although other studies in the literature have used many types of norm-nudges (e.g., Dal Bó & Dal Bó, 2014; Krupka & Weber, 2009; Masclet et al., 2003) to improve cooperation outcomes, we find that it is the presence of a shared normative belief which is necessary and sufficient to improve initial group cooperation, and not just the saliency of the cooperation norm at the individual level, since cooperation involves inter-dependent choices. A follow-up finding is that this effect nonetheless is short-lived.

Secondly, we provide a deeper understanding of how the influence of normative messaging influences cooperation outcomes, through an indirect effect of improved expectation of other group members' behaviour. Norm-compliance relies on both empirical and normative information (e.g., Bicchieri, 2005; 2016). We show the effect of the interplay of these elements

in compliance with a cooperation norm. We show that saliency through a shared normative belief influences group members' empirical beliefs of each other's contribution, which in turn raises the groups' (initial) cooperation.

Third, we add to the understanding of the normative standard of (unconditional) cooperation shaped by local (group-level) experiences. Recent studies have shown that subjects find full cooperation as the most socially appropriate action in the presence of strategic uncertainty but display a norm of conditional cooperation with other's contribution choices. These studies use between-subject designs, where norm elicitations are carried out with a different set of subjects than from the ones that take part in the actual cooperation games. (e.g., Kimbrough & Vostroknutov, 2016; Kölle & Quercia, 2021). In contrast, we incorporate a within-subject design, where norms and behavioural outcomes are extracted from the same subjects. Our findings show that although all groups experience cooperation decay, the ones that perform better also elicit a stronger norm of (unconditional) cooperation post play. This adds to the literature on cooperation norms highlighting that people's views of normative standard of unconditional cooperation are based on their own group level experiences.

2.4 Experimental design

We exogenously manipulated the saliency of a cooperative norm through the Krupka-Weber (KW) norm-elicitation (2013) in a PGG. Our experiment is comprised of three main stages as shown in the illustration below (Figure 2.1). We had three main treatments HIGH, LOW and NO, with 20 groups (or 60 subjects) per treatment.

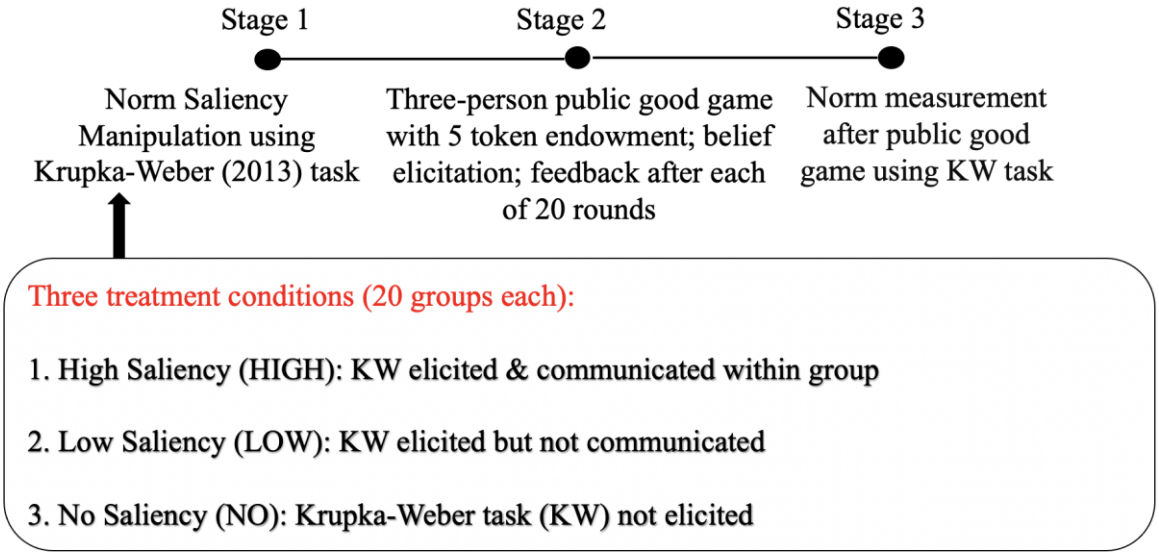


Figure 2.1: Overall design of the experiment

2.4.1 Overview of the design

The treatment manipulation was in Stage 1, with varying degrees of norm-saliency through the KW task (details below). In all treatments, Stage 2 was a three-person repeated PGG, with partner matching, belief-elicitation, and feedback, played over 20 rounds. Subjects were endowed with 5 tokens, which they could contribute to a group account in increments of 1 token. To see if saliency of a cooperative norm had any transient or long-lasting effect on one's own cooperative beliefs and cooperation choices, we compare these stage 2 contributions across treatments.

In all treatments, participants were assigned randomly to groups of three at the beginning of the experiment. Every participant was given 5 tokens at the beginning of each round, which they could keep in a private account or contribute to a group account. The payoff function is defined as follows:

$$p_i = 5 - g_i + 0.5 \sum_{j=1}^{3} g_j$$

Where g_i is $i's$ contribution; and g_j is the contribution of subject ***j,*** with a marginal social benefit factor of 1.5. The exchange rate was set at 1 point in the PGG = £0.08.

Before the start of round 1, participants were given detailed instructions about game play followed by an illustrative example. Participants were then asked a series of control questions, which served the purpose of a comprehension check. No participant could proceed further unless everyone in his/her group cleared the control questions.

In each round, participants were asked to state how much they wanted to contribute to the group account from the 5 tokens that they were endowed with in each round, and they also had to state their beliefs about how much they thought the other two group members would contribute to the group account in total (any number between 0-10). After every round, subjects were given feedback from the previous round, including their contribution to the group account, the sum of contributions from all members, points earned from the private account, points earned from the group account, earnings from those round, and cumulative earnings from all the previous rounds.

The cooperation norm in stage 1 and stage 3 was elicited using the KW method in the context of a one-shot PGG that subjects actually played in 20 rounds repeated PGG in stage 2[5]. For each of the six possible contribution choices, each participant had to select one of four possible responses: "very socially appropriate", "somewhat socially appropriate", "somewhat socially

[5] We wanted to elicit a normative standard of cooperation using the KW elicitation, not necessarily defined in terms of group members' compliance with cooperation outcomes.

inappropriate" and "very socially inappropriate". We defined socially appropriate as "*what most people in the room today would agree is the right thing to do"*. In other words, participants had to evaluate each choice in the decision environment as "socially appropriate" and "consistent with moral or proper social behaviour" or "socially inappropriate" and "inconsistent with moral or proper social behaviour".

Subjects were informed at the beginning of the experiment that they would be paid for only one of the stages that would be chosen at random. In NO, each of the PGG and Stage 3 KW had a 50:50 chance of being selected, while, in LOW and HIGH, the chances of Stage 1 KW, PGG and Stage 3 KW were 25%, 50%, and 25% respectively.[6] In the Stage 1 KW task, a subject's random choice was matched with the choice of another random subject in the session, while in the Stage 3 KW task, the matching was with anyone not from the same group but the same session.[7] If the choices matched, subjects were paid £15 and if they didn't match, they were paid £5. If the PGG was selected for payment, then subjects' payoff was the total cumulative payoff (exchanged in Points = £) at the end of the 20 rounds.

[6] Since Stage 2 PGG earnings were based on the cumulative earnings of 20 rounds which could range from £8 to £12 and in Stage 1 or Stage 3 KW elicitation earnings could be either £5 or £15, depending on the response they provided, these probabilities were computed to make sure that the expected earnings from a 60-minute session was around £10 in all treatments.

[7] This was to avoid the potential issue of a comparative advantage of group-members in HIGH, since they would have received feedback from a similar task in stage 1, and hence would have more information to coordinate better than group-members from NO and LOW.

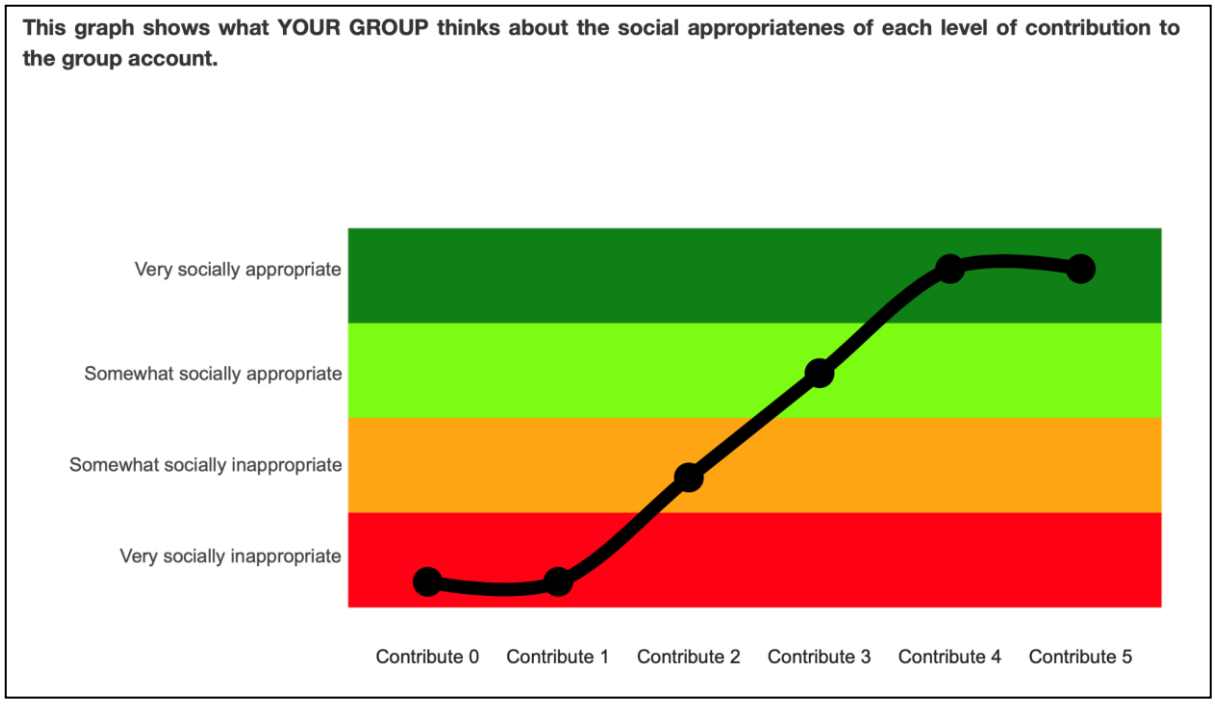


Figure 2.2: Group feedback from stage 1 KW elicitation in HIGH: This is from one of the 20 groups. Each point on the line graph reflected the group average normative rating for each possible contribution to the group account.

In our HIGH treatment, the cooperative norm was elicited through the KW task in Stage 1 for each contribution decision, and these evaluations were made public to the group members before they could make actual cooperative decisions. In other words, HIGH salience aimed to reflect the notion of a shared normative belief such that everyone within a group was aware of the collective norm, which may reinforce one's own perceptions of cooperation.

In HIGH, the norm was elicited and communicated within the group in the form of a colour-coded figure (see Figure 2.2 above). This feedback was given to the group members before round 1 of the Stage 2 PGG. The colour-coded graph displayed the group's average social appropriateness ratings for each level of contribution from 0 tokens to 5 tokens to the group account, with a gradual increment of 1 token. Thus, subjects in HIGH were not only made

aware of a cooperation norm but were also informed of what their group members collectively thought about the social appropriateness of each possible level of cooperation.

In LOW, the norm was also elicited, but the results were not communicated within the group. In other words, in LOW, subjects were forced to think about the collectively perceived social appropriateness of all possible levels of cooperation, but these perceptions were not communicated and hence were only private. Saliency was restricted to the individual and was hence lower than in HIGH, where collective perceptions of social appropriateness were common knowledge.

In NO, there was no norm elicitation in Stage 1, meaning subjects directly started with the Stage 2 PGG, with no saliency of a cooperative norm at all. In this condition, subjects might or might not have thought about the social appropriateness of various levels of cooperation. Because subjects were not forced to think about it, normative considerations were not salient. Therefore, NO served as our baseline treatment.

We elicited norms in a one-shot PGG and observed behaviour over repeated interaction in 20 rounds in the PGG. We wanted to keep the design simple in the sense that we wanted to manipulate the degree of saliency only, in terms of the common knowledge of the norm, and observe its effect on people's belief of others' behaviour and own behavioural outcomes. Moreover, we were primarily interested in eliciting a normative behavioural standard of (unconditional) cooperation, before and after the experience in the 20 rounds, and not a norm defined by group members' compliance.[8]

[8] Other studies that elicit a cooperation norm in repeated interaction find evidence of a norm of conditional cooperation, where matching the decision of other group members is considered the most appropriate action (e.g., Kimbrough & Vostroknutov, 2016; Kölle & Quercia, 2021)

We incorporated a Stage 3 KW norm-elicitation, similar to Stage 1, but after the end of the 20 rounds of PGG, to see if the experience in the game as a function of the initial degree of saliency could affect perceptions of a cooperation norm.

At the end of the stages in NO, LOW and HIGH, participants had to complete a brief questionnaire. This contained some basic demographic questions such as age, gender, field of study, political orientation, and ethnic background. Lastly, they were shown their earnings based on the stage selected at random. They were then called one by one individually and paid this amount privately.

The participants were recruited using ORSEE (Greiner, 2015). They were students from the University of Nottingham. We had a total of 180 subjects, with 60 subjects and 20 groups per treatment. All the sessions lasted about an hour and the average payment was £10 per subject. The experiment was conducted with the software LIONESS Lab (Giamattei et al., 2020).

2.5 Results

We first compare the Stage 1 KW ratings in LOW and HIGH, to check for differences in prior normative beliefs of cooperation before the start of the PGG. Following KW, we assign the following values to the possible responses in the KW method: -1 to "very socially inappropriate", -0.33 to "somewhat socially inappropriate", +0.33 to "somewhat socially appropriate" and +1 to "very socially appropriate". Figure 2.3 shows the average rating of the six actions in the PGG in LOW and HIGH.[9] In both conditions, subjects' prior beliefs regarding

[9] See Appendix: Tables A2.1 and A2.2 for mean rating and distribution of responses for the 6 possible choices in LOW and HIGH.

a cooperative norm were represented as follows: giving all 5 tokens was considered as the most socially appropriate action, and with declining contribution to the group account, the actions were rated as more and more socially inappropriate, with contributing 0 tokens being considered as the most socially inappropriate action[10].

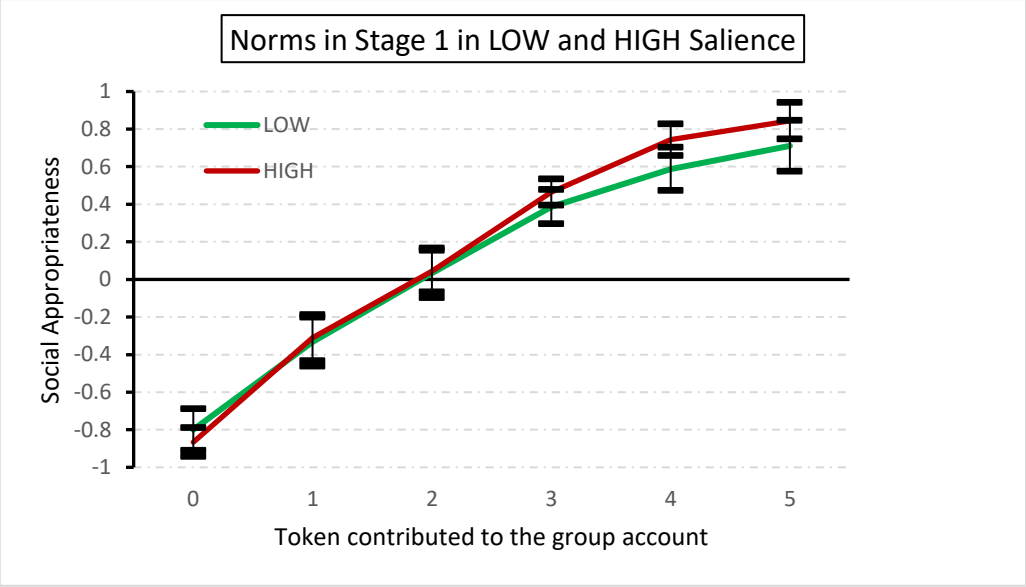


Figure 2.3: Mean appropriateness rating in Stage 1 KW in LOW and HIGH.

For all six contribution choices, there was no significant difference in the distribution of the responses between LOW and HIGH, taken at the group level, as is also reflected by the overlapping mean rating lines in Figure 2.3.[11] Subjects in both conditions had a similar shared belief regarding a norm of cooperation, when elicited before the start of the PGG. The vast majority of subjects i.e., 109 out of the 120 (91%) individual norm elicitations from LOW and HIGH displayed an increasing monotonic norm of cooperation, meaning that most subjects

[10] Other studies with one-shot PGG find a similar monotonic pattern, although they were elicited using a different set of subjects that those who participated in the actual PGG (e.g., Kimbrough & Vostroknutov, 2016; Kölle & Quercia, 2021).

[11] We report the *p*-values of Mann-Whitney tests adjusted for 6 multiple comparisons using the False Discovery Rate procedure, which is used for all comparisons through the analysis (Benjamini & Hochberg, 1995). All $p > 0.347$.

elicited a similar cooperative norm, implying an implicit consensus regarding the content of the norm.[12]

2.5.1 The influence of saliency on cooperation in round 1

We look at the effect on round 1 contributions and beliefs to see if the varying level of saliency introduced in stage 1 had any immediate effect on subjects' own cooperation attitudes as well as their perceptions of their group members' cooperation[13]. Figure 2.4 shows how varying salience in LOW and HIGH shifts the beliefs and contribution outcomes towards higher contribution levels relative to NO. Groups' average round 1 beliefs were 3.34, 2.74 and 2.17 and average round 1 contributions were 3.23, 2.85 and 2.31 in HIGH, LOW and NO respectively. There were significant differences in these distributions, supported by Kruskal-Wallis and bilateral Mann-Whitney tests[14]. We also checked for any correlation between these beliefs and cooperation decisions and found that they were positively correlated.[15]

Based on these observations, we checked if saliency causally influenced round 1 beliefs and round 1 contributions and, if so, then if the effect on contributions was moderated through beliefs. Table 2.1 shows these causal effects: the regressions show that both LOW and HIGH had a statistically significant positive effect on round 1 beliefs of the other two group members'

[12] Nine subjects (7.5%), 7 from LOW and 2 from HIGH, displayed a norm whereby they rated both extreme choices of contributing all 5 tokens and contributing 0 tokens to the group account as socially inappropriate, with the remaining contribution choices as socially appropriate; the 7 from LOW also contributed between 1-3 tokens in round 1, implying that their contribution decision was based on their "perceived norm" of cooperation. However, for the other 2 subjects displaying a similar norm from HIGH, contributed 4 and 5 tokens respectively in round 1. This may be due to the feedback of the Stage 1 norm from the group (in a visual illustrated colour coded graph) that modified their perception of the norm. Two subjects in LOW (1.6%) rated all six choices as socially appropriate.

[13] Number of subjects with 0 contribution in round 1 were NO (11) > LOW (10) > HIGH (5); likewise with round 1 beliefs NO (5) > LOW (2) > HIGH (0). This shows that at an individual level, subjects' perception of what the other 2 group members would contribute and their own choices were affected by the degree of norm saliency introduced prior to the game.

[14] Kruskal Wallis for beliefs and contribution: $\chi^2(2)=15.257$, $p=0.0005$; and $\chi^2(2)=7.765$, $p=0.0206$; Mann Whitney for beliefs and contributions: NO vs HIGH: Z=-3.922, p=0.0000, Z=-2.570, p=0.0279; NO vs LOW: Z=-1.831, p=0.0677, Z = -1.374, p = 0.1732; LOW vs HIGH : Z = -1.997, p= 0.0677, Z =-1.785, p = 0.1128.

[15] NO: r (18) =0.716, p=0.004; LOW: r (18) = 0.4663, p = 0.0382; HIGH = r (18) =0.571, p = 0.0086.

contribution. HIGH and LOW improved the average belief by 1.094 tokens and 0.567 tokens respectively even after controlling for other socio-demographic factors (Col 1). We also found that although saliency appeared to improve round 1 contribution (Col 2), these effects disappeared (Col 3) after controlling for round 1 beliefs. In other words, the effect of saliency on round 1 contribution was moderated through its effect on round 1 beliefs.

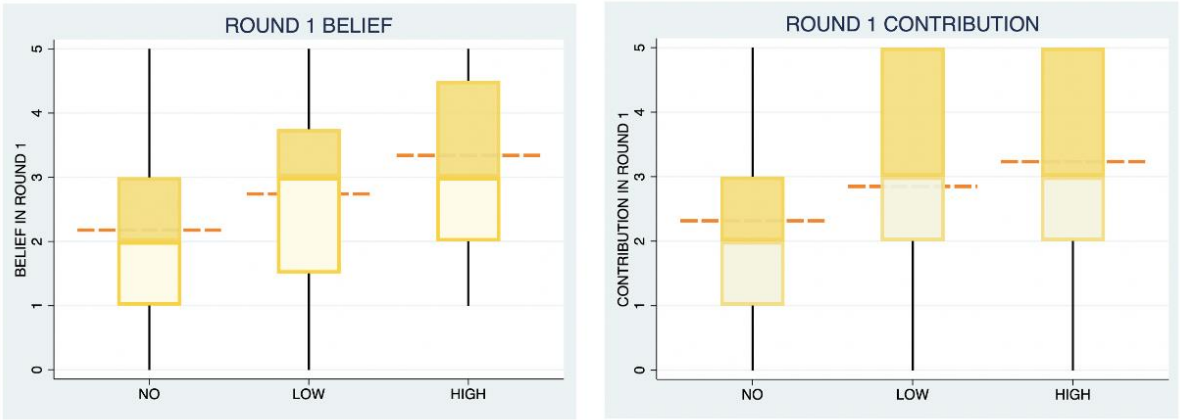


Figure 2.4: Boxplots of round 1 beliefs and contributions. The orange lines reflect the average contributions across the conditions which increases from NO < LOW< HIGH for both beliefs and contribution decisions. The median in the boxplot demarcates the interquartile range as represented by the shaded yellow areas. The median in NO is 2 tokens, while for LOW and HIGH is 3 tokens respectively, for both Round 1 belief and contribution. We asked subjects to state their belief of the sum of the other two group members' contribution, which was on a scale of 0-10 tokens. To keep beliefs on the same scale as contribution in our illustrations and analysis, we divide the stated beliefs by 2.

VARIABLE: Round 1	(1) Belief	(2) Contribution	(3) Contribution
HIGH	1.094***	0.897**	0.124
	(0.251)	(0.362)	(0.303)
LOW	0.567*	0.570*	0.169
	(0.286)	(0.327)	(0.278)
Belief			0.707***
			(0.0741)
Constant	1.214*	1.855**	0.996*
	(0.660)	(0.823)	(0.594)
Controls	Yes	Yes	Yes
Observations	180	180	180
R-squared	0.134	0.077	0.384

Table 2.1: OLS regression. Dependent variables are belief and contribution in round 1. Standard errors in parentheses, adjusted for intra-group correlation (groups are used as independent clustering units, with 60 groups). Significance levels: *** p<0.01, ** p<0.05, * p<0.1. We control for age, gender, ethnicity, background in Economics, and political orientation.

2.5.2 Beliefs and cooperation over rounds

In figure 2.5, we show the effect of saliency on average belief and cooperation[16] across all 20 rounds in the repeated PGG. We make a few observations. First, as we move across the rounds, average group beliefs stay consistently higher than NO in the 20 rounds, while average contributions stay higher in all rounds except round 6. Second, although round 1 contribution were higher in LOW than NO in round 1 and subsequent rounds, then fell below NO at round 12 and remained similar to the baseline condition. Likewise, the average contributions were higher in LOW than NO intermittently in some rounds, but stayed lower than NO, after round 10. Third, there is a mild decay in the average beliefs while this decay is sharper with regards to average contribution across the 20 rounds for all the three conditions.

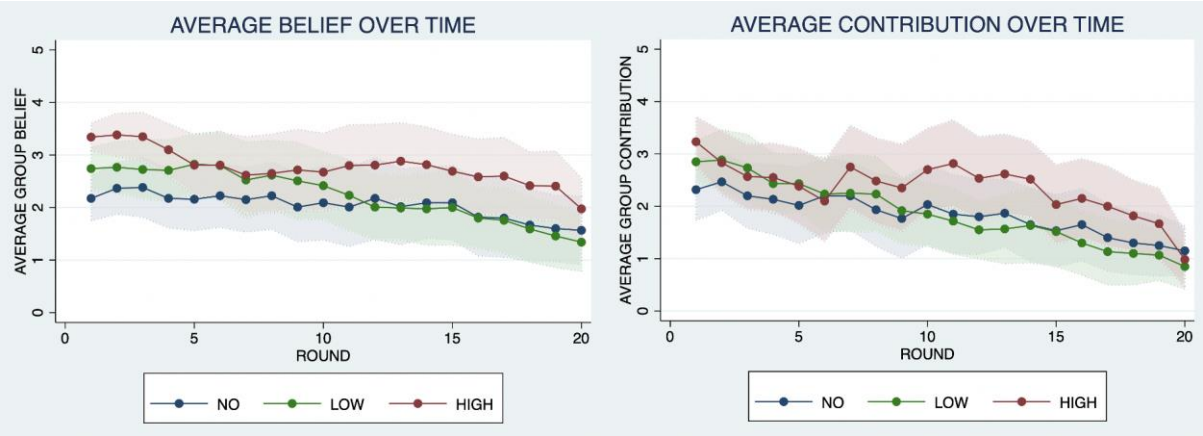


Figure 2.5: Average beliefs and contributions. This is over 20 rounds in fixed groups of three in NO, LOW and HIGH. The shaded area reflects the overlapping confidence limits of the average beliefs and contributions choices.

We formally express the results using GLS regressions in Table 2.2 below. In Model (1) and (2), the dependent variable is average group belief across the 20 rounds of PGG, and we regress these on the main treatment variables HIGH and LOW, and round as well as the interaction of these terms to see the influence of rate of belief decay in the treatments on average beliefs. We

[16] We found high positive correlation in the groups' average belief and contribution decisions in the 20 rounds: NO: r (18) = 0.9567, p=0.000; LOW: r (18) =0.8725, p= 0.000; HIGH = r (18) =0.9505, p=0.0000.

make the following observations: firstly, HIGH increases the average group beliefs by 0.669 tokens (p = 0.073) while LOW increases the average belief by 0.193 tokens, but this is statistically insignificant (p = 0.588). With passing rounds, the average beliefs decline by 0.0539 tokens (p = 0.000).

In Model (2), with the interaction terms, we find that with each passing round average beliefs in HIGH decline by 0.0478 tokens, although the effect is insignificant (p=0.558) while in LOW, the rate of belief decay is 0.0791 tokens is significant (p = 0.042). In other words, we find a sharper belief decay in LOW compared to NO, across the 20 rounds, whereas HIGH and NO witness a similar rate of decay.

In Model (3), (4) and (5), the dependent variable is average group contribution, and we regress our model on the main treatments HIGH, LOW, round variable, and its interaction with HIGH and LOW, as well as on group beliefs. We find no effect of HIGH (p = 0.227) and LOW (p= 0.850) on average contribution, but a significant effect of round (p = 0.000), signifying declining cooperation with passing rounds, by 0.0759 tokens, in Model (3). On the other hand, in Model (4), we find that an increase in group belief by 1 token increases the group contribution by 0.614 tokens (p = 0.000). This effect stays significant (p = 0.000) in Model (5) with the interaction variables of round x HIGH and round x LOW.

To sum up, we find that saliency only has a transient effect on cooperation, which does not sustain over time, but it has a more long-lasting effect on group member's beliefs regarding group-members' contribution choices across the 20 rounds (at 10% in the HIGH treatment only). However, this effect does not help with improved group cooperation over time. Irrespective of the degree of saliency introduced prior to the game and its effect on round 1

outcomes, there is both belief decay regarding group members' contribution, as well as cooperation decay with passing rounds in all treatments.

VARIABLE: Round 1-20	(1) Belief	(2) Belief	(3) Contribution	(4) Contribution	(5) Contribution
HIGH	0.669*	0.806**	0.465	0.0549	0.00775
	(0.372)	(0.351)	(0.385)	(0.176)	(0.267)
LOW	0.193	0.660*	0.0656	-0.0528	0.121
	(0.356)	(0.371)	(0.347)	(0.173)	(0.256)
Round	-0.0539***	-0.0347*	-0.0759***	-0.0428***	-0.0391***
	(0.00891)	(0.0177)	(0.00719)	(0.00548)	(0.00893)
Round x HIGH		-0.0131			0.00477
		(0.0223)			(0.0133)
Round x LOW		-0.0444**			-0.0164
		(0.0219)			(0.0119)
Belief				0.614***	0.610***
				(0.0376)	(0.0385)
Constant	1.744***	1.543**	1.479**	0.408	0.375
	(0.598)	(0.612)	(0.653)	(0.487)	(0.506)
Controls	Yes	Yes	Yes	Yes	Yes
Observations	3,600	3,600	3,600	3,600	3,600
Number of subjects	180	180	180	180	180

Table 2.2: Random effects Regression. Dependent variables are average belief and contribution in 1-20 rounds. Standard errors in parentheses, adjusted for intra-group correlation (groups are used as independent clustering units, with 60 groups). Significance levels: *** p<0.01, ** p<0.05, * p<0.1. We control for age, gender, ethnicity, background in Economics, and preference for political orientation.

2.5.3 The influence of saliency on norms after experience in the game

To see if the varying degree of saliency had any effect on Stage 3 KW evaluations through the experience of cooperation, we compare the distributions of normative ratings elicited in Stage 3 across the three conditions. Figure 2.6 shows the average norm rating in NO, LOW and HIGH for each of the six possible contribution decisions in the PGG. We compared the distributions and found no significant differences among them: (all p values > 0.167, bilateral Mann-Whitney tests with p-values adjusted for 18 multiple comparisons). In all conditions, giving 0 tokens to the group account, is rated as the least socially appropriate action, while giving all

five tokens is the most socially appropriate action, and the actions become more and more socially appropriate with an increasing number of tokens contributed to the group account[17,18].

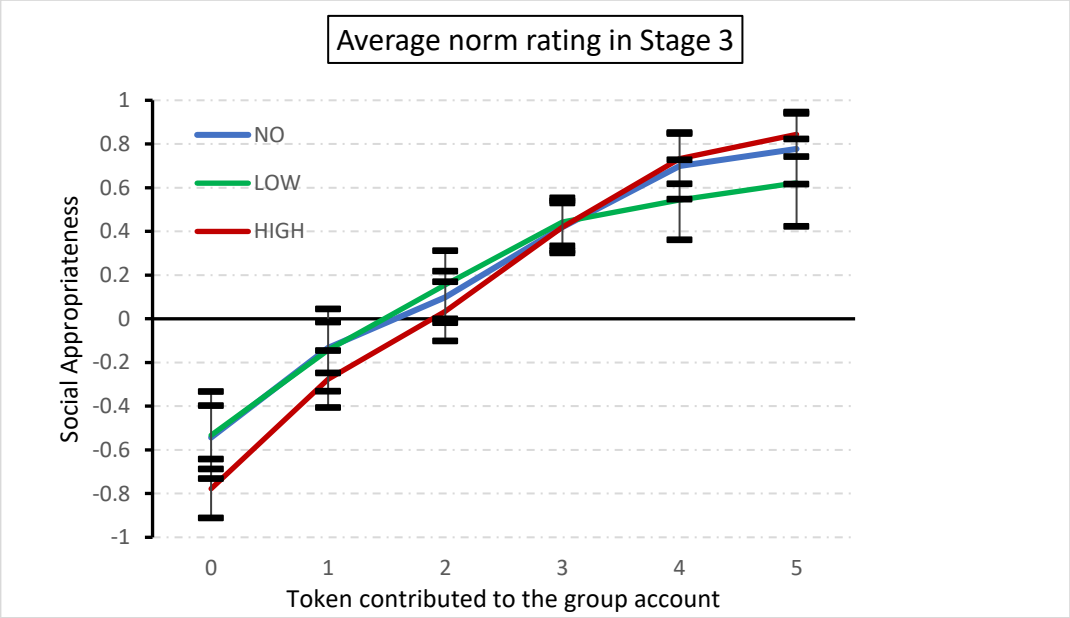


Figure 2.6: Average norm-rating across 6 contribution decisions in NO, LOW and HIGH.

To check if the experience in the 20-round fixed group PGG had any effect on the post-play KW ratings, we looked into the association of group's average contribution in the 20 rounds and the Stage 3 average norm rating for the six contribution choices as shown in the scatterplots in Figure 2.7. We pooled data from all 60 groups in NO, LOW and HIGH since all groups witnessed similar cooperation decay in the 20 rounds and saliency of a cooperative norm influenced the initial contribution decision only, and not the overall gameplay. From the scatterplot, we find negative correlation between average contribution in the 20 rounds and the norm ratings for the contribution choices 0, 1, 2, and 3 tokens to the group account, and positive

[17]See Appendix: Tables A2.3, A2.4, A2.5 for the detailed distributions.
[18] We also compare the Stage 1 and Stage 3 KW ratings within groups from LOW and HIGH to check if the experience in the game had any effect on the Stage 3 ratings and find similar patterns of increasing social appropriateness with higher contribution choices; see Figure A2.1 in the Appendix.

correlations between average contribution and norm rating for contribution choices of 4 and 5 tokens to the group account. This shows that groups that witnessed better cooperation in the repeated rounds, also appeared to provide slightly more positive rating for higher contribution choices and harsher norm rating for low or no contribution to the group account, irrespective of the initial degree of saliency of a cooperative norm. In other words, groups that cooperated better than others elicited a stronger perception of a cooperative norm at the end of the 20 rounds. [19,20]

For a more detailed view of this result, we also looked into the groups that experienced the highest and lowest cooperation in the 20 rounds in NO, LOW and HIGH and their respective KW norm ratings. We computed the average contribution in the 20 rounds from the 20 groups each in NO, LOW and HIGH, and then ranked the groups from highest to lowest based on these average contributions within the condition. We then took the top 5 groups and bottom 5 groups (based on 15 subjects each) from NO, LOW and HIGH[21] and checked their KW evaluations both prior and post PGG as shown in Figure 2.8[22].

[19] Additionally, to see if group average ratings change from Stage 1 to Stage 3, we took the difference between stages 1 and stage 3 ratings for all the 6 contribution choices. 0 indicates no difference, positive difference shows actions being rated more harshly in Stage 3 after the experience in the game compared to Stage 1, and negative difference indicate actions being rated less harshly in Stage 3: We find similar correlation between average contribution and difference in norm-rating as shown in Figure A2.2 in the Appendix.

[20] We also run a GLS regression and find that subjects' experience in the game influenced their stage 3 ratings: this is reported in Table A2.6 in Appendix.

[21] Although we pooled the data from all treatments to check the correlations in PGG and stage 2 KW, we looked into the top/bottom 5 each within NO, LOW and HIGH. We didn't take the top/bottom 5 from across the treatments, as saliency manipulation did affect round 1 and average contributions differently: positively in round 1 and no effect overall.

[22] The magnitude of the top 5 and bottom 5 groups' average contribution is mentioned in Figure 2.8 in the parenthesis.

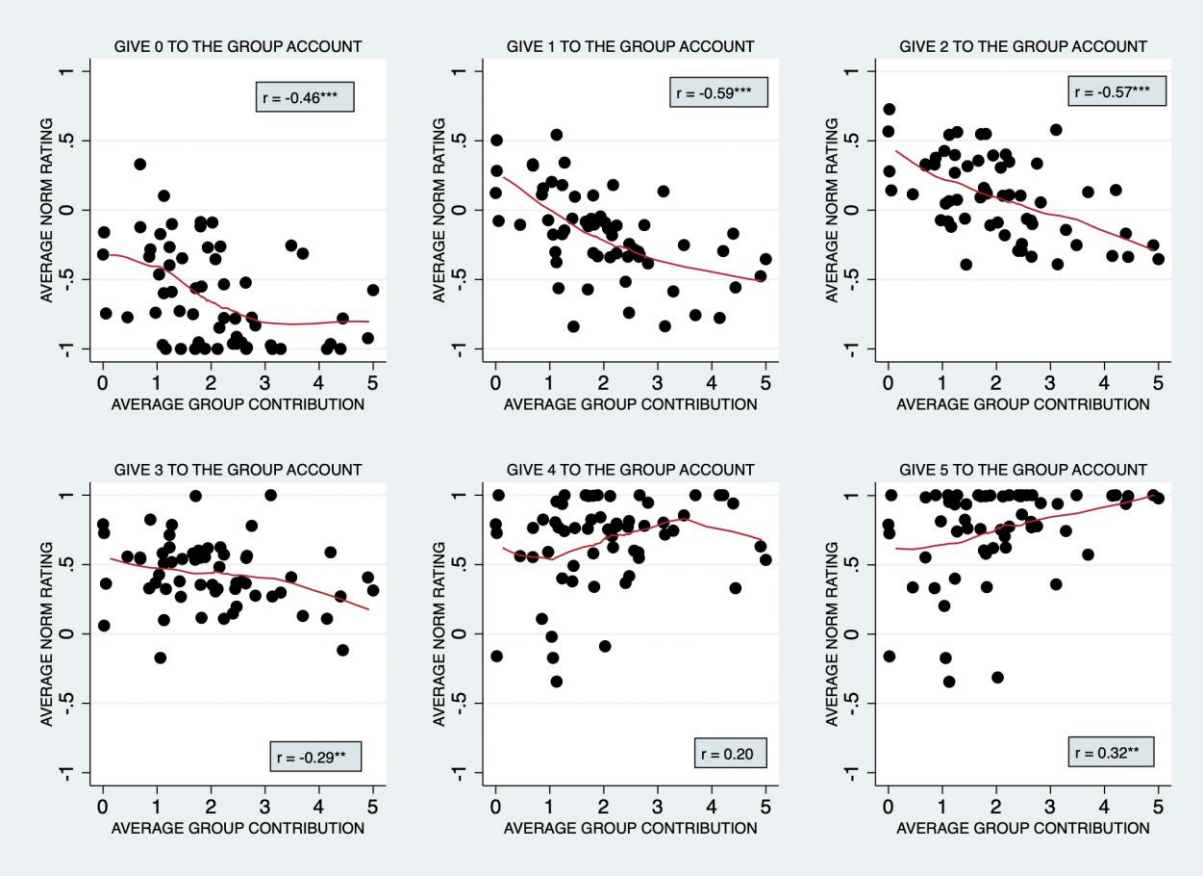


Figure 2.7: Scatterplot of norms and contribution. From Top left: Group average contribution and norm rating for 0 contribution in the group account are moderately negatively correlated (r (58) = -0.4589, *p* = 0.0002); similar negative correlation between group's experience in the game and norm rating for 1 token (r (58) = -0.5943, *p* = 0.0000); 2 tokens (r (58) = -0.5691, *p* = 0.0000); and 3 tokens (r (58) = -0.2912, *p* = 0.0240) respectively. On the other hand, we find a low positive correlation between contribution and norm rating for contribution choice of giving 4 tokens (r (58) = 0.2007, *p* = 0.1241) and contribution choice of 5 tokens to the group account: (r (58) = 0.3157, *p* = 0.0140)[23]. The norm ratings in the experiment were in the range of -1 to 1 and contribution tokens were in the range of 0-5.

We observed that the top 5 groups in all the conditions started at a higher contribution choice and belief in round 1 compared to the bottom 5 groups. Additionally, although the top 5 groups also experienced cooperation decay, they were able to maintain a higher level of cooperation and beliefs about group members' contribution choices throughout the 20 rounds compared to the bottom 5 groups. As far as the normative evaluations are concerned, we found a few

[23] We find a similar pattern for group average beliefs and the rating for the 6 contribution choices in stage 3, reported in Figure A2.3 in Appendix: moderate negative correlation between groups' average belief and contribution rating for 0, 1, 2 and 3 tokens, and positive correlation between groups' belief and rating for contribution choice of 5 token to the group account.

interesting observations: firstly, groups that experienced higher cooperation also reported a stronger norm of cooperation in stage 3 compared to groups that experienced low cooperation in all treatments. This is reflected by the slightly steeper purple lines in NO, LOW and HIGH in Figure 2.8.

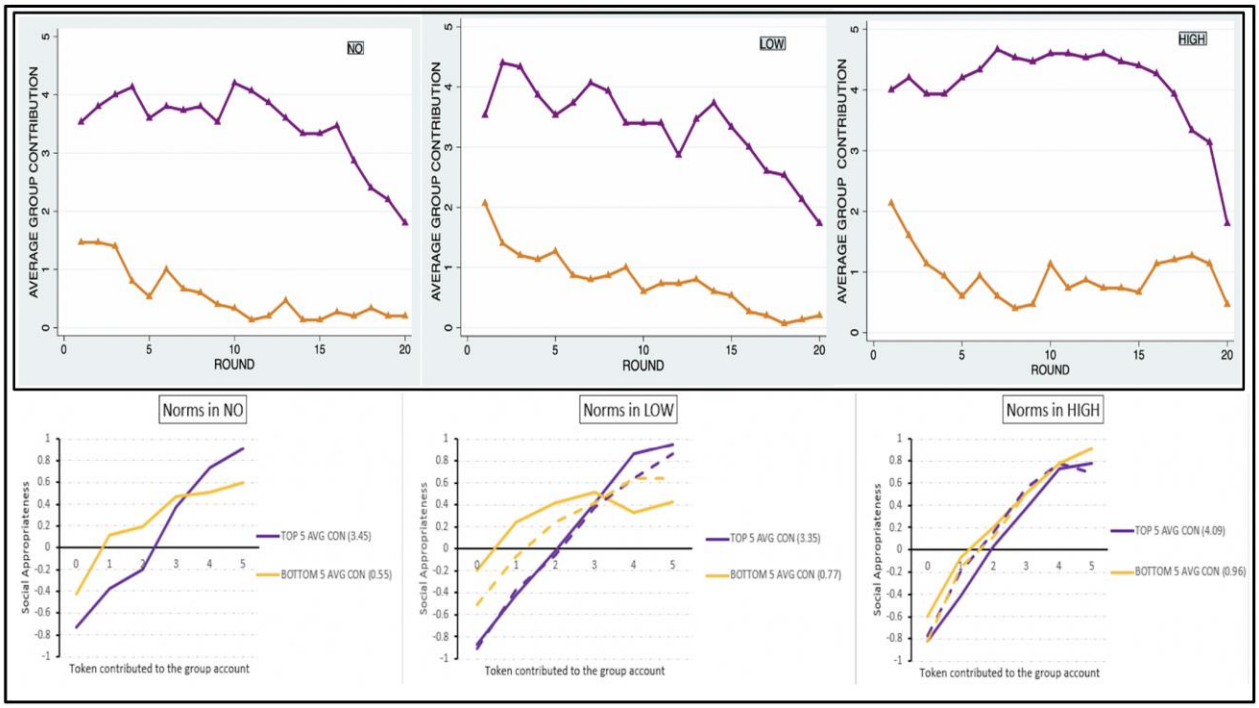


Figure 2.8: KW elicitation of the top 5 and bottom 5 groups in NO, LOW and HIGH. The lines represent the average contributions, shown at the top portion of this panel figure. The dashed lines represent the Stage 1 KW ratings in LOW and HIGH, and the solid lines indicate stage 3 norms. The values in parentheses are the average group contribution in the 20 rounds for the top 5 and bottom 5 groups, which are similar in NO, LOW and HIGH.

Secondly, we see in LOW that for the bottom 5 groups, the norms were weaker in Stage 1 compared to the top 5 groups as demonstrated by the flatter dashed yellow line compared to the dashed purple line, which may have led to less cooperation in these groups in the PGG. Additionally, for these groups, the norms become even weaker in Stage 3, after the experience of low cooperation as is shown by the flatter yellow line. In other words, the bottom 5 groups consider giving nothing or less to the group account as more socially appropriate in Stage 3 compared to the top 5 groups. On the other hand, for the top 5 groups, the norms stay relatively strong and stable as is evidenced by steeper (purple line) and overlapping lines from stage 1

and stage 3 average norm ratings. In HIGH, although the norms are slightly stronger for the top 5 groups than the bottom 5 groups as in the other conditions, the stage 1 (dashed) and stage 3 (solid) lines somewhat overlap for the top as well as the bottom 5 groups, reflecting a higher consistency in a norm of cooperation. This may be due to a shared consensus regarding the appropriateness of cooperation within group members' which may have reinforced their perception of a normative belief of cooperation.

2.5.4 Study 2: Online replication

To examine the robustness and replicability of the results, we implemented the experiment a second time using an online setup. We recruited participants from the University of Nottingham subject pool, using ORSEE (Greiner, 2015) but this time instead of a physical laboratory, the experiment was conducted online.[24] The implementation of the online experiment only differed minimally from the lab experiment (see Appendix: Experimental Instructions for details). We recruited 270 subjects in this online experiment, with 30 groups per treatment. The experiment lasted around 60 minutes. We ran NO, LOW and HIGH randomly across the sessions. We estimated the sample size based on the effect of HIGH vs NO on round 1 contribution from Study 1 (0.56) and assuming $\alpha = 0.05$ and a power of 0.80.

We checked subjects' prior cooperation norm through Stage 1 KW task (see panel A in Figure 2.9) and found a similar pattern as in Study 1 for the six contribution choices: the responses shift from very socially appropriate to very socially inappropriate as we move from 5, 4, 3, 2, 1, and 0 tokens[25].We found no difference in the distribution of responses in the six contribution choices in LOW and HIGH: (Mann-Whitney test after adjusting for 6 comparisons, $p > 0.15$).

[24] This data was collected in November-December 2020, during the Covid-19 pandemic, when conducting experiments in physical laboratory was not feasible.
[25] Refer to the Tables A2.7, A2.8 in Appendix for the distribution of responses for 6 contribution choices.

We also checked if saliency had any immediate causal effect on round 1 beliefs and contribution decisions and found that, as in Study 1, HIGH increased round 1 beliefs of other 2 group members' contribution by 0.609 tokens. Likewise, the effect of HIGH on round 1 contributions was through its effect on round 1 beliefs.[26] Differently from Study 1 (where saliency influenced group beliefs across the 20 rounds at 10%), we found that there was no such causal effect on the average group beliefs and contribution choices across the 20 rounds. All groups experienced beliefs and contributions decay: average group beliefs declined by 0.0195 tokens, while average group contributions declined by 0.027 tokens (after controlling for beliefs), with every passing round (see panel B in Figure 2.9)[27].

Lastly, we also compared the Stage 3 KW ratings in NO, LOW and HIGH, and found no difference in the distributions, results supported using bilateral Mann-Whitney tests (see panel C in Figure 2.9; all $p > 0.1$; never significant for all 18 comparisons)[28]. We also compared Stage 1 and Stage 3 norms within groups in LOW and HIGH to check for order effects and found no significant change in the distribution of responses (see panel D in Figure 2.9): Result supported by bilateral signed rank tests, (all $p > 0.46$) and never significant. The group-level experience in the game had no effect on the shift in Stage 3 norms, reflecting a stable norm of cooperation.

To see if the experience of cooperation affected Stage 3 norm ratings, we checked the correlations between average group contributions and average Stage 3 ratings for the six contribution choices in all groups. Like in Study 1, we found negative correlation between average contribution and norm ratings for the contribution choices of 0,1,2 and 3 tokens, and

[26] See Table A2.9 in Appendix for the regression analysis
[27] See Table A2.10 in Appendix for the regression analysis
[28] See Tables, A2.11, A2.12, A2.13 in Appendix for detailed distribution

positive correlation for the contribution choices of 4, and 5 tokens (see panel E in Figure 2.9) [29]. We also took the top 5 and bottom 5 groups in terms of their average contribution in the 20 rounds in NO, LOW and HIGH and observed that, in all conditions, the top 5 groups showed a slightly stronger norm of cooperation in stage 3, compared to the bottom 5 groups, indicated by the steeper purple lines (see panel F in Figure 9).

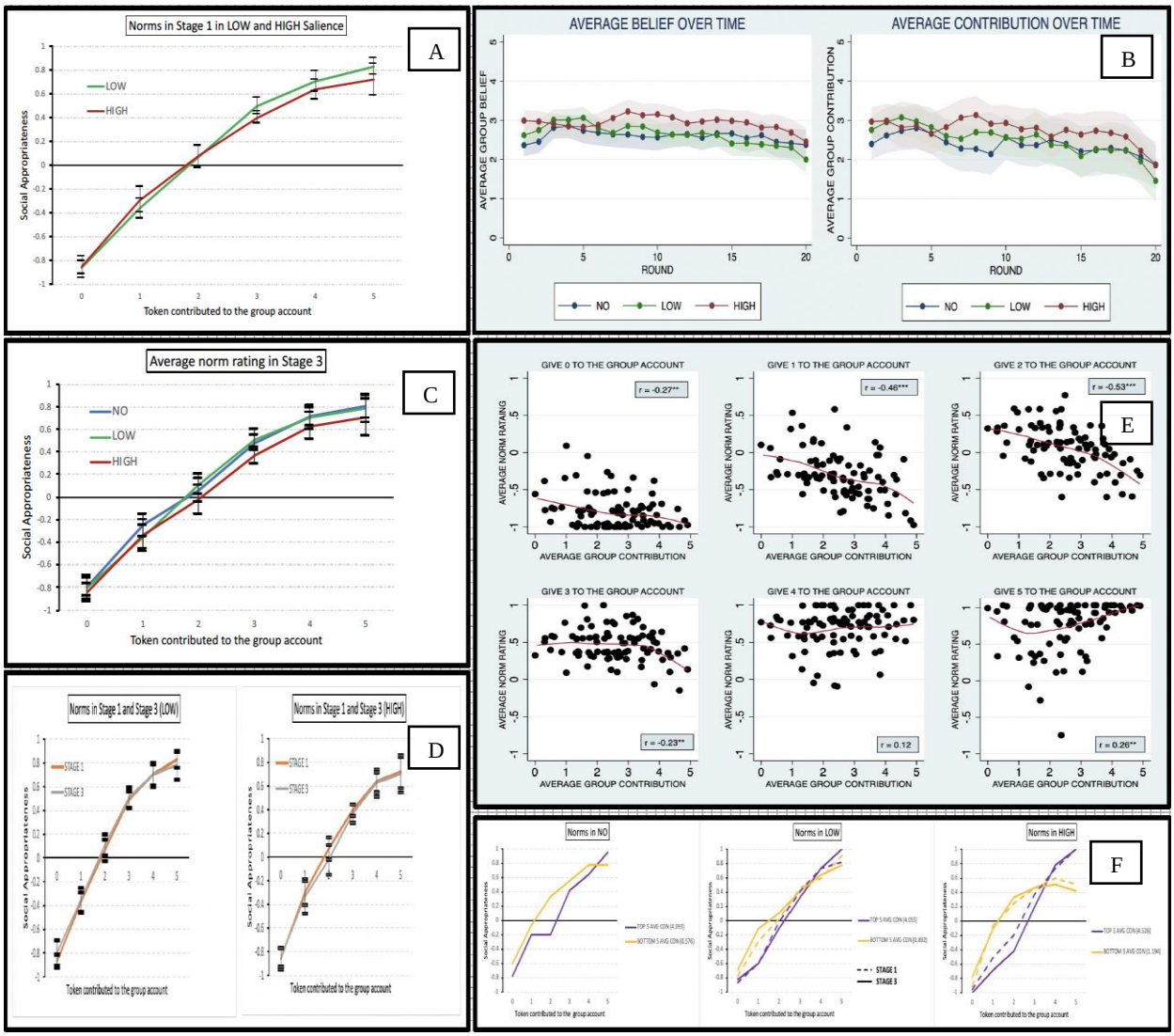


Figure 2.9: Panel showing the behavioural and norm results from Study 2.

[29] See table A2.14 in Appendix for the regression analysis.

In summary, our replication showed that only saliency of a shared normative belief led to an increase in average group contribution in round 1 of the PGG through a mediated effect of group members' beliefs of their expected contribution. Second, this effect was not long-lasting, as cooperation declined over the 20 rounds. Finally, there was no effect on the norm of cooperation post play[30].

2.6 Discussion and Conclusion

To study the impact of norm-saliency in social dilemmas, we exogenously manipulated the degree of saliency of a cooperative norm to see if saliency improves cooperative beliefs and choices and also if the experience of actual cooperation influences the normative evaluation of cooperation and free riding. We implemented a three-stage experiment, with saliency manipulation in Stage 1, followed by a 20 round three-person PGG with belief elicitation and feedback in Stage 2, followed by norm measurement after PGG in Stage 3. Both Stage 1 and Stage 3 were carried out using the KW norm-elicitation method in a three-person PGG.

Our results show that HIGH and LOW both improved round 1 beliefs regarding group members' contribution choices, which in turn mediated the effect on round 1 cooperation outcomes. We observed similar result in the online replication only with HIGH. In other words, the saliency of a normative standard of cooperation especially through a shared normative belief, led to an improved expectation of group members' contribution. Group feedback of a shared consensus regarding a norm of cooperation in HIGH, made groups form improved empirical beliefs, and both types of beliefs aligned with each other leading to increased cooperation, at least initially (round 1).

[30] We pool all groups from both studies and look into the overall influences of saliency on cooperation outcomes: See Table A2.15 in Appendix

The group feedback of the normative standard in HIGH may have helped in two ways: firstly, a similar collective evaluation of a norm of cooperation (increasingly monotonic) by group members might have "reinforced" a subjects' own judgement of the acceptable normative standard of cooperation and secondly, a shared consensus may have led to an increased sense of "like-mindedness" which in turn might have had a role in raising the cooperative attitude, affecting both the empirical belief of group members' contribution as well as one's own contribution choice in round 1. This could also explain why the effect of LOW saliency was not always effective in improving round 1 outcomes.

Moreover, HIGH influenced average group beliefs in Study 1 only, but had no effect on average group cooperation in the 20 rounds, in both studies. On the other hand, LOW never influenced either group beliefs or contribution outcomes. In both studies, the groups' average (negative) empirical beliefs influenced group's cooperation decisions. Additionally, neither LOW or HIGH affected belief and contribution decay. In summary, this meant that although the information of shared normative beliefs improved belief of other's contribution (Study 1), this didn't influence cooperation choices over a period of time.

A possible explanation of this may be that people have heterogeneous cooperative attitudes (Burlando & Guala, 2005; Fischbacher & Gächter, 2010; Ones & Putterman, 2007; Thöni & Gächter, 2015) and varying levels of norm-sensitivity, i.e., the utility derived from following a norm varies across individuals (Bicchieri, 2016; Fehr & Schurtenberger, 2018; Kimbrough & Vostroknutov, 2016) and the ones that have lower levels of sensitivity may act as trend-setters (Bicchieri, 2018). Here, individuals with low norm-sensitivity towards cooperation, implying a low cooperative attitude, may gradually contribute less with each passing round.

Consequently, subjects update and lower their empirical beliefs of group members' contribution choices as a result of the now increased salience of the heterogeneous cooperative preferences, inferred through the groups' feedback of contribution choices from the preceding rounds. With passing rounds, the positive effect of norm-saliency dissipates and the lowered empirical beliefs of group member's contribution, through the selfish actions of a few, lowers the average group cooperation and leads to cooperation decay. Negative empirical feedback and consequently lowered updated beliefs of low and no contribution of a few group members impacts cooperation choices negatively and more strongly that positive cooperation (with moderate to high) cooperation choices, as it caters to selfish-interest motives, leading to an asymmetric empirical effect (e.g., Croson & Shang, 2008; de Oliveira et al., 2014; Dimant, 2019; Gino et al., 2009; Thöni & Gächter, 2015) and thus lowers the average cooperation outcomes across the 20 rounds with the gradually increased motivation to act as income-maximisers towards the end rounds (Fischbacher & Gächter, 2010). When normative and empirical beliefs seem to contradict each other, it is usually the empirical information and beliefs that carry more weight and drive behaviour (Bicchieri & Xiao, 2009; Keizer et al., 2008; Lindenberg & Steg, 2007; Smith et al., 2012).

Stage 3 normative elicitations could be affected by treatment manipulation of norm saliency in Stage 1, or the experience of declining cooperation in fixed groups in stage 2, or both. We found that HIGH and LOW had no effect on the post-play evaluation of a cooperative norm, in Study 1 and Study 2. This is not surprising, as the effect of saliency of a normative standard dissipates over time as illustrated by the cooperation decay in Stage 2 behavioural outcomes. However, irrespective of the initial degree of saliency, groups undergo varying experience of cooperation dynamics due to heterogeneous cooperative attitudes. Groups that start with high(er) contribution outcomes, although also experience a gradual decline in cooperation,

perform consistently better in terms of average group beliefs and cooperation choices across the 20 rounds. On looking into the association between Stage 3 norms and such varying cooperation experiences, we find negative correlations, i.e., groups that experience higher cooperation, also provide a harsher rating for no and low contribution choices, and a stronger positive rating for high cooperation choices in Stage 3. This shows that although all groups elicit a similar content of the norm in Stage 3, illustrating the stability of a socially acceptable cooperative standard, groups exhibiting improved cooperation outcomes, elicit a strengthened norm of cooperation post-play. On the other hand, groups that displayed low initial and overall cooperation dynamics, provided a weaker normative standard of cooperation at the end. This may be due to a combination of self-serving bias among the selfish individuals and an updated normative belief through a pessimistic view of a norm of cooperation, among group members influenced through such negative peer effects.

Our results raise a few important questions regarding the study of cooperation in the presence of a shared normative belief. For instance, varying degrees of norm-saliency of an injunctive norm of cooperation could be introduced intermittently across repeated rounds to observe its effects of cooperative beliefs and outcomes over repeated interaction. In this case, the results on post-play norms can be further evaluated in the presence of experience of cooperation as a function of norm saliency, if the repeated reminder of norms influences cooperation and prevents decay. The role of group size can also be investigated in the context of high salience of a cooperative norm, in the sense if a shared normative belief with a bigger group exerts a stronger influence on cooperative attitudes. The evidence of group size on cooperation is somewhat mixed (e.g., Barcelo & Capraro, 2015; Duffy & Xie, 2016; Nosenzo et al., 2015) and since norm-talk relies on communication within group-members regarding a shared

consensus, which reinforces the norm, it would be worthwhile to observe how high salience in a bigger group with shared belief affects cooperation.

2.7 Appendix

2.7.1 Supplementary Analysis and Figures

	N	Mean	-1	-0.33	+0.33	+1
Give 5	60	0.71	3.3%	8.3%	16.7%	71.7%
Give 4	60	0.58	1.7%	5%	46.7%	46.7%
Give 3	60	0.38	1.7%	5%	76.7%	16.7%
Give 2	60	0.03	3.3%	51.7%	31.7%	13.3%
Give 1	60	-0.33	23.3%	58.3%	13.3%	5%
Give 0	60	-0.79	76.7%	20%	0%	3.3%

Table A2.1: Appropriateness ratings in Stage 1 KW task in LOW. For each possible contribution choice, we show the average rating, and the frequency of responses, where -1 is "very socially inappropriate"; -0.33 is "somewhat socially inappropriate"; +0.33 is "somewhat socially appropriate" and +1 is "very socially appropriate". The shaded cells represent the most frequent response for each contribution choice. We find that giving 5 or 4 or 3 tokens are considered socially appropriate, while giving 2 or 1 or 0 are considered as socially inappropriate actions.

	N	Mean	-1	-0.33	+0.33	+1
Give 5	60	0.84	1.7%	1.7%	15%	81.7%
Give 4	60	0.74	0%	0%	38.3%	61.7%
Give 3	60	0.46	0%	0%	80%	20%
Give 2	60	0.04	1.7%	46.7%	45%	6.7%
Give 1	60	-0.31	21.7%	55%	21.7%	1.7%
Give 0	60	-0.86	81.7%	16.7%	1.7%	0%

Table A2.2: Appropriateness ratings in Stage 1 KW task in HIGH. Most participants (modal responses) are as follows: contributing 5 and 4 tokens are very socially appropriate; contributing 3 tokens is somewhat socially appropriate, contributing 2 and 1 are somewhat socially inappropriate while contributing 0 is very socially appropriate.

Action	N	Mean	-1	-0.33	+0.33	+1
Give 5	60	0.77	5%	3.3%	11.7%	80%
Give 4	60	0.69	3.3%	5%	25%	66.7%
Give 3	60	0.42	0	8.3%	70%	21.7%
Give 2	60	0.09	1.7%	36.7%	56.7%	5%
Give 1	60	-0.13	10%	56.7%	26.7%	6.7%
Give 0	60	-0.54	55%	31.7%	3.3%	10%

Table A2.3: Distribution and mean appropriateness rating in Stage 3 in NO. Modal responses (as identified by the grey cells) for the 6 possible contribution choices are as follows: contributing 5 and 4 tokens are very socially appropriate; contributing 3 and 2 tokens is somewhat socially appropriate, contributing 1 token is somewhat socially inappropriate while contributing 0 is very socially appropriate.

Action	N	Mean	-1	-0.33	+0.33	+1
Give 5	60	0.62	6.7%	11.7%	13.3%	68.3%
Give 4	60	0.54	5%	10%	33.3%	51.7%
Give 3	60	0.44	0%	11.7%	60%	28.3%
Give 2	60	0.15	0%	40%	46.7%	13.3%
Give 1	60	-0.14	21.7%	43.3%	20%	15%
Give 0	60	-0.53	60%	20%	10%	10%

Table A2.4: Distribution and mean appropriateness rating in Stage 3 in LOW. Modal responses (as identified by the grey cells) for the 6 possible contribution choices are as follows: contributing 5 and 4 tokens are very socially appropriate; contributing 3 and 2 tokens is somewhat socially appropriate, contributing 1 token is somewhat socially inappropriate while contributing 0 is very socially appropriate.

Action	N	Mean	-1	-0.33	+0.33	+1
Give 5	60	0.84	1.7%	1.7%	15%	81.7%
Give 4	60	0.73	1.7%	1.7%	31.7%	65%
Give 3	60	0.42	3.3%	3.3%	70%	23.3%
Give 2	60	0.03	1.7%	48.3%	48.3%	6.7%
Give 1	60	-0.27	15%	63.3%	20%	1.7%
Give 0	60	-0.77	78.3%	11.7%	8.3%	1.7%

Table A2.5: Distribution and mean appropriateness rating in Stage 3 in HIGH. Modal responses (as identified by the grey cells) for the 6 possible contribution choices are as follows: contributing 5 and 4 tokens are very socially appropriate; contributing 3 tokens is somewhat socially appropriate, however 48.3% responses indicate the action of contributing 2 tokens as somewhat socially appropriate while another 48.3% rate it as somewhat socially inappropriate; contributing 1 token is somewhat socially inappropriate while contributing 0 is very socially appropriate.

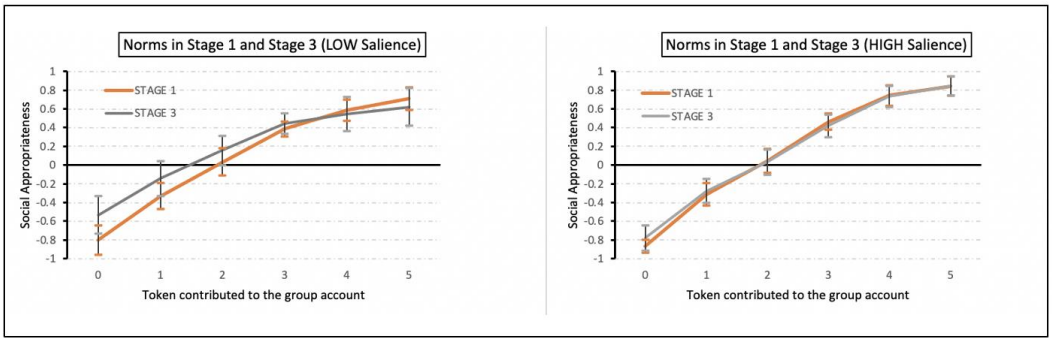


Figure A2.1: Average norm rating within groups. This includes the stage 1 and stage 3 KW elicitations in the 40 groups in LOW and HIGH.

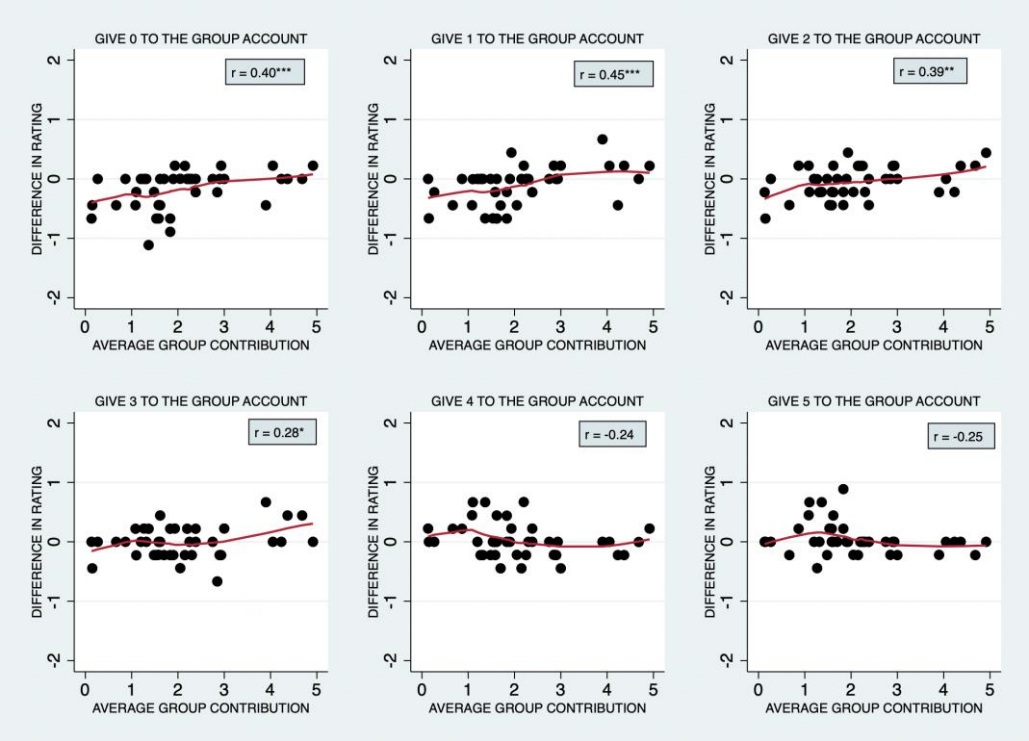


Figure A2.2: Scatterplot of average contribution and difference in norms. The panel shows group average contribution and difference in norm-rating between norm rating between Stage 1 and Stage 3 elicitations for the 6 contribution decisions with 40 groups from LOW and HIGH. 0 difference reflects no difference in average rating, positive difference reflects harsher rating in Stage 3 and negative difference shows weaker judgement in Stage 3. Significance levels: *** p<0.01, ** p<0.05, * p<0.1. Groups that experienced higher level of cooperation in LOW and HIGH, provided a harsher normative evaluation for no and low (giving 0,1,2,3 tokens) contribution choices in Stage 3 compared to Stage 1.

A positive correlation here means that as average contribution increases, the difference in rating becomes more positive as is evidence for the contribution choices 0, 1, 2 and 3 tokens in the group account. More specifically, groups that experienced higher level of cooperation, also elicited slightly harsher rating for no and low contribution choices in Stage 3 compared to their initial Stage 1 ratings. On the other hand, for contribution choices "Give 4" and "Give 5" to the group account, there was a negative relationship between difference in rating and average group contribution. In other words, groups that witnessed lower cooperation in the 20 rounds, gave a harsher rating for high contribution levels (4 and 5 tokens) in Stage 3 compared to Stage 1 KW elicitations, although these associations were not significant.

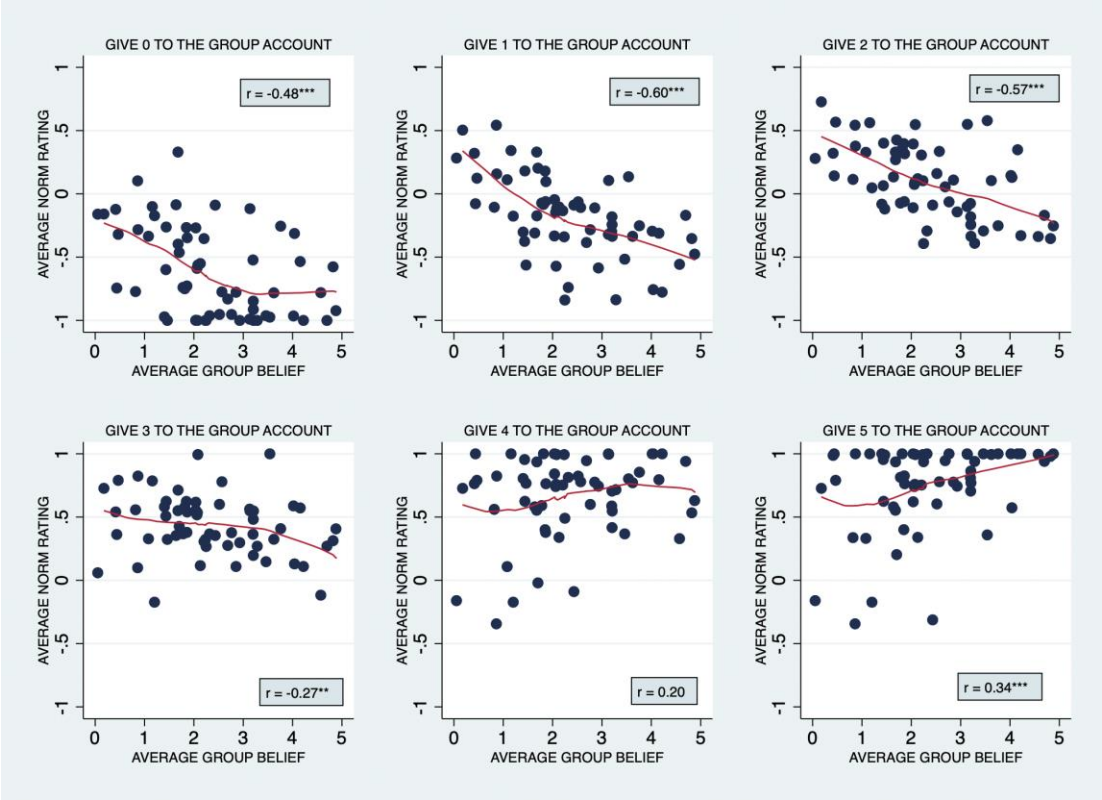


Figure A2.3: Scatterplot between average group belief and stage 3 norms. From top left: negative correlation exists between average group belief and average group norm rating for the contribution choices of 0 token (r (58) = -0.4786, *p* = 0.0001); 1 token (r (58) = -0.6036, *p* = 0.0000); 2 tokens (r (58) = -0.5681, *p* = 0.0000; 3 tokens (r (58) = -0.2713, *p* = 0.036. On the other hand, we find positive correlation between group belief and average norm rating for the contribution choices of 4 tokens (r (58) = 0.0252, *p* = 0.1157) and 5 tokens (r (58) = 0.342, *p* = 0.0075

To see if saliency and the experience in the 20 rounds causally influenced the Stage 3 ratings, we ran OLS regressions of Stage 3 norm ratings in groups from LOW and HIGH as depicted in Table A2.6. Here, columns (1-6) represent the 6 contribution choices as the 6 different dependent variables present in the KW decision environment. For each contribution choice, we regress our model using HIGH, LOW to check for treatment effects and GRAVG CON, which is the average group contribution in the 20 rounds, to see if local experience in the game affected stage 3 norm ratings post-play. We find three main results: first, HIGH and LOW had no effect on the norm rating for any of the six possible contribution choices. However, in HIGH, with declining contribution choices, the rating were more negative, for instance, norm rating for "Give 0" was rated more harshly by 0.204, although this was statistically insignificant

($p = 0.129$). We also find that groups' experience in the repeated game, depicted by group average contribution in 20 rounds (GRAVG CON) influenced Stage 3 ratings: groups that experienced higher cooperation outcomes, provided a harsher judgement for contribution choices of 2 tokens by 0.113 ($p = 0.000$); 1 token by 0.117 ($p = 0.000$) and 0 tokens by 0.089 ($p = 0.012$). In other words, better experience of cooperation led such groups to be more critical of no and low cooperation choices: contributing 0-40% of the given tokens to the public account. On the other hand, there was no causality between average group cooperation and norm rating for contribution choices of 3, 4, and 5 tokens. There was no effect of declining cooperation in the game on the norm evaluation of moderate to full cooperation to the group account.

VARIABLE: Stage 3 KW	(1) Give 5	(2) Give 4	(3) Give 3	(4) Give 2	(5) Give 1	(6) Give 0
HIGH	0.0311	0.0283	0.0220	-0.00664	-0.0758	-0.204
	(0.0825)	(0.0856)	(0.0690)	(0.0728)	(0.0682)	(0.0916)
LOW	-0.157	-0.149	0.0208	0.0593	-0.0151	0.00505
	(0.112)	(0.114)	(0.0713)	(0.0705)	(0.0825)	(0.107)
GRAVG CON	0.0431	0.0163	-0.0486	-0.113***	-0.117***	-0.0890**
	(0.0247)	(0.0313)	(0.0215)	(0.0212)	(0.0274)	(0.0316)
Constant	0.854**	0.756**	0.625***	0.263	0.226	-0.674
	(0.371)	(0.316)	(0.231)	(0.220)	(0.300)	(0.415)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	180	180	180	180	180	180
R-squared	0.119	0.061	0.042	0.167	0.189	0.140

Table A2.6: OLS Regression with stage 2 KW choices in Study 1. Dependent variable is Stage 2 KW contribution choice(s). Standard errors in parentheses, adjusted for intra-group correlation (groups are used as independent clustering units with 60 groups). Significance levels: *** p<0.01, ** p<0.05, * p<0.1. We control for age, gender, ethnicity, background in Economics, and preference for political orientation.

	N	Mean	--	-	+	++
Give 5	90	0.83	0%	4.44%	16.67%	78.89%
Give 4	90	0.70	1.11%	1.11%	38.89%	58.89%
Give 3	90	0.49	1.11%	4.44%	63.33%	31.11%
Give 2	90	0.06	2.22%	40%	53.33%	4.44%
Give 1	90	-0.37	16.67%	73.33%	8.89%	1.11%
Give 0	90	-0.87	81.11%	17.78%	1.11%	0%

Table A2.7: Stage 1 KW ratings in LOW in Study 2. Most participants (modal responses) are as follows: contributing 5 and 4 tokens are very socially appropriate; contributing 3 and 2 tokens are somewhat socially appropriate, contributing 1 token is somewhat socially inappropriate while contributing 0 is very socially appropriate.

	N	Mean	--	-	+	++
Give 5	90	0.72	5.56%	5.56%	14.44%	74.44%
Give 4	90	0.64	0%	8.89%	36.67%	54.44%
Give 3	90	0.39	0%	7.78%	74.44%	17.78%
Give 2	90	0.07	1.11%	46.67%	42.22%	10%
Give 1	90	-0.29	16.67%	64.44%	15.56%	3.33%
Give 0	90	-0.86	83.33%	13.33%	2.22%	1.11%

Table A2.8: Stage 1 KW ratings in HIGH in Study 2. Most participants (modal responses) are as follows: contributing 5 and 4 tokens are very socially appropriate; contributing 3 tokens is somewhat socially appropriate, contributing 2 and 1 tokens are somewhat socially inappropriate while contributing 0 is very socially appropriate.

VARIABLE: Round 1	(1) Belief	(2) Contribution	(3) Contribution
HIGH	0.609***	0.548**	0.0778
	(0.210)	(0.258)	(0.169)
LOW	0.235	0.356	0.175
	(0.191)	(0.259)	(0.187)
Contribution			
Belief			0.772***
			(0.0484)
Constant	2.369***	2.084***	0.255
	(0.504)	(0.603)	(0.501)
Controls	Yes	Yes	Yes
Observations	270	270	270
R-squared	0.079	0.049	0.467

Table A2.9: OLS Regression from Study 2. Dependent variables are belief and contribution in round 1. Standard errors in parentheses, adjusted for intra-group correlation (groups are used as independent clustering units, with 90 groups). Significance levels: *** p<0.01, ** p<0.05, * p<0.1. We control for age, gender, ethnicity, background in Economics, and preference for political orientation.

VARIABLE: Round 1-20	(1) Belief	(2) Belief	(3) Contribution	(4) Contribution	(5) Contribution
HIGH	0.328	0.373	0.372	0.149	0.184
	(0.273)	(0.257)	(0.293)	(0.127)	(0.173)
LOW	0.0265	0.334	0.143	0.125	0.227
	(0.267)	(0.260)	(0.295)	(0.142)	(0.170)
Round	-0.0195***	-0.00833	-0.0403***	-0.0270***	-0.0227***
	(0.00740)	(0.0135)	(0.00732)	(0.00390)	(0.00832)
Round x HIGH		-0.00433			-0.00331
		(0.0170)			(0.00985)
Round x LOW		-0.0293			-0.00967
		(0.0192)			(0.00999)
Belief				0.681***	0.680***
				(0.0329)	(0.0332)
Constant	3.128***	3.010***	2.885***	0.727***	0.687***
	(0.381)	(0.368)	(0.397)	(0.258)	(0.257)
Controls	Yes	Yes	Yes	Yes	Yes
Observations	5,400	5,400	5,400	5,400	5,400
Number of subjects	270	270	270	270	270

Table A2.10: Random effects Regression from Study 2. Dependent variables are average belief and contribution in 1-20 rounds. Standard errors in parentheses, adjusted for intra-group correlation (groups are used as independent clustering units, with 90 groups). Significance levels: *** p<0.01, ** p<0.05, * p<0.1. We control for age, gender, ethnicity, background in Economics, and preference for political orientation.

	N	Mean	-1	-0.33	+1	+3
Give 5	90	0.79	2.22%	5.56%	12.22%	80%
Give 4	90	0.71	0%	3.33%	36.67%	60%
Give 3	90	0.48	0%	2.22%	73.33%	24.44%
Give 2	90	0.06	2.22%	45.56%	43.33%	8.89%
Give 1	90	-0.25	17.78%	55.56%	23.33%	3.33%
Give 0	90	-0.79	74.44%	22.22%	1.11%	2.22%

Table A2.11: Distribution of Stage 3 norms in NO, Study 2. Modal responses (as identified by the grey cells) for the 6 possible contribution choices are as follows: contributing 5 and 4 tokens are very socially appropriate; contributing 3 tokens is somewhat socially appropriate, contributing 2 and 1 token are somewhat socially inappropriate while contributing 0 is very socially appropriate.

	Obs	Mean	-1	-0.33	+0.33	+1
Give 5	90	0.78	3.33%	4.44%	13.33%	78.89%
Give 4	90	0.70	3.33%	2.22%	30%	64.44%
Give 3	90	0.50	3.33%	2.22%	58.89%	35.56%
Give 2	90	0.11	1.11%	40%	50%	8.89%
Give 1	90	-0.35	18.89%	68.89%	8.89%	3.33%
Give 0	90	-0.79	77.78%	16.67%	3.33%	2.22%

Table A2.12: Distribution of Stage 3 norms in LOW, Study 2. Modal responses: contributing 5 and 4 tokens are very socially appropriate; contributing 3 and 2 tokens are somewhat socially appropriate, contributing 1 token is somewhat socially inappropriate while contributing 0 is very socially appropriate.

	N	Mean	-1	-0.33	+0.33	+1
Give 5	90	0.70	6.67%	4.44%	15.56%	73.33%
Give 4	90	0.63	2.22%	4.44%	40%	53.33%
Give 3	90	0.37	0%	7.78%	78.89%	13.33%
Give 2	90	-0.02	4.44%	51.11%	37.78%	6.67%
Give 1	90	-0.34	21.11%	63.33%	11.11%	4.44%
Give 0	90	-0.84	81.11%	15.56%	2.22%	1.11%

Table A2.13: Distribution of Stage 3 norms in HIGH, Study 2. Modal responses: contributing 5 and 4 tokens are very socially appropriate; contributing 3 tokens is somewhat socially appropriate, contributing 2 and 1 tokens are somewhat socially inappropriate while contributing 0 is very socially appropriate.

As in Study 1, we looked into the effect of HIGH, LOW, and average contribution, i.e., experience in the game on post-play norms and found similar results: LOW and HIGH (other than effect on 3 tokens) had no effect on the norm rating of the contribution choices. Additionally, groups that experienced higher cooperation outcomes than others, irrespective of the treatment they belonged to, provided a harsher evaluation for contribution choices of 0 token by a reduced average of 0.048, (p = 0.048) 1 token by 0.104 (p = 0.000) and 2 tokens and -0.102 (p = 0.000). This indicated that better cooperation led to stronger judgment of low contribution and free riding.

VARIABLE: Stage 3 KW	(1) Give 5	(2) Give 4	(3) Give 3	(4) Give 2	(5) Give 1	(6) Give 0
HIGH	-0.108	-0.0897	-0.0979**	-0.0442	-0.0504	-0.0385
	(0.0906)	(0.0700)	(0.0481)	(0.0686)	(0.0789)	(0.0509)
LOW	-0.0277	-0.0131	0.0381	0.0660	-0.0905	-0.00427
	(0.0755)	(0.0614)	(0.0545)	(0.0618)	(0.0638)	(0.0620)
GRAVG CON	0.0475**	0.0281	-0.0290	-0.102***	-0.104***	-0.0480**
	(0.0217)	(0.0221)	(0.0194)	(0.0206)	(0.0227)	(0.0196)
Constant	0.971***	0.802***	0.587**	0.197	-0.218	-1.155***
	(0.317)	(0.247)	(0.235)	(0.144)	(0.205)	(0.179)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	270	270	270	270	270	270
R-squared	0.062	0.022	0.050	0.108	0.101	0.078

Table A2.14: OLS Regression with norms, Study 2. Dependent variable is Stage 2 KW ratings for the 6 contribution choices. Standard errors in parentheses, adjusted for intra-group correlation (groups are used as independent clustering units, with 90 groups). Significance levels: *** p<0.01, ** p<0.05, * p<0.1. We control for age, gender, ethnicity, background in Economics, and preference for political orientation.

VARIABLE: Round 1-20	(1) Belief	(2) Contribution	(3) Contribution
HIGH	0.490**	0.432*	0.110
	(0.221)	(0.233)	(0.104)
LOW	0.107	0.111	0.0416
	(0.217)	(0.227)	(0.110)
Yr20	0.386**	0.483**	0.229**
	(0.186)	(0.193)	(0.0922)
Round	-0.0333***	-0.0545***	-0.0327***
	(0.00584)	(0.00543)	(0.00332)
Belief			0.657***
			(0.0250)
Constant	2.402***	2.177***	0.556**
	(0.328)	(0.343)	(0.234)
Controls	Yes	Yes	Yes
Observations	9,000	9,000	9,000
Number of subjects	450	450	450

Table A2.15: Random effects Regression with pooled data. Dependent variables are average beliefs and contributions across rounds 1-20 in the PGG, pooled from study 1 and study 2. Standard errors in parentheses, adjusted for intra-group correlation (groups are used as independent clustering units, with 150 groups). Significance levels: *** p<0.01, ** p<0.05, * p<0.1. We control for age, gender, ethnicity, background in Economics, and preference for political orientation.

CHAPTER 3 Methods in rule following

3.1 Abstract

Rule following is a social process (Gächter et al. 2021), where compliance is driven by what others do and expect to be done. Gächter et al. (2021) use a simple rule-following decision environment to show that behaviour is subject to strong peer effects: when subjects are shown the real behaviour of peers who take part in the same task, the decision of whether to comply or not with the rule depends on what the peers do. If peers violate, subjects are more likely to violate. If peers comply, subjects are slightly more likely to comply. In this paper, we are interested in testing the effectiveness of two different methods to elicit this type of conditional behaviour. These methods do not require real-time information about what peers do, making them simpler and more efficient in carrying out such types of elicitation tasks. They are a variant of the strategy method (S-method) adapted from Gächter et al. (2021) and the direct approach (D-method) based on anchoring procedures, adopted from Abeler et al. (2019). Between the two methods, the S-method captures subjects' preference to violate rules conditional on increasing rate of others' violation outcomes. Additionally, this method is also useful to observe preference heterogeneity in subjects with rule-compliance and elicit types of decision makers.

3.2 Introduction

Social norms are behavioural patterns that define the informal shared rules of a society. As Elster (1989) pointed out, the simplest rule is of the type "Do X", or "Don't do X". Norm driven behaviour is conditional on two main types of societal beliefs, which are i) what most people do (empirical beliefs), ii) what most people approve of being done (normative beliefs). If the behavioural outcome is a result of both types of societal beliefs, i.e., empirical and normative beliefs, then it is said to be a social norm (Bicchieri, 2005; 2016) Thus, norm-compliance requires people's preference to conform, conditional on such social beliefs.

This chapter presents a methodological study where the main objective is to test different methods to measure the extent to which norm compliance behaviour is conditional on empirical beliefs. The decision environment consists of a simple rule-following task, adapted from Gächter et al. (2021). This involves an individual decision-making task, where the decision maker can control a circle figure. On a single click, the circle figure moves from the left end of the animated screen and stops at the centre, on a grey bar, at the top of which is a cross (X). Subjects are told that the rule is to "wait in the grey area until the X disappears". The decision-maker is initially endowed with 20 experimental points and with each passing second, his points decline by 1 point. In other words, waiting for the cross to disappear and following the rule is costly. The decision-maker can choose to either wait till the cross disappears and follow the rule, earning less points, or to move before the cross disappears and violate the rule, earning more points.

To test if the preference to follow or violate a rule is conditional on social beliefs, we use two experimental methods. In both cases, we study conditionality with respect to empirical

information. We focus on empirical beliefs because, in an experiment with peer effects, Gächter et al. (2021) have shown that rule compliance in the rule-following task is conditional on others' behaviour. Thus, we have a benchmark situation in which we expect to observe preferences for compliance that are conditional on empirical information. Our two methods are as follows.

Firstly, we use a version of the strategy method adapted from Gächter et al. 2021 (S-method henceforth). Here we inform subjects of four possible scenarios that differ in the proportion of other people who break the rule. These scenarios are a) 0-25% other MTurkers broke the rule (LOW violation information); b) 26-50% other MTurkers broke the rule (MID LOW violation); c) 51-75% other MTurkers broke the rule (MID HIGH violation) and d) 76-100% other MTurkers broke the rule (HIGH violation). In each scenario, we ask subjects to provide a response between following the rule and breaking the rule.

However, a key issue with the strategy method is that is has been considered as a "psychologically cold" method of preference elicitation (Brandts & Charness, 2011). Moreover, subjects may feel compelled to condition their response to information they are provided with within-subject designs (Zizzo, 2009). To investigate preference conditionality more realistically and to circumvent the issues with within-subject design of the S-method, we also use a direct-elicitation method (D-method henceforth), adapted from Abeler et al. (2019), based on an anchoring approach (Tversky & Kahneman, 1974). Here subjects are shown possible graphical representations that also vary in the proportion of other people who break the rule from a hypothetical study that serve as "informational anchors" after which they take part in the rule-following task in real time. To keep the fraction of violation rates comparable with the S-method, the D method involves four between-subject treatment groups of increasing

violation rates, namely LOW, MID LOW, MID HIGH and HIGH. Both studies were programmed using LIONESS (Giamattei et al., 2020) and carried out using Amazon MTurkers from the USA.

We found the following result from the S-method: there was an increase in the average violation across the scenarios with increasing proportion of other MTurkers' violations. The average violation rate was 28% when asked to state their preference in a scenario where 0-25% of previous MTurkers broke the rule (LOW); 32% in a scenario where 26-50% broke the rule (MID LOW); 62% where 51-75% broke the rule (MID HIGH) and 70% where 76-100% broke the rule (HIGH). Additionally, we found that the increase in the preference to violate a rule was a function of the increasing information of rule-violation. The regression estimates show that a 1% increase in violation increased the elicited preference to violate by 0.62%. This was higher than that of 0.19% from Gächter et al (2021). This finding is indicative of preference conditionality or subjects' preference towards conformity with others' (increasing) violation outcomes.

Additionally, the strategy method provides a measure of the subject's preference schedule (to follow or violate a rule) for different possible rates of (increasing) violation around them. We found four main types of decision makers namely: i) unconditional rule followers, ii) unconditional rule-violators, iii) conditional rule-followers/violators and iv) unclassified type. 18% were unconditional rule-followers who always followed the rule. 14% were unconditional rule violators who always broke the rule. 46% were conditional rule-followers/violators: they first elicited a preference for compliance when presented with the scenario of LOW violation (0-25% broke the rule), but then they switched to violation when presented with a higher violation rate. Some of these conditional violators switched to breaking a rule with MIDLOW

level of violation, reflecting a high sensitivity to empirical information of violation. They maintained this preference with MID HIGH and HIGH. Some others made a similar switch from compliance to violation in MID HIGH and maintained this in HIGH. Few others switched only in HIGH reflecting a low(er) sensitivity to empirical information of violation. Additionally, 22% were considered unclassified. These were based on responses that chose to sometimes follow and sometimes break the rule in no particular order. This finding reflects that most individuals display conformity with rule-following behaviour with varying degrees of norm-sensitivity.

We found some interesting results from the D method: after informing subjects of the "possible outcomes from a potential study", we elicited their empirical belief (of rule violation), to check if the anchoring manipulation worked. We found an increase in the elicited belief of violation across the treatment groups: 30% in LOW, 42% in MID LOW, 48% in MID HIGH and 63% in HIGH, and this trend was significant. Thus, the informational anchors significantly shifted beliefs. However, following this, the D-method failed to elicit the positive conditionality in behaviour in the rule-following task. In fact, with increasing information of rule-violation across the treatment groups, subjects displayed a decreased preference to violate the rule. The violation rates from the rule-following task were 51%, 58%, 38% and 34% across LOW, MID LOW, MID HIGH and HIGH. The regression estimates show that a 1% increase in violation decreased violation by 0.28%.

Between the two elicitation methods, the S-method captured subjects' preference conditionality as observed in previous research (Gächter et al., 2021) on rule-violations accurately. Here, subjects were shown the empirical information (of increasing fraction of other MTurkers who broke the rule) the same time as they were asked to elicit their preference

between following a rule and breaking a rule. This design informed the subjects about other MTurkers' behaviour and also helped them update their normative beliefs of rule-violations. When subjects "witness" more rule violation, they likely infer that most people around them approve of such violation, and as a result they shift (lower) their normative beliefs of violation (Bicchieri et al., 2020). Both the information of increasing rule-violation and the (inferred) updated normative belief of the appropriateness of such violation led to the increase in the elicited preference to violate rule. This mechanism is similar to the real-time peer decisions in Gächter et al. (2021). On the other hand, although there was a shift in the empirical beliefs in the D-method, this was not long-lasting enough to impact the behaviour in the rule following task.

The organization of the chapter is as follows: Section 3.3 discusses the related literature and our contribution. This is followed by section 3.4 that outlines the experimental design. Section 3.5 provides the results, and Section 3.6 discusses and concludes.

3.3 Related Literature and our contribution

Peer-effects have been studied extensively in the Social Sciences to understand how contagious good and bad behaviour are. Studies have found that such peer-effects drive the selfish outcomes in cooperation dilemmas (e.g., de Oliveira et al., 2014; Fischbacher & Gächter, 2010); dishonesty (e.g., Gino et al., 2009; Innes & Mitra, 2013); littering (Cialdini et al., 1990; Keizer et al., 2008); and corruption (Dong et al., 2012; Schram et al., 2022) outcomes. Both negative and positive peer effects exist, but negative effects are stronger (e.g., Baumeister et al., 2002; Bicchieri et al., 2022; Croson & Shang, 2008; Dimant, 2019; Gächter et al., 2021). Recent evidence points to such conformity attitude being a domain-general behavioural trait, reflected through honesty and cooperation games (Isler & Gächter, 2022).

Gächter et al. (2021) found evidence of people's preference to comply or violate in a simple rule-following task functional on peers' violation decisions around them. In their study, they assigned subjects randomly to one of the 28 treatment groups. One of these was the baseline group, where subjects made a decision to either follow or violate a rule without any additional information of how others behaved. In the other 27 treatment groups, subjects could see other circles figures, in addition to their own. These figures were reflective of the movement of "peers" or actual other previous participants who took part in a similar task before. The multiple treatment groups in their design allowed them to investigate how varying compliance rates and group size influenced people's rule-compliance behaviour.

Their study reflected a few interesting findings. They found an asymmetric peer effect on violation outcomes. More specifically, they found that when subjects observed full compliance i.e., when all "peers" followed the rule, there was an insignificant increase in rule-compliance compared to the baseline group. On the other hand, when subjects observed full violation i.e., when all peers violated a rule, subjects significantly violated the rule themselves. When subjects observed a combination of both violation and compliance outcomes by peers, they were more influenced by the negative information of violation, as was evident by the decrease in compliance compared to the baseline group. Their regression estimates showed that a 1% increase in violation by other peers significantly increased violation by 0.19%.

Their results provide proof of preference conditionality through peer effects, specifically with regards to rule violation. When subjects were exposed to the information of increasing size of violation by other peers, they preferred to violate more too. This aligns with the ideas of "social proof" (Cialdini, 1993) and "rational herding" theories (e.g., Banerjee, 1992; Bikhchandani et

al., 1998). Although the results convey meaningful explanations, this method of extracting subjects' preferences is slightly impractical. It involved a two-step design where first they collected data from actual subjects which they then used as peer's behavioural responses in real-time in the rule following task with animated circle figures. Additionally, to capture the effect of varying size of peers' violation on subjects' behavioural decisions, the design consisted of multiple treatment groups which is potentially costly from the viewpoint of data collection.

In this chapter, we are interested in testing two methods to elicit this type of conditional behaviour, without using real-time information about what peers do, which may be easier to use in experimental and survey applications. We test the effectiveness of two other methods namely the S-method adapted from Gächter et al. (2021) and D-method adapted from Abeler et al. (2019) in capturing conditionality in rule-violation. The S-method employed in this study has been considered helpful for various reasons. Firstly, this is cost-effective, being a within-subject elicitation method. Moreover, it provides contingent responses for all possible scenarios of a decision environment from the same individuals and thus seems highly informative. However, it does suffer from some potentially critical drawbacks. For instance, it has been considered a cold treatment (e.g., Brandts & Charness, 2011; Roth, 1995) as it involves the presence of only hypothetical situations, and thus, may be ineffective in invoking any emotions that may be otherwise needed for the decision-environment. This raises external validity concerns. Additionally, the framing of the decision-environment may also attribute to the experimenter-demand effect (EDE) and thus potentially lead to biased outcomes. EDE(s) are the "changes in behaviour by experimental subjects due to cues about what constitutes appropriate behaviour" (Zizzo, 2009, pp 75). This can be a valid problem, mainly when the outcomes dictated by such effects are positively associated with the predicted outcomes of the

study. The strategy method involves a within-subject design, and the order or sequence of questions may lead to the potential dangers of EDE. Additionally, this elicitation method requires decisions for all possible subgames, not just those resulting from the actual gameplay, which might lead to the elicitation of biased preferences.[31]

To counter these issues, direct elicitation methods have also been extensively used to elicit individual preferences[32]. The between-subject design of such methods reduces the risk of EDE. Additionally, these methods are more involving and have the ability to induce visceral factors that are fundamental in the decision-making process. In fact, some studies employ both strategy and direct procedures to investigate effects on preferences based on the presentation of the choice environment.[33] The overall conclusion regarding the equivalence of preferences elicited using these methods across different games remains mixed.

We employ a variant of the direct method elicitation by adopting a newer approach used by Abeler et al. (2019) which relies on the idea of "anchoring" (Tversky & Kahneman, 1974). Abeler et al. (2019) criticize the use of empirical information based on the actual behavioural outcomes of subjects from previous studies, as the distribution of this information may be highly selected and not necessarily representative of the general population. In their study involving eliciting truth-telling preferences, they provided subjects information from a "potential" study and then ask them to imagine the possible outcomes from the said study.

[31] In an ultimatum game experiment, Bazerman et al. (1999) found that second movers accepted more unequal offers if they had to decide for each potential offer, as opposed to stating the minimum threshold for an offer to be acceptable from their end. Moreover, there have been issues with regards to timing and order of play in such an elicitation method (Güth et al.,1998; Rapoport & Fuller, 1998; Weber et al., 2004)

[32] See Brandts and Charness (2011) for a comparative review between strategy and direct elicitation methods.

[33] Some studies find no significant differences in elicited preferences with either the strategy or the direct method: (Cason & Mui., 1998; Brandts & Charness., 2000; Fischbacher et al., 2012) while others do (e.g., Schotter et al., 1994; Brosig et al., 2003., Weber et at., 2004; Güth et al., 2001).

Following this they elicited subjects' empirical beliefs and found a shift in stated beliefs which showed that the anchoring manipulation worked.

This is a useful method as it involves a between-subject design and thus avoids the issues of EDE from the S-method. Moreover, it appears more practical than the real-time information with peers from Gächter et al. (2021) as it captures the effect of varying rates of rule-violation with just 4 treatment groups.

This study contributes to the experimental literature investigating methods in preference elicitation in the context of rule compliance. Firstly, we elicit conditional preferences using the S-method in the framework of a simple rule following task and find comparable results to the study by Gächter et al. (2021). Additionally, we uncover the types of individuals and their proportion in the population, with most subjects being conditional violators. The sequencing of the task with the elicitation of contingent responses to increasing violation rates helps visualise subjects' sensitivity to rule compliance. Within these conditional violators, while some shift to the selfish outcome at low levels of others' violations, others need a very high violation rate to prefer this selfish outcome themselves.

Hence, the S-method is useful in capturing the notion of subjects' preference to violate the rule conditional on others' behaviour. We also employ the D-method, looking at the potential risks of the S-method. However, this elicitation method fails to reflect preference conditionality with violation outcomes in the rule following task. Based on our findings, we suggest using the S-method to elicit such conditional preferences in other experimental contexts too. This is a cost-effective method, and the design allows to isolate and look into the influences on subjects' preferences from empirical information alone, which is beneficial for proof-of concept studies.

Moreover, the within-subject design allows us to implement a single experiment instead of running multiple treatment conditions, unlike the D-method and the study with real-time peer effects.

3.4 Experimental setup and design

3.4.1 Decision Environment: The rule following task

The decision environment we use in both preference elicitation methods is the rule following task, introduced by Gächter et al. (2021). This is a simple individual decision-making task where the subject controls a circle figure (see Figure 3.1), placed at the left side of the screen. At the beginning of the task, a subject is given 20 endowment points. This is shown at the top of the screen. Each endowment point is $0.05, and so 20 points = $1. On clicking the Move button for the first time, the circle figure moves from the left side of the screen to the grey bar at the centre of the screen and stops there. At the same time, with each passing second in the task, the endowment declines by 1 point. In the instructions, subjects are told that the rule is to "wait in the grey area until the cross (X) disappears". On the second click, the circle figure moves from the grey bar into the black bar on the right side of the screen. This is the end of the task.

Waiting for the cross to disappear leads to the loss of more endowment points: the longer a subject waits, the more points he/she loses. In other words, following the rule is costly. The earnings from this task are determined by the points remaining at the end of the task. For instance, if a subject waited for the cross to disappear and completed the task in 15 seconds,

then he or she would earn 20 – 15 = 5 * $0.05 = $0.25[34]. Gächter et al. (2021) use this decision environment in real time[35] to show that behaviour is subject to strong peer effects.

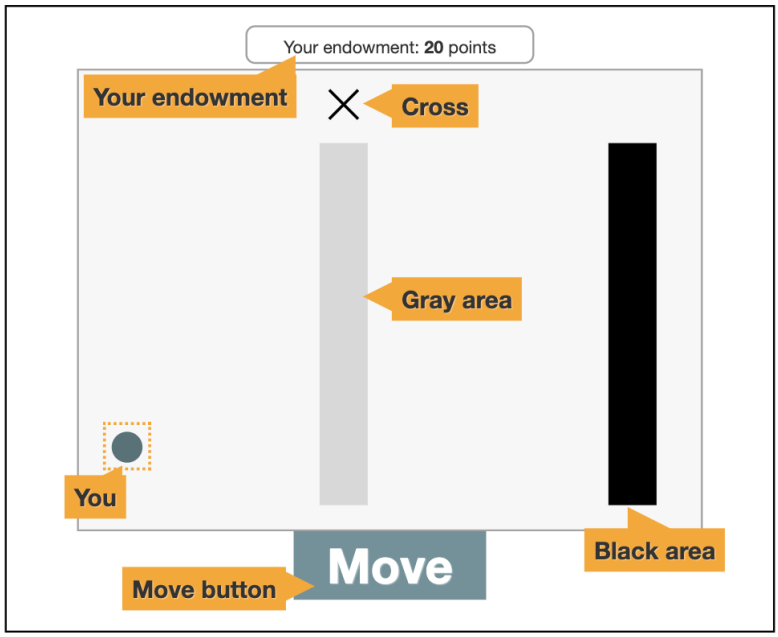


Figure 3.1: Screenshot of the decision environment of the rule following task.

Our experiments were programmed using LIONESS (Giamattei et al., 2020) and carried out on Amazon MTurk, with n = 250 US MTurkers. Subjects or MTurkers first got access to HIT(s). These are human intelligence tasks, which comprise short surveys, experiments, or studies. After accepting our HIT, subjects were directed into the LIONESS interface to participate in the tasks.

3.4.2 Study 1: S-Method

Subjects were first shown the decision environment of the rule following task (Figure 3.1) and informed of the rule of this decision task. They were also told about the scenarios of complying

[34] In the actual task, the cross would disappear after 12 seconds. Therefore, any click after 12 seconds would be considered a compliance outcome, while any click before 12 seconds would be considered a violation outcome.
[35]By real-time, we mean that the peer behaviour is shown to a subject to reproduce what a peer actually did in a previous iteration of the experiment.

and violating the rule and the potential points they could earn from these outcomes. Specifically, it was mentioned that "if you wait until the cross disappears, it will take your circle 11 seconds to reach the black area, and your earnings will be 9 points. If you move before the cross disappears, your circle will take 5 seconds to reach the black area, and your earnings will be 15 points. After this, subjects were informed of a hypothetical study which reads as follows:

*"Over **100** MTurkers from the USA previously participated in a similar HIT as the one you are participating in today. These MTurkers also could choose whether to wait until the cross (X) disappears or to move before it disappears. In today's HIT, we will show you four possible outcomes of this previous HIT. We will ask you, for each outcome, whether you want to wait until the cross disappears or to move before it disappears".*

The four possible outcomes were in increasing size of previous MTurkers who broke the rule and moved before the X disappeared. The scenarios were i) Between **0-25%** of previous MTurkers moved before the cross disappeared (LOW) ii) between **26-50%** of previous MTurkers moved before the cross disappeared (MID LOW) iii) between **51-75%** of previous MTurkers moved before the cross disappeared (MID HIGH) and iv) between **76-100%** of previous MTurkers moved before the cross disappeared (HIGH). This served as our empirical manipulation with increasing violation rate.

For each possible scenario, subjects could choose between two possible actions: wait in the grey area until the cross disappears and follow the rule or move before the cross disappears and violate the rule. Thus, subjects in the S-method did not actually participate in the rule-following task described previously. Instead, we elicited a subject's conditional preference to either

follow or break a rule across four scenarios. These scenarios differed in the fraction of the previous MTurkers who broke the rule and moved before the cross (X) disappeared. Therefore, we elicited four responses from each subject based on an increasing violation rate.

Before starting the task, we also informed the subjects about the normative belief of rule violation, based on a previous study with another set of MTurkers (Gächter et al., 2021). More specifically, we stated the following "We also ran a version of the study where we asked a different group of MTurkers from the USA to tell us how appropriate they think it is to move before the cross disappears. 82% told us that moving before the cross disappears is inappropriate". This was to ensure that the normative beliefs were fixed and did not influence the preference outcomes. In other words, the aim was to observe the effect on the subject's preference to comply with a rule conditional on varying degrees of empirical information (of violation) alone.

The preference elicitation was incentive compatible. We told subjects that one out of the four possible responses would be used for computing the bonus earnings. The bonus earnings were computed as follows: at the end of the study, the subjects were informed of the actual outcome. i.e., the percentage of previous MTurkers who broke the rule and moved before the cross (X) disappeared. Their choice of a response corresponding to this scenario would be then used to compute the bonus earnings. The actual violation rate was 37% in the rule-following task from Gächter et al. (2021) study. As a result, subjects' responses from the scenario: "between 26-50% of previous MTurkers moved before the cross disappeared" or MIDLOW would be used to compute the bonus earnings. If a subject's response were to wait and follow the rule in this scenario, they would be paid 9 points. On the other hand, if he/she decided to move and violate

the rule, they would be paid 15 points. [36] A screenshot of the preference elicitation task is shown in Figure 3.2 below.

Based on the results from the study on peer-effects (Gächter et al., 2021), we expect an increase in the average elicited preference to violate. Specifically, we expect an increase in the proportion of elicited violation preference as we move across the four scenarios that vary in the fraction of previous MTurkers that moved before the cross disappeared.

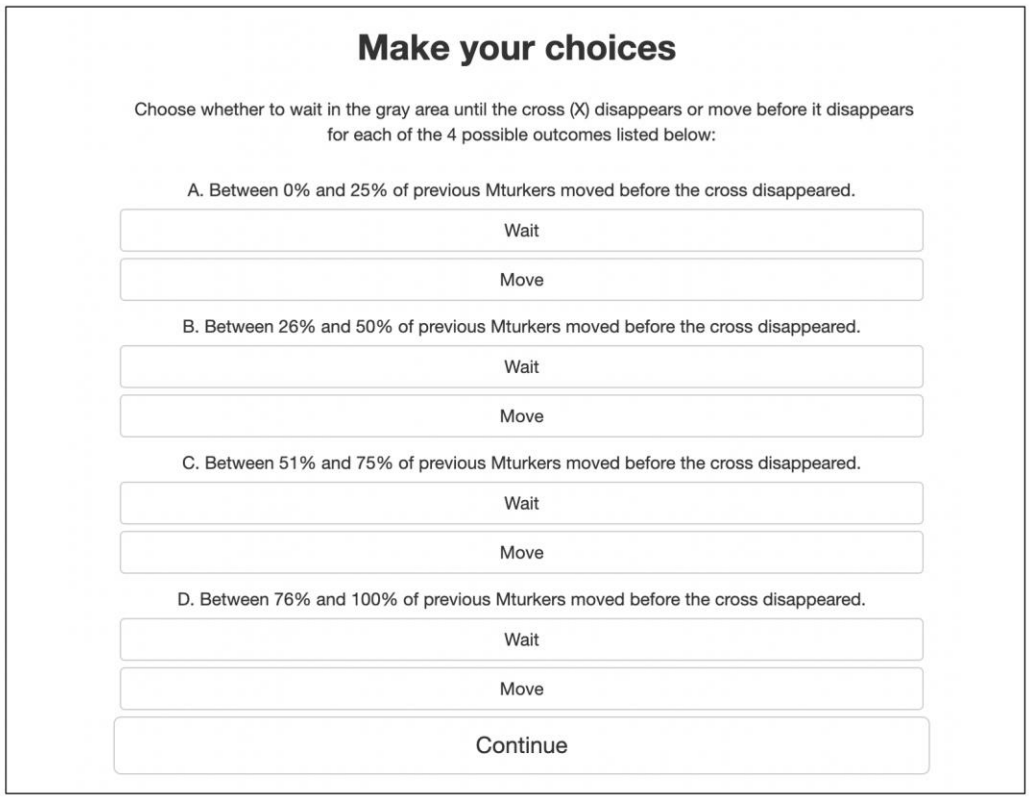


Figure 3.2: Screenshot of the S-method. The four scenarios show the increasing empirical information of others' (other MTurkers) violation rates. For each scenario, subjects could make a choice between Wait (and follow the rule) and Move (and violate the rule).

[36] This was based on the study by Gächter et al (2021), with the rule-following task with a different sample of MTurkers. Based on the results, most subjects who followed the rule completed the task in 11 seconds and earned 9 points. On the other hand, most subjects who broke the rule completed the task in 5 seconds and earned 15 points. We used these modal earnings as a reference for the incentives in the S-method.

3.4.3 Study 2: D-Method

This method involved a between-subject design with four treatment groups of increasing empirical information of rule-violation, namely LOW, MID LOW, MID HIGH and HIGH, comparable to the S method. It is based on the idea of anchoring (Tversky & Kahneman, 1974), which helps establish an expected standard in situations of vagueness or ambiguity.

As in the previous case, subjects were first familiarised with the decision environment of the rule following task. Subjects were also informed of the potential earnings from the previous study's possible compliance and violation outcome[37]. This was to ensure that the information given to the subjects from the decision environment of the task stays comparable between the two treatments. Subjects were then asked if they understood the task. If they answered "yes", they could move to the next stage. Here they were asked to imagine different possible outcomes from a hypothetical study. A screenshot from one of the treatment groups is shown in Figure 3.3 below[38].

In LOW, subjects were asked to imagine behaviour where in one case only 8 and in another case, 17 out of 100 violated the rule from the rule-following task. These imaginary behavioural outcomes were shown using graphical representations as shown in Figure 3.3. In MID LOW, subjects were shown scenarios where 33 and 42 MTurkers (out of 100) violated the rule. In MID HIGH, subjects were shown scenarios where 58 and 67 (out of 100) violated the rule. Lastly, in HIGH, subjects saw scenarios where 83 and 92 (out of 100) violated the rule[39].

[37] This was based on the responses of the control questions from the previous study: Subjects here were shown the correct responses to the control questions from the previous study that was informative of the points earned from compliance and violation: if they moved after the cross disappeared and finished the task in 11 seconds, they would earn 9 points, and if they moved before the cross disappeared and finished the task in 5 seconds, they would earn 15 points.

[38] The screenshots of the other treatment groups and subject responses from the survey questions are provided in the Appendix as figures A3.1, A3.2 and A3.3

[39] These numbers were picked to keep the data-analysis between the methods comparable: the mid-values of the averages of these numbers equalise with the mid values of the four fractions of violation rate in the S-method.

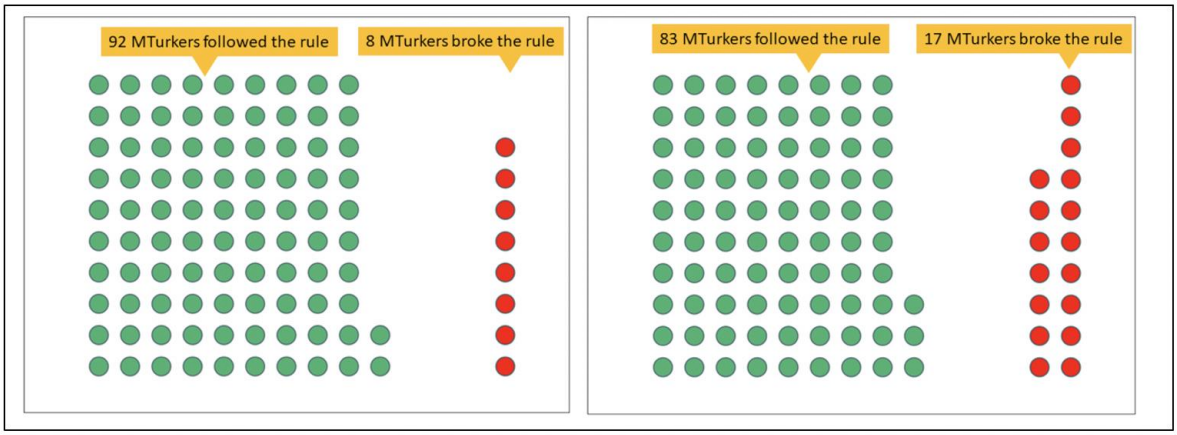


Figure 3.3: Screenshot from D-method of LOW treatment group. These graphical representations show "possible outcomes" from a potential study. The green circles show rule followers, and the red circles show rule violators. These figures were shown one after the other. Subjects were first shown the figure on the left and asked a few questions, and then shown the figure on the right and asked a few more questions regarding their thoughts on these rule followers and violators. This served to anchor their empirical beliefs of rule-violation.

After each visual illustration, subjects were asked a few questions to elicit their thoughts about the possible outcomes from the hypothetical study. More specifically, in LOW, subjects were asked:

"8 MTurkers broke the rule and moved before the cross had disappeared. Can you imagine why they would do that?"

"92 MTurkers followed the rule and moved after the cross had disappeared. Can you imagine why they would do that?"

Additionally, they were asked to think about how satisfied the MTurkers must have felt who broke the rule and the MTurkers who followed the rule. They could provide their response on a 5- point scale ranging from very dissatisfied to very satisfied. The same protocol was followed in the other treatment groups as well.

These mid-values are relevant for the regression analysis to show the conditional relationship between the elicited response and empirical information of rule-violation.

After these steps were carried out, subjects from all treatment groups were told that a similar study had already been carried out using a different group of MTurkers. They were then asked to estimate the fraction of subjects they thought would have moved before the cross had disappeared. In other words, subjects were asked to elicit their empirical belief of rule violation in the rule-following task. They were told that if their estimate is accurate, i.e., if their response lies within +/- 3 percentage points, then they would receive an additional bonus of $0.50 at the end of the task. Thus, subjects were incentivised to provide accurate beliefs of rule violation. This was to ensure whether our anchoring manipulation was successful in shifting subjects' empirical beliefs of rule violation.

After this, subjects were directed to the main task. A visual illustration of the task is shown in Figure 3.4 below. In the beginning, there is a reminder of the rule of the task in a pop-up window. After reading the rule, subjects could click on OK (see A in the panel), after which the pop-up message disappeared, and subjects would be allowed to start the game (See B in the panel). On the first click on the Move button, the circle moves into the grey area (C in the panel). On the second click, it moves from the grey area to the black area on the left side of the screen, and this is the end of the task (D in the panel). Subjects would lose endowment points with each passing second in the task. In other words, following the rule is costly.

Before the start of the task, subjects were informed of the normative information, as in the S-method. This was again to make sure that a normative standard was fixed, and the decision outcome was conditional on the empirical information manipulation alone. After all the information was given to them, they participated in the rule-following task. Each subject played the task once only.

Subjects were paid additional bonus incentives based on belief elicitation and their decision in the task. Note that in the rule-following task, subjects started with 20 points (20 endowment points =$1) and with each passing second, points declined by 1 point.

Following the evidence of preference conditionality from the results from Gächter et al (2021), we expect an increase in the proportion of violation outcome in the rule-following task with increasing information of others' (previous MTurkers) violations. Since the D-method incorporates a between-subject design, we expect a higher proportion of violation decisions from subjects as we move across the treatment groups from LOW, MID LOW, MID HIGH and HIGH.

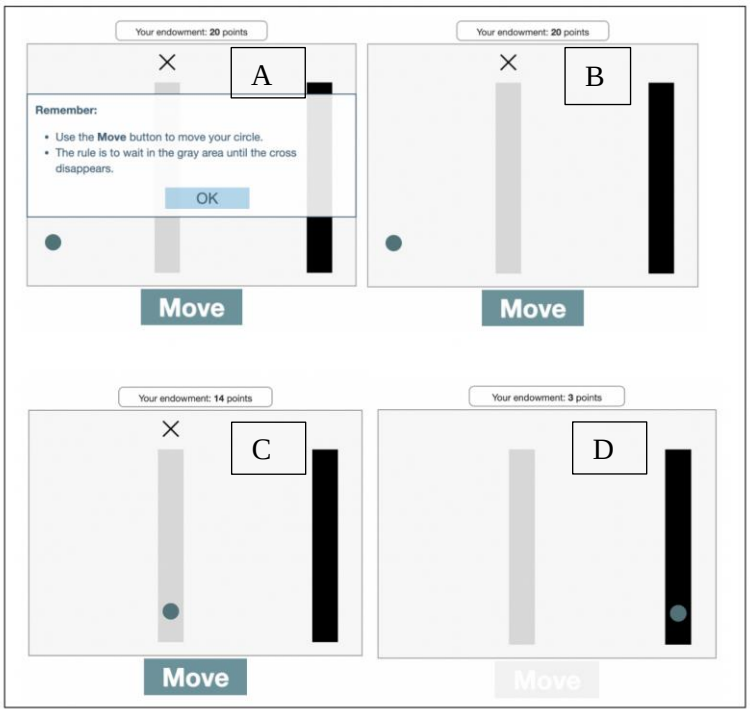


Figure 3.4: Timeline of the rule following task. (A) shows the initial decision environment, which reminded subjects of the rule of the task, and this reminder message disappeared within a few seconds, as shown in (B). On the first click on the Move button, the circle figure moved from the left side into the grey area and stopped there. In this illustration, the subject followed the rule and waited in the grey area until the X disappeared. After this, on the second click on the Move button, the circle moved from the grey area into the black area on the right side of the screen. This was the end of the task.

3.4.4 Procedure

We recruited 50 MTurkers for the S method and 200 MTurkers for the D-method (all from the USA). We had 50 subjects per treatment (LOW, MID LOW, MID HIGH and HIGH) in the D-method, to keep the number of subjects consistent across the two methods. The overall demographic information of these MTurkers is given in Table 3.1 below. Col (1) reflects the information from the S-method, and Col (2) - Col (5) display the information from the four treatment groups of the D-method. At the beginning of each task, subjects were informed of the show-up fee, the potential to earn more points in the tasks in the study, and the approximate duration of the study. At the end of the task for both S and D-methods, subjects filled up a brief questionnaire stating their age, gender, education and income level, weekly earnings from Mturk and political orientation. We also asked for their feedback about the study, especially regarding the ease of understanding the rules on a scale of 1-10.

	S-method (1)	D-method (2)	D-method (3)	D-method (4)	D-method (5)
Characteristics		LOW	MID LOW	MID HIGH	HIGH
Average Belief	N. A	30	42	48	63
Average age	33	34	34	35	33
Male %	70%	60%	58%	50%	67%
White %	62%	67%	70%	75%	81%
Bachelor's degree	44%	51%	62%	65%	55%
Annual income: $10000-$60000	54%	67%	60%	65%	65%
More than $20 (weekly earnings)	48%	55%	52%	48%	53%
Number of subjects (N)	50	50	50	50	50

Table 3.1: Demographic characteristics of the MTurkers from all treatment groups

3.5 Results

3.5.1 S-Method

We first report subjects' average violation preference rate across the four scenarios of increasing violation (by previous MTurkers) from the S-method. The average violation rate was 28% when asked to state their preference in a scenario with LOW violation (where 0-25% previous MTurkers broke the rule); 32% in a scenario with MID LOW violation (where 26-50% broke the rule), 62% with MID HIGH violation (51-75% broke the rule) and 70% with HIGH violation (where 76-100% broke the rule). This was reflective of an increased preference to violate a rule, with increasing rate of others' rule-violation. This increased preference to violate is displayed in Figure 3.5. As we expected to find differences in the elicited proportion of violation preferences across the four brackets of empirical information, we ran a Friedman test, the non-parametric alternative to analysis of variance. Under this test, the null hypothesis states that there is no difference in elicited responses (of violation) across the four brackets while the alternative states that there would be differences. The results supported the alternative hypothesis: there were significant differences in the elicited distributions, χ (3) = 16.032; p = 0.0011. We also tested for bilateral comparisons between the distributions of the responses and found significant differences as reported in Table 3.2 below.

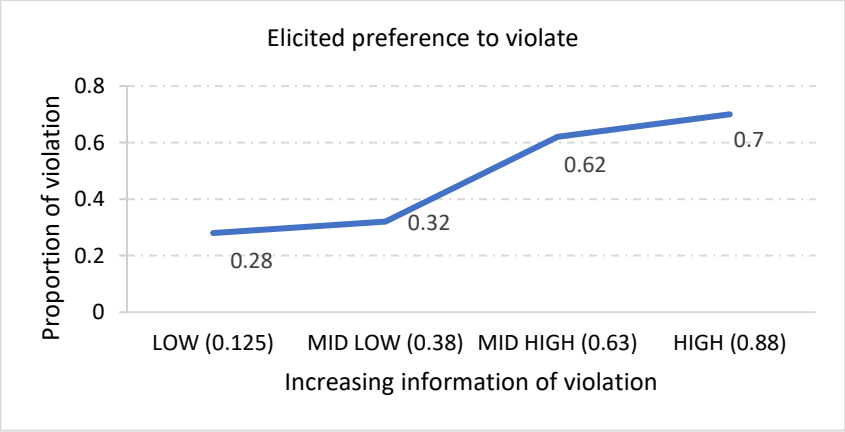


Figure 3.5: Elicited preference from S-method: increasing preference to violate with increasing empirical information of violation denoted by the blue line. The values in the parentheses denote the mid-values of the information brackets.

	Signed Rank Z	*p*	Adjusted *p*
LOW vs MID LOW	-0.577	0.774	0.774
LOW vs MID HIGH	-3.272	0.001	0.003
LOW vs HIGH	-4.041	0.000	0.000
MID LOW vs MID HIGH	-3.128	0.002	0.004
MID LOW vs HIGH	-3.657	0.000	0.000
MID HIGH VS HIGH	-1.414	0.289	0.346

Table 3.2: Distributional differences in S-method. We report the results from signed rank test with adjusted *p* values for bilateral multiple comparisons using the False Discovery rate procedure (Benjamini & Hochberg, 1995). The groups are classified as follows: LOW: 0-25% broke the rule; MID LOW: 26-50% broke the rule; MID HIGH: 51-75% broke the rule; HIGH: 76-100% broke the rule. We found significant differences between all possible distributions except LOW and MID LOW and MID HIGH and HIGH.

We then investigated the causal effect of such information on subjects' preferences. We converted the four possible scenarios into a continuous factor to check the effect of an increasing rate of violation on subjects' violation preference. We regressed the preference to violate on the midpoint of each possible scenario. In other words, we regressed the binary choice of violation with a rule on an increasing rate of others violation based on the values of 12.5, 38, 63 and 88 (these are the mid-point values of the four scenarios). This was to observe if there exists a linear effect of others' violation on subject's own preferences.

We report the results in Table 3.3. In Col (1) and Col (2), we report the results from a Linear Probability Model (without and with controls) and find a positive effect of increasing violation rate on one's preference to violate a rule and this is significant at 1%. In other words, a 1% increase in others' violation, increases the probability of violation preference by 0.6%. In Col (3) we report the results from a Logit model, where the effect stays positive and significant at 1%. In Col (4), we report the results from the odds ratio: there is an increase in the odds of violating a rule by 3% with a percentage increase in other's violation, and this effect stays significant at 1%, after controlling for age, gender, education, income level, weekly MTurk earnings and political orientation. Thus, subjects have an increased probability to violate if they

observe that a higher fraction of other MTurkers violate a rule. This reflects how others' violation (empirical information) influences preference for conformity in a rule-following scenario.

VARIABLE: Preference to violate	**(1)** LPM	**(1')** LPM	**(2)** Logit	**(2')** Odds ratio
% of others violation	0.00620***	0.00620***	0.0308***	1.0312***
	(0.00126)	(0.00129)	(0.00697)	(0.00719)
Constant	0.168**	-0.465	-4.955***	0.0070***
	(0.0735)	(0.310)	(1.744)	(0.1228)
Other controls	No	Yes	Yes	Yes
Observations	200	200	200	200
R-squared	0.122	0.230		

Table 3.3: Linear Probability and Logit Regression, S method. Dependent variable is a preference to violate which is a binary outcome. Standard error in parentheses, adjusted for 50 clusters in subject. Significance levels: *** p<0.01, **p<0.05, *p<0.1. We control for age, income, weekly earnings from MTurk, political orientation, and ethnic background.

The increase in the elicited proportion of violation preferences across the four fractions of increasing violation rate may not reflect a straight linear relationship as shown in Figure 3.5. This may be indicative of the fact that some of these fractions were more influential in shifting subject's preferences from compliance to violation outcomes.

To check if this was the case, we again ran linear probability and logit regression models with dummy variables for the four fractions of increasing violation rate, reported in Table 3.4. LOW (0-25% of other MTurkers violate) served as the baseline group. In Col (1) we report the result from a LPM regression: there is an increase in the probability of a violation preference in MID LOW by 4% but this is not statistically significant. However, with the information of MID HIGH (51-75% of other MTurkers violate), there is an increase in the probability of switching from compliance to violation preference (from LOW) by 34% and this is significant at 1%. Likewise, there is an increase in the probability of eliciting a violation preference by 42% in

the HIGH scenario (when 76-100% of other MTurkers violate), and this is significant at 1%[40]. Col (2) and Col (2') report the results from Logit regression and Odd ratios.

VARIABLE: Preference to violate	(1) LPM	(2) Logit	(3) Odds ratio
MID LOW	0.0400	0.215	1.239
	(0.0716)	(0.376)	(0.4657)
MID HIGH	0.340***	1.636***	5.134***
	(0.0955)	(0.486)	(2.497)
HIGH	0.420***	2.041***	7.6998***
	(0.0884)	(0.478)	(3.6841)
Constant	-0.493*	-5.075***	0.0062***
	(0.284)	(1.654)	(1.0133)
Other controls	Yes	Yes	Yes
Observations	200	200	200
R-squared	0.231		

Table 3. 4: Linear Probability and Logit Regression 2. Dependent variable is a preference to violate which is a binary outcome. Robust standard error clustered on subjects in parentheses. Significance levels: *** p<0.01, **p<0.05, *p<0.1. We control for age, income, weekly earnings from MTurk, political orientation, and ethnic background.

To summarize, these results show i) evidence of preference conditionality based on others' violation and that ii) a majority violation rate (more than 50% others violate) influenced subjects to shift from compliance to violation preferences.

Additionally, we also investigated the type of preferences by looking into the responses within the subjects across the four scenarios. We define unconditional rule violators (UV) as subjects who always violated the rule, irrespective of the empirical information given to them; unconditional rule followers (UF) as the subjects who always chose to follow the rule in the four scenarios of increasing violation; conditional rule followers (CF) as the subjects who switched from following the rule (costly action) to violating the rule (selfish action) with

[40] We report the pairwise comparisons of the marginal linear predictions in Table A3.1 in Appendix.

increasing information of violation from previous MTurkers. We classified the rest of the elicited preferences (that didn't fall under the previous three categories) as unclassified since these were erratic responses that switched between compliance and violation intermittently or when preferences changed from violation to compliance with increasing violation. Subjects' preference schedules are shown in Figure 3.6 below.

14% were UV, 18% were UF and 22% were unclassified. 46% of the subjects (or 23 out of 50) were CF, with different degree of information sensitivity: while a few made the shift with MID LOW violation rate (3 out of 23 or 13%), the rest shifted their preference with MID HIGH (18 out of 23 or 78%), and HIGH (2 out 23, or 9%). This confirms the above result that most CF require at least 50% or more incidence of violation around them to shift from the socially desirable to the selfish action. Additionally, this also shows that was a heterogeneity within the conditional rule-followers in their norm-sensitivity.

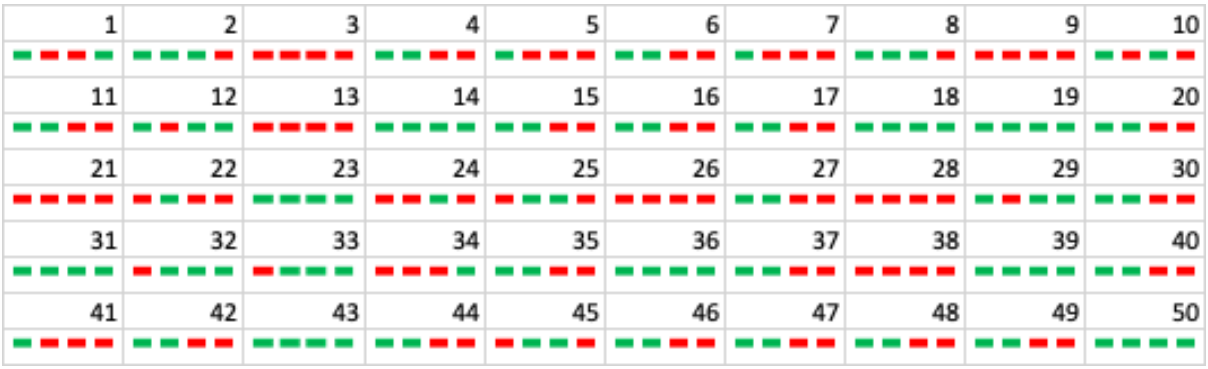


Figure 3.6: Elicited responses of the 50 subjects in the S-method. Green signifies compliance and red signifies violation preference. The four bars represent the four fractions of increasing violation information shown to the subjects (LOW, MID LOW, MID HIGH and HIGH). Unconditional Rule followers (9): subjects 14, 18, 19, 23, 31, 36, 39, 43, 50 ;Unconditional Rule violators (7): subjects 3, 9, 13, 21, 26, 28, 38; Conditional rule followers (3) (shift in MID LOW): subjects 5, 7, 41; Conditional rule followers (18) (shift in MID HIGH): subjects 4, 6, 11, 15, 16, 17, 20, 27, 30, 35, 37, 40, 42, 44, 46, 47, 48, 49; Conditional rule followers (2) (shift in HIGH): subjects 2, 8; Unclassified (11): subjects 1, 10, 12, 22, 24, 25, 29, 32, 33, 34, 45.

3.5.2 D-Method

Next, we present the results from the D-method. Firstly, to check if the anchoring manipulation worked, we elicited subjects' empirical beliefs on rule violation. We report the average elicited beliefs and the standard deviation across the treatments as follows: LOW = 30.11 (SD= 26.65); MID LOW: 42.16 (SD= 18.9); MID HIGH = 48.6 (SD= 17.29); HIGH = 63.52 (SD= 28.05). This shows that there was a shift (increase) in the elicited beliefs across the treatment groups of increasing violation information.

We also checked for a causal effect of anchoring on the elicited beliefs. To convert the four treatment groups into a continuous factor, we took the mid-point value of the two scenarios of violation illustrated to them. For instance, for LOW, when 8 and 17 (out of 100) violated the rule, the average violation rate is 8+17= 25/2 = 12.5. Likewise for MID LOW, MID HIGH and HIGH, the values are 37.5, 62.5 and 87.5. This was comparable to the mid-point values from the S-method.

We regressed the elicited beliefs as a function of the % of violation rate that was "anchored" at the beginning of the study. This served as a manipulation check to investigate if the graphical representations shifted the empirical belief of violation. We report the results from an OLS estimation in Table 3.5: In Col (1), we find that with a 1% increase in others violation rate, there is an increased empirical belief of violation by 0.427% and this is significant at 1%. This effect stays significant after controlling for other socio-demographic variables, as is illustrated in Col (2). Thus, the anchoring procedure was successful in shifting subjects' beliefs of violation.

VARIABLE: Elicited belief	(1) OLS	(2) OLS
% of others' violation	0.427***	0.446***
	(0.0660)	(0.0681)
Constant	24.75***	36.83***
	(3.630)	(14.00)
Other controls	No	Yes
Observations	202	202
R-squared	0.212	0.234

Table 3.5: OLS Regression. Dependent variable is elicited empirical beliefs. Standard error in parentheses. Significance levels: *** p<0.01, **p<0.05, *p<0.1. We control for age, income, weekly earnings from MTurk, political orientation, and ethnic background. None of the control variables are significant.

We then investigated the effect of this belief shift on subjects' behavioural decisions in the rule-following task. Interesting, there was a decline in the violation outcome in the actual task, even with a shift in beliefs: the violation rates were: 51% in LOW; 58% in MID LOW; 38% in MID HIGH and 34% in HIGH. We graphically show the violation rates across the four treatment groups in Figure 3.7 below. We expected to find an increased trend in violation across the 4 treatment groups of (increasing) empirical information of violation, based on the results from Gächter et al. (2021) and the S-method above. Since the D-method generates independent observations, to investigate the differences in the proportion of violation outcomes across the groups, we ran a Jonckheere- Terpstra test, which is used to test for an ordered difference in the medians of the groups. Here, the null hypothesis states no difference in the violation outcomes across the four treatment conditions. On the other hand, the priori alternative hypothesis states that the violation outcomes increase as we move across the brackets of increasing violation rate (based on previous evidence). We find the following result: Jonckheere- Terpstra statistic test Z= -2.254; p = 0.0222. Our priori alternative expects violation to increase, but we find the opposite trend: there is a decrease in the proportion of violation with (increasing) empirical information of others violation. Thus, the alternative hypothesis is rejected in favour of the opposite.

Thus, the trend displayed in Figure 3.7 and the result from the statistical test suggest that there was a lowered preference to violate with increasing information of violation, which was unexpected and contrary to the results from Gächter et al (2021) and S-method.

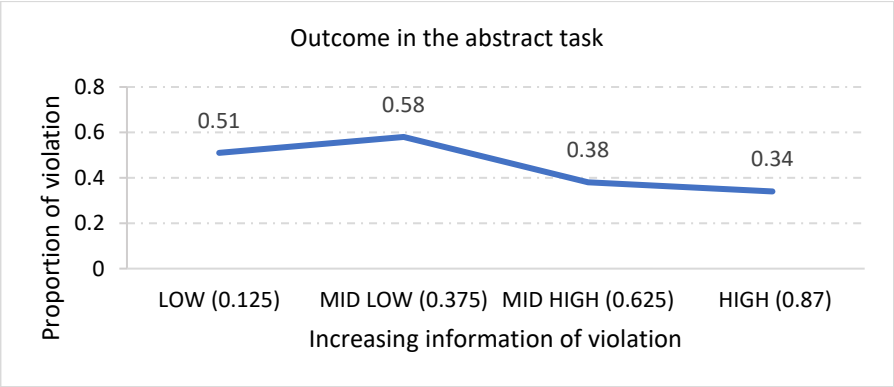


Figure 3.7: Elicited responses from D-method. **The blue line shows the reported proportion of violation outcomes across the four treatment groups.**

We also ran a regression analysis to check for causality. This is reported in table 3.6 below: In Col (1), we report the effect of other's violation rate on subjects' own violation in the rule-following task using a linear probability model. We find that with a 1% increase in other's violation, there is a decrease in the probability of violation preference by 0.283% and this is significant at 5%. The effect becomes significant (at 1%) after controlling for other socio-demographic characteristics, as is depicted in Col (1'). Col (2) shows a similar negative effect at 1% from a Logit model, after controlling for demographics. In Col (2'), we report the results from the odds ratio: there is a decrease in the odds of violating a rule by 1.5% with a percentage increase in other's violation.

VARIABLE: Preference to violate (D)	(1) LPM	(1') LPM'	(2) Logit	(2') Odds Ratio
% of others' violation	-0.00283**	-0.00344***	-0.0155***	0.985***
	(0.00124)	(0.00127)	(0.00587)	(0.00578)
Constant	0.599***	0.619**	0.536	1.709
	(0.0717)	(0.258)	(1.134)	(1.939)
Other controls	No	Yes	Yes	Yes
Observations	202	202	202	202
R-squared	0.025	0.109		

Table 3.6: Linear Probability and Logit Regression: D method. Dependent variable is a preference to violate which is a binary outcome. Standard error in parentheses. Significance levels: *** p<0.01, **p<0.05, *p<0.1. We control for age, income, weekly earnings from MTurk, political orientation, and ethnic background.

To check if some category of % others' violation influenced the decline in violation preference more than others, we ran linear probability and logit regression models with dummy variables for the four fractions of increasing violation rate, reported in Table 3.7. LOW served as the baseline condition. In Col (1) we report the result from a LPM regression: there is an increased probability of a violation preference in MID LOW by 3.36% but this is not statistically significant. However, with MID HIGH, there is a decrease in the probability to violate by 17.3% and this is statistically significant at 10%. With HIGH, there is a decrease in the probability to violate by 22% and this is statistically significant at 5%[41]. Col (2) shows a similar negative effect at 10% (for MID HIGH) and at 5% (for HIGH) from a Logit model, after controlling for demographics. Col (2'), reports the odds ratio: there is a decrease in the odds of violating a rule by 54.6% in MID HIGH (significant at 10%) and by 63.2% in HIGH (significant at 5%).

To summarise: Although the anchoring procedure was successful in shifting subjects' beliefs, there was a decline rather than an increase in violation outcomes across the treatment groups.

[41] We report the pairwise comparisons of the marginal linear predictions in Table A3.2 in Appendix

Moreover, a majority of others' violation (more than 50%) influenced subjects' violation decisions. This is the opposite of what we expected and observed in previous research (Gächter et al. 2021) and in the S-method (see Figure. 3.5).

VARIABLE: Preference to violate (D)	(1) LPM	(2) Logit	(2') Odds Ratio
MID LOW	0.0336	0.143	1.154
	(0.0981)	(0.427)	(0.493)
MID HIGH	-0.173*	-0.791*	0.454*
	(0.0985)	(0.438)	(0.199)
HIGH	-0.220**	-1.000**	0.368**
	(0.101)	(0.461)	(0.170)
Constant	0.537**	0.153	1.166
	(0.256)	(1.133)	(1.321)
Controls	Yes	Yes	Yes
Observations	202	202	202
R-squared	0.118		

Table 3. 7: Linear and Logit Regression 2. Dependent variable is a preference to violate which is a binary outcome. Standard error in parentheses. Significance levels: *** p<0.01, **p<0.05, *p<0.1. We control for age, income, weekly earnings from MTurk, political orientation, and ethnic background.

We also compared the proportion of violation reported from both methods. A summary of the proportion of violation preferences from the methods is given in Table 3.8. While in the S-method, preference to violate increased as a function of increasing violation information, in the D-method, the violation preference actually decreased from LOW to HIGH. We ran a test of difference of proportion for each of the four scenarios between the methods and found that there were significant differences between elicited proportion of violation preference from the two elicitation methods, and this was significant at 5% (LOW and MID HIGH) or 1% (MID LOW and HIGH). The null hypotheses of no difference between the elicited responses between the two methods is rejected for all four cases.

% Of Violation	S-method	D-method	Z value	Adjusted *p* value
LOW	28% (14/50)	52% (27/52)	-2.4635	0.0138
MID LOW	32% (16/50)	58% (29/50)	-2.6131	0.0090
MID HIGH	62% (31/50)	38% (20/52)	2.3768	0.0175
HIGH	70% (35/50)	34% (17/49)	3.5172	0.0004

Table 3.8: Proportion test to compare preferences in S and D methods. We do find significant differences at 5% in the elicited preferences (of violation) across the four scenarios. We also report the percentage of elicited preference from S and D-method. Note that LOW represents 0-25% MTurkers violate, MID LOW is 26-50%, MID HIGH is 51-75% and HIGH is 76-100%. While the proportion of violations increases in the S-method, it decreases in the D-method, as we move from LOW to HIGH.

3.6 Discussion and Conclusion

Although Gächter et al. (2021) find evidence of preference conditionality with peer effects, their elicitation procedure comprises of multiple treatment groups and can be cumbersome from the viewpoint of collecting experimental data. This study aimed to investigate the effectiveness of two preference elicitation methods: the S and the D-method, to test such preference to follow rules conditional on an increasing rate of others' violation outcomes. We used the rule-following task as the decision environment, adapted from Gächter et al. (2021). The S-method involved a within-subject analysis, where subjects elicited four responses based on an increasing fraction of violation information presented to them. On the other hand, the D-method incorporated a between-subject component. There were four treatment groups of different fractions of empirical information (proportion of violation) presented to subjects, after which subjects participated in the actual rule-following task.

In the S-method, we found evidence of such preference conditionality as we move across the brackets of increasing violation rate. This is consistent with the findings from Gächter et al. (2021). This design is also comparable to the real-time response of peer effects from their study as empirical information and preference elicitation were presented to the subjects at the same

time. An increasing incidence of the behavioural outcome of rule violation may have lowered subjects' normative considerations of the observed behaviour (Bicchieri et al., 2020). In other words, by observing a higher incidence of rule violation, subjects may have inferred that high(er) rule violation of others is indicative of the idea that rule-violation may be socially acceptable behaviour. As a result, they updated and lowered their normative standard towards rule-compliance. Thus, the high prevalence of violation and its subsequent influence on the underlying normative consideration could have causally affected subjects' preference to violate a rule. Additionally, we found a bigger share of conditional rule-violators, i.e., subjects who shift from a socially appropriate to a selfish preference as they witness more violations around them. This also indicates that rule compliance may not be internalised as morally correct behaviour and is rather based on conformity with others' actions.

Additionally, the S-method was helpful to observe preference heterogeneity with regards to rule compliance and elicit subject "types". For instance, if some subjects always chose to follow the rule, irrespective of the increasing empirical violation information, they could be classified as unconditional rule-followers because of an internalised logic of rule-following due to high moral standards. On the other hand, if some subjects displayed a preference always to violate a rule, irrespective of the empirical information given to them, they were considered as unconditional rule-violators, as their selfish self-interest drove their preference, and they didn't care about what others around them did. Both these types of subjects can be considered as trend-setters, as they didn't get affected by social influence. However, if subjects elicited a preference to follow a rule, but this preference shifted from following a rule (costly action) to breaking a rule (selfish action) by observing high(er) rates of rule-violation around them, then they were classified as conditional rule-followers (or violators) as their preference to act a certain way was influenced by the social information of violation around them. Additionally,

various studies have suggested that norm-followers exhibit different degrees of norm-sensitivity (Bicchieri, 2016; Fehr & Schurtenberger, 2018; Kimbrough & Vostroknutov, 2016) and we do find differences within these conditional followers as to how much violation they need to witness around them to shift from the socially appropriate to the selfish action. Therefore, the S-method is useful to observe the heterogeneity in subjects with regards to their norm-sensitivity.

In the D-method, we provided subjects with graphical representations involving certain fractions of rule-violation to influence beliefs and subsequently subject's behavioural outcomes in the rule-following task. Although these informational anchors were successful in shifting subjects' beliefs of rule violation, they did not induce positive preference conditionality. Subjects' empirical beliefs of rule violation around them was not long-lasting enough to causally influence the behavioural outcomes in the task. Here, the graphical representations were followed by empirical belief elicitation (which served as a manipulation check), which was then followed by the actual task in real-time. [42]

Between the two methods, we suggest the S-method for eliciting conditional preferences for a couple of reasons. Firstly, it is a more economical method of preference elicitation, as we can

[42]These illustrations with the green and red coloured circles could have activated the normative thoughts regarding the social appropriateness of following a rule and social inappropriateness of violating a rule, which may have had some influence effect on subjects' decisions in the rule-following task. We had asked subjects the possible reasons for why subjects from the hypothetical study may have followed the rule and violated the rule just after showing them the illustration. Subject responses are given in the Appendix. For rule-followers, some subjects stated the following reasons as possible motivations for following a rule: a clear understanding of the task, being fair-minded, honest, patient, and having a moral inclination to do the right thing. Some also suggested that people may have followed the rule due to a fear of possibly being punished. On the other hand, for the rule-violators, a few subjects suggested the following possible motivations: being unclear about the task; being selfish, impatient, or anxious. Some even suggested that people violated the rule, as there was no punishment. These subjects thus might have had a preference to dissociate themselves from the rule-violators and comply with the rule, out of self-image or social image concerns, which have been shown to affect behavioural outcomes (Akerlof & Kranton, 2000).

collect multiple responses per subject. Secondly, it is helpful to observe the proportion of different types of preferences, i.e., conditional violators, as well as unconditional rule followers and rule-violators and also norm-sensitivity within subjects. Thirdly, in this case, the shift in empirical beliefs and its effect on the normative beliefs of appropriate (or inappropriate) behaviour happens at the same time as the actual decision. As a result, the beliefs have an immediate role in subjects' elicited preferences. On the other hand, the D-method has a more complex effect on preference elicitation, where the graphical representations may have influenced both empirical and normative thoughts, resulting in lowered violation outcomes with an increasing violation, possibly out of self-image concerns. However, this is merely speculative at this stage, and further testing is required to make more meaningful conclusions. For instance, eliciting direct responses in multiple rounds would help establish the stability of this result. Additionally, it would also be helpful in illustrating the types of preferences in this setup. Moreover, it would be beneficial to replicate the D-method, using illustrative cues in a more abstract setup with no colour-coded display of rule-violators and followers.

3.7 Appendix

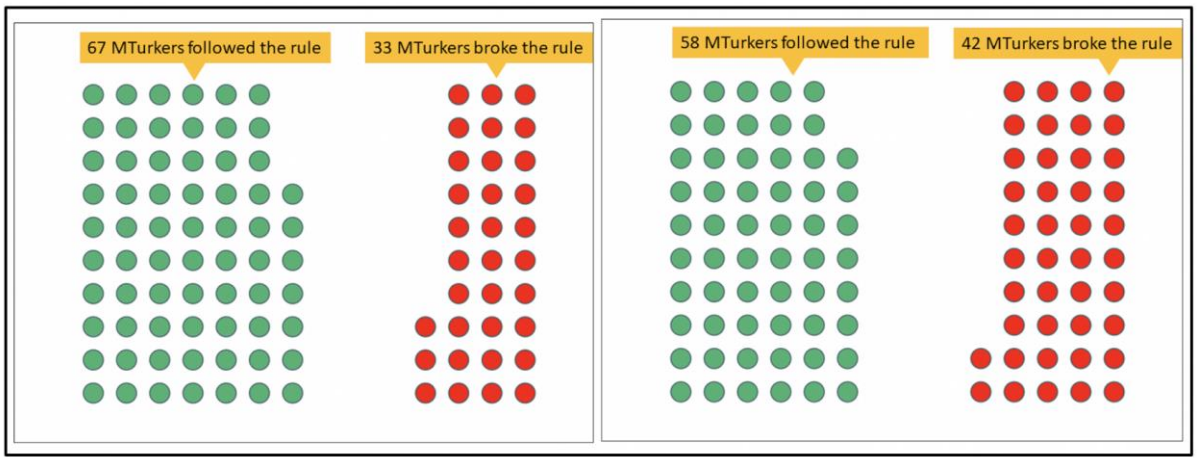
3.7.1 Supplementary Analysis and Figures

Scenario	Contrast	Standard Error	T statistic	Adjusted *p* value
MID LOW vs LOW	0.04	0.0715	0.56	0.579
MID HIGH vs LOW	0.34	0.0954	3.56	0.001
HIGH vs LOW	0.42	0.0883	4.75	0.000
MID HIGH vs MID LOW	0.3	0.0891	3.36	0.001
HIGH vs MID LOW	0.38	0.0921	4.12	0.000
HIGH vs MID HIGH	0.08	0.0574	1.39	0.170

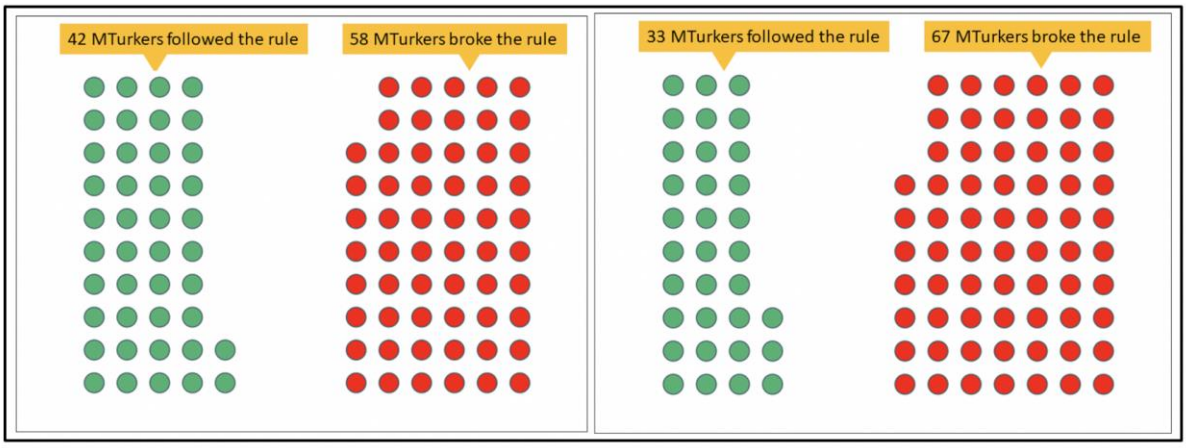
Table A3.1: Pairwise comparisons of marginal linear predictions from S method. The *p* values are adjusted for multiple comparison using the FDR correction.

Scenario	Contrast	Standard Error	T statistic	Adjusted *p* value
MID LOW vs LOW	0.03	0.0981	0.34	0.733
MID HIGH vs LOW	-0.17	0.0984	-1.76	0.080
HIGH vs LOW	-0.22	0.1011	-2.18	0.031
MID HIGH vs MID LOW	-0.20	0.0927	-2.23	0.027
HIGH vs MID LOW	-0.25	0.0955	-2.66	0.009
HIGH vs MID HIGH	-0.04	0.0946	-0.50	0.621

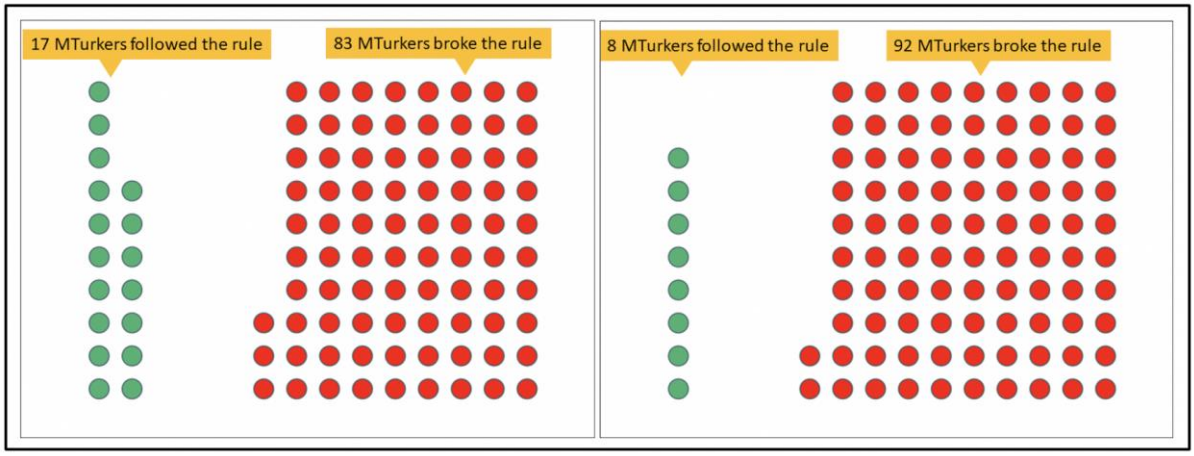
Table A3.2: Pairwise comparisons of marginal linear predictions from D method. The *p* values are adjusted for multiple comparison using the FDR correction.



A3.1: Graphical representations of MIDLOW.



A3.2: Graphical representations of MIDHIGH.



A3.3: Graphical representations of HIGH.

3.7.2 Subject Responses

Treatment	Subject	"Why do you think they broke the rule?"	"Why do you think they followed the rule?"
LOW	1	They did not read the rules.	They had read the instructions.
LOW	2	Maybe they want to finish the study quickly	Maybe they follow the rule clearly for getting bonus
LOW	3	To try to maximize their earnings to complete the task quicker	So, they are not disqualified from the study
LOW	4	Because they want to reach black destination early.	They followed the rule of when the cross disappears then only you have to start.
LOW	5	They either broke the rules at the potential of making more money, or they simply did not understand the rules.	Most people tend to follow rules that are given without an attempt at cheating for more.
LOW	6	There was no stated punishment for moving before the X disappeared. Or they didn't fully understand the rules. Or they did it by accident.	They did that because they felt that was the best way to get the best score.
LOW	7	It appears that they get paid even if they break the rule, so they break the rule to get paid more. It's also possible they didn't fully understand the instructions, or they have limited motor skills. Or they just don't care.	
LOW	8	I would guess they did that because they do not fully understand the rules.	I feel they understand the rules and are paying attention to what is going on.
LOW	9	They likely were so tense about clicking the move button as soon as the cross disappeared that they accidentally clicked pre-emptively. It's also very possible that they thought breaking the rules wouldn't matter and they could earn more money by clicking move immediately.	Most likely, they were worried about losing the bonus money, so they were very careful to follow the rules to avoid being disqualified.
LOW	10	Yes, to receive more points per the question/answer given.	Yes, they followed the rules.
LOW	11	I believe the 8 MTurkers broke the rule because they wanted to earn more points. If they got to the black bar earlier, they'd have more points banked and would get a better outcome. That's as opposed to following the rule but getting less points. It really comes down to whether more points or following the rules is more important to the player. The total points were more important to the 8 MTurkers who broke the rule.	I think that to them, it was more important to play fair and to follow the rules of the game as opposed to just going for personal gain. They placed a greater value on being a good player rather than just being greedy or just out for themselves. They wanted to be able to feel good about themselves and how they conducted themselves over the course of the game. That was a bigger reward for them as opposed to getting a higher point total.
LOW	12	because they wanted to earn the extra points or attempt to	To stay safe and get some points.
LOW	13	Because they are rewarded rather than punished for breaking the rule. The ones who broke the rule earned the most points.	Because they are used to following the rules, and possibly assume there is some sort of punishment or disincentive for disobeying, even if they are not told what it is.
LOW	14	they didn't read the instructions, or they just got impatient, or they didn't really care about the money and just wanted to get finished with the task	they don't mind following rules and they like maximizing their gain
LOW	15	Either they were in a hurry, or they didn't understand the rule.	They were paying attention.
LOW	16	this a correct way to clear the circle	this is a correct way
LOW	17	To increase their earnings output	They were told not to move
LOW	18	They did not understand the rules or did not understand how to stop and move the ball. They could have simply made a mistake.	They read the rules correctly and understood how the game was played
LOW	19	Maybe they thought they would get a greater number of points, and therefore money	They were attentive to the rules and didn't want to break them.
LOW	20	They are probably. Thought that it would take less time and leave them more points.	They wanted to follow the rules and win the game.
LOW	21	They were impatient and clicked to move, due to either misunderstanding the cross rule, or just wanting to maximize profits. Another reason is a bot advanced enough to start the game, but not understand the rules at all.	They read the instructions and followed them properly.
LOW	22	They are anxious and not concentrating	They read the instructions very well and know what they are doing.
LOW	23	They could have accidentally moved, or not read the instructions correctly	Because they read the instructions and it seemed like the safe bet
LOW	24	They thought they would get more points.	They wanted to follow the instructions.

LOW	25	yes, point in the part of stay	the point in low point
LOW	26	To see if they could advance without penalty	To follow the rules of the game and receive a bonus
LOW	27	They did not follow the instructions, or they wanted to break the rules therefore earning more points by completing the task in less time.	They followed the rules and by doing so they did not want to be disqualified. They believe in playing fair.
LOW	28	I would guess that they either did not read the instructions completely, or they read them quickly and rushed through the HIT in order to get paid. It's possible that they had trouble comprehending the rules and that it was an innocent mistake. Or maybe the cross didn't appear in their study due to an error of some kind and after sitting long enough to almost lose all of their money, they moved anyway. The final reason I can think of is because perhaps those 8 were given different instructions.	They read the instructions closely and wanted to maximize their profits.
LOW	29	To try and get to the black area in the quickest time.	They did not want to jeopardize the amount they would get paid.
LOW	30	I Imagine they broke the rule because of some disturbance that happened in that time.	They followed the rule because of they can understand every instruction carefully and obey the rules.
LOW	31	They were quick to assume they could go and therefore get a faster end time with more money.	They wanted to be patient and follow the rules, making whatever they could for a time they spent.
LOW	32	After reading all the directions I'm assuming 8 broke the rules trying to get to the blackline faster. Since I haven't played the game, yet I understand the subtleties by the descriptions and picture guides but not actually participating yet has to make me think that though the game is simple there might be more to the game than initially described. Perhaps the controls are as described but maybe there is a timer delay. But to me the answer is to make more money	The 92 workers moved after the cross left to follow the rules and to get paid something
LOW	33	They wanted the highest possible number of points	Maybe they were concerned that there would be a penalty for breaking the rules; therefore, they would lose points.
LOW	34	I think they would move before the cross had gone away because they wanted to try to finish the task quicker, which would mean a higher bonus.	The 92 probably waited because they wanted to follow directions and be sure they got paid for the task.
LOW	35	They want more point, so they quickly move the circle	They follow the instruction correctly.
LOW	36	I imagine that they were broke the rule because of some kind of disturbance happen in the certain time period.	They follow the rule because they understand the instructions very quick and better way.
LOW	37	Because they didn't read the directions closely enough.	They were attentive and followed the directions.
LOW	38	They wanted to maximize their potential earnings.	They wanted to comply with the rules or felt it was risky to not comply.
LOW	39	for money	they are afraid
LOW	40	I suspect they are just feeling anxious to move their circle	Because that's what the instructions asked us to do
LOW	41	The wanted to get the max amount of money possible. I believe that they thought it was better to break the rules, and earn more money, than to watch their money decrease.	They were following the rules laid out before them. The wanted to complete the task successfully and were not as concerned about the decrease in points.
LOW	42	They were anxious	They weren't anxious
LOW	43	They earn more points if they move before the cross disappears.	Because they tend to follow the rules.
LOW	44	they might have	not sure
LOW	45	Some people have a hard time following direction.	Most people like to follow the rules.
LOW	46	They didn't read or understand the instructions.	They read and understood the instructions. And they wanted to complete the study.
LOW	47	This is done because there doesn't seem to be a penalty for moving before the cross disappears.	Because they aren't thinking outside of the box and are trying to be rule followers.
LOW	48	They didn't read the directions properly?	They read the instructions.
LOW	49	They probably didn't read the instructions.	They read the instructions
MID LOW	50	Yes, they thought that by breaking the rules, they would be able to maintain as many points as they possibly could. They also thought that by maintaining as many points as they possibly could, they would be able to make as much money for the HIT as possible. According to the rules, it says to	They were simply following the rules of the game.

		wait in the grey area until the cross disappears, and you will be paid according to time. It does not say anything about being disqualified if you fail to stay in the grey area until the cross disappears.	
MID LOW	51	They wanted more money and weren't being judged on following the rules but by feeling faster.	They wanted to follow the rules given by the task.
MID LOW	52	They did not read the instructions, or it was an accident	They read the rules and care about getting paid.
MID LOW	53	They broke the rule to have more points and earn a bigger bonus	They did this to follow the study rules rather than worry about the bonus
MID LOW	54	they wanted to get through the study as fast as possible or did not read directions	because they read the directions and knew what needed to be done/when
MID LOW	55	They did not fully wait until the cross hit the grey area because they wanted to make more bonus gains.	They did not risk waiting a bit more, even if it means less bonus.
MID LOW	56	To get a higher bonus payment	To get a higher bonus payment
MID LOW	57	don't like rules	Like rules
MID LOW	58	It pays better to ignore the instructions and move before the cross disappears	They followed the instructions or thought there would be repercussions for moving before it disappeared.
MID LOW	59	Impatient or did not understand the instructions. May have been wanting to go faster to get the reward.	Trying to follow the rules as to not be disqualified from the study.
MID LOW	60	to use less time and hence lose less points out of the 20	Because they are following the instructions you gave to wait for the cross to disappear.
MID LOW	61	Probably tried to anticipate when the cross may disappear. Kind of like a sprinter who starts too early...he/she is trying to get that tiny advantage by jumping ahead a slight bit. A few probably just did it by accident	They're probably calmer, level-headed, or patient. You're going to get something no matter what, so they're probably not too concerned with squeezing out that extra penny or two.
MID LOW	62	To earn more on the bonus	They might feel they may be rewarded for following the rules.
MID LOW	63	Because they want to move as soon as possible. All thirty-three workers didn't have patience for the disappearance of the cross.	Because they waited and succeeded.
MID LOW	64	They didn't understand the instruction	They understood the instructions
MID LOW	65	They did not read the directions. They did not care about the directions. They wanted to see what would happen.	They are good at following directions.
MID LOW	66	to make it to the black area more quickly	because they chose to play by the rules set before them
MID LOW	67	I think they not like the rules, so they broke the rules.	They will clear the rules and follow the instructions, so they followed the rules.
MID LOW	68	They wanted to do it quickly	They were very careful.
MID LOW	69	They decided that even though there was a rule that prohibited this, there was no punishment for breaking the rule. It was stated that you would be paid for how long it took, and nothing was said about not getting paid if you did not follow the rules. Those who caught this decided that it was worth the extra money to break the rules.	They either do not like to break rules in fear of punishment or did not catch that you could indeed break the rules in this exercise.
MID LOW	70	Accidental clicking, impatience, distraction	To comply with the survey requirements, to succeed, to continue the game
MID LOW	71	They accidentally started too soon	They had very accurate timing.
MID LOW	72	They attempted to earn more points	The read the instructions properly and understood the rule.
MID LOW	73	They were too eager	They waited calmly
MID LOW	74	I feel that the 33 MTurkers that broke the rule may have broken it because they did not fully read or understand the instructions.	I feel that the 67 MTurkers that followed the rule and moved their circle after the cross had disappeared did this because they fully read and understood the instructions.
MID LOW	75	They didn't understand the rules	They read and understood the instructions
MID LOW	76	because they didn't follow the rules	because they didn't follow the rules
MID LOW	77	Because they did not see a consequence of breaking the rule and performing faster would yield more credits.	They probably believed that there was some unknown consequence for breaking the rule and did not want to incur any time of penalty.
MID LOW	78	Because it was not stated that they would not be paid if they broke the rule	They followed the rule since they feared that they would not be paid if they didn't.
MID LOW	79	They didn't read the instructions carefully.	They did read and understand the instructions and wanted to do well on the task.
MID LOW	80	Because the 67 people followed instructions & the 3 people did not.	Because they followed instructions

MID LOW	81	To earn more points, although it is a rule, moving quicker still gives you more points.	Because they believe that following the rules is more important than getting more points.
MID LOW	82	To maximize their points.	They want to follow the rules.
MID LOW	83	They want the higher reward. $.50 is not a lot and many people do this because they need money.	Some people might have thought that if they follow the rules, there was another reward.
MID LOW	84	they are too eager to minimize the time they spend	they want to stick to the rule
MID LOW	85	They wanted to maximize their earnings and didn't care about rules on a frivolous survey.	They were just doing as they were told and weren't concerned about losing a small amount of loose change.
MID LOW	86	They weren't following the directions.	Because they were able to follow directions
MID LOW	87	They were in a hurry and being careless	They were ethical
MID LOW	88	very interesting they average answer	good interest
MID LOW	89	I think the people who broke the rule might have determined to get as much bonus as possible, since there was no apparent punishment to breaking the rule. It is also possible that, as the seconds ticked by, some people wondered if the actual test involved the X not disappearing at all but was rather measuring when people would stop waiting. I would wonder that myself and wouldn't want to miss out on a bonus but would also want to be honest.	I think most of those that followed the rule determined to follow it no matter what, because they take pride in their work and want to do a good job on the task. I also think it's possible that a group of people would have determined at what point they stop waiting for the X to disappear, and just go get a bonus to at least pay for their time.
MID LOW	90	It was not explained whether there would be a consequence for breaking the rule. They were trying to get the task done as fast as possible to get the highest reward.	They were told that there was a rule, and they probably assumed that there would be a consequence for breaking the rule.
MID LOW	91	Some were guessing that there were no repercussions (most of the 33) and would make more. Some accidentally went forward.	They followed the directions
MID LOW	92	They didn't read the instructions, or they weren't being patient.	They read the instructions and wanted more points.
MID LOW	93	Biggest reason is that they did not fully read the rules. MTurkers are often looking to complete HITs quickly. Not paying attention to the rules means they do not complete the task correctly. They are more interested in maximizing their potential bonuses.	They followed the directions and understood that they can only collect their bonus if the task is done correctly.
MID LOW	94	Because the instructions were unclear, and they were impatient	Because they didn't comprehend the instructions
MID LOW	95	Those MTurkers disregarded the rule in order to maximize their points. By ignoring the stop sign, they were able to gain more points and thus a larger bonus.	They were honest and followed the rules of the study. They knew they would get a bonus and played honestly so they wouldn't jeopardize that bonus.
MID HIGH	96	58 probably broke the rule because they want the highest earnings. However, the highest earnings are against the rules since you're supposed to move after the cross disappears.	They simply followed the rules.
MID HIGH	97	They would break the rule because they did not read it or understand it	They followed the rule because they read it and understood it
MID HIGH	98	They didn’t fully understand	They understood the rules
MID HIGH	99	To make more money	To ensure they don't get disqualified from the study
MID HIGH	100	They probably didn't carefully read the instructions, since there was a lot written on the first page. Alternatively, perhaps they were very concerned with the bonus timer ticking down that they decided to break the rule in hopes that they would receive a larger bonus.	Because they follow the rules and wanted to maximize the possibility of getting their bonus. There were a lot of rules, and it might have been easy to get lost, which is why only 42% of participants were able to follow them, but these people are likely the ones who are most concerned with getting the bonus and are therefore paying the most attention.
MID HIGH	101	They just wanted to end the survey as fast as possible.	These ones like to follow instructions.
MID HIGH	102	They probably did not read the instructions completely.	They did read the instructions completely.
MID HIGH	103	They wanted to get a head start on the others. Which they are not supposed to do that.	They read the rules and wanted to play by them
MID HIGH	104	Because they were anxious	because they were intentional about the instructions
MID HIGH	105	They get impatient or anticipate when the cross will disappear and move too early.	They are following instructions and want the full points
MID HIGH	106	I imagine they did so to increase their reward.	I guess they wanted to follow the rules or were worried what might happen if they broke the rule.
MID HIGH	107	Maybe they failed to read the instructions and did not know they needed to wait until the x disappeared	They read the instructions and wanted to win the game

MID HIGH	108	Because the rule requires waiting for the cross disappear, causing them to lose points for every second they wait for the cross. Players want to maximize their points, which explains why they wouldn't want to lose points by waiting for the cross to go away.	Because they might fear there is some sort of point penalty for breaking the rule, thus reducing the amount of money they earn.
MID HIGH	109	They didn't understand the instructions or were impatient	They understood the instructions and were patient.
MID HIGH	110	They didn't read the rules?	They read the rules
MID HIGH	111	Yes, they want to finish faster and earn more money	Yes, they want to follow the rules and earn more money and not be punished for starting earlier
MID HIGH	112	These MTurkers wanted to maximize their bonus earnings after realizing that they could break the waiting rule successfully.	The rule was established before the start of the game. These MTurkers did not want to break the rules as they may not have noticed that one could move before the cross disappears.
MID HIGH	113	They make more	They're too honest
MID HIGH	114	Yes, to earn more point and thus make more money.	They followed the rule which is to wait until the cross disappears.
MID HIGH	115	Probably to get the maximum amount of money they could. If you're paid out by the points you have at the end of the task, it would make more sense to do the task as fast as possible even if you broke the rules. However, I imagine if you break the rules, you get no pay out.	They wanted to follow the rules set by the requesters. It probably is also more incentivized to do that versus trying to break the rules.
MID HIGH	116	Perhaps, they took this as a game and played. People were in a rush to complete the game.	They were diligent in completing the survey. They stuck to the professional code for MTurkers.
MID HIGH	117	They wanted to maximize the amount of bonus they received.	They believe they should follow the rules of the game even if it is not to their advantage.
MID HIGH	118	They think there will be a secret benefit to moving faster.	They want to follow the rules.
MID HIGH	119	Because they do not read the instructions.	Because they are correctly following the instructions.
MID HIGH	120	to get more points	because they aren't rule breakers
MID HIGH	121	They didn't follow the directions or perhaps were stressed by the situation by jumping the gun.	They understood clearly what they purpose of the study was and accomplished it according to plan.
MID HIGH	122	Of the 100 MTurkers who play the task, 42 follow the rule and move after the cross disappears, while 58 break the rule and move before the cross disappears.	42 MTurkers followed the rule and moved their circle after the cross had disappeared. It is 42+58 total 100 MTurkers followed the rule moved.
MID HIGH	123	because they thought they might make more money or did not want to take the time to wait	because they thought they would make more money if they waited
MID HIGH	124	To increase the pay-out at the end of the task.	They felt following the rules as they perceived them would grant them the optimal pay-out.
MID HIGH	125	The wanted to minimize the time it took them in order to maximize their points.	They wanted to follow the rules and play an honest game.
MID HIGH	126	They were eager to earn points.	They were patient and calculated.
MID HIGH	127	they want the maximum amount of money	they probably worry that breaking the rules will have a negative impact on their payment
MID HIGH	128	They were impatient and wanted to do it fast	They like to follow rules
MID HIGH	129	They probably wanted to increase the amount of money they would get by decreasing the amount of time they would have to wait in the grey area.	They were following the rules and maybe they did not occur to them to break the rules in order to get more money.
MID HIGH	130	58 MTurkers broke the rule and moved their circle before the cross had disappeared	42 MTurkers followed the rule and moved their circle after the cross had disappeared
MID HIGH	131	I think they would do that because they didn't read the instructions or maybe they wanted to finish it really fast.	they probably did that to earn extra points and in turn more $
MID HIGH	132	To get the points	They wanted to follow the rules
MID HIGH	133	I imagine they did that because they did not fully understand the rules or did not have the patience to wait.	Because these were the rules of the study to earn a bonus
MID HIGH	134	They did that cause there were nervous about trying to impress themselves with this game because of the bonus involved	These people were paying attention and deserved to win and make some bonus money
MID HIGH	135	I imagine they would do that because they got impatient or didn't read the rules, therefore they did whatever they could do without thinking about it.	They followed the rules and didn’t let any kind of anxiety or impatience get to them to do anything of that sort.
MID HIGH	136	Didn't read carefully.	They read carefully
MID HIGH	137	They would make more money because they have more time	They are rule-followers and want to do things the right way.

MID HIGH	138	To maximize their points.	They wanted to follow the rules.
MID HIGH	139	To earn more points	To play a fair game
MID HIGH	140	They wanted to maximize their pay-out.	Because they are honest people.
MID HIGH	141	They did not see any cross so they had to imagine when it would disappear.	They calculated the amount of time it took to get to the cross and repeated that time crossing to the black line.
MID HIGH	142	they didn't follow the instructions as required	they must have read the instructions very carefully to notice they had to wait for the marker
MID HIGH	143	It's more difficult to cross	yes, i imagine is difficult
HIGH	144	To earn 15 points instead of 9 points	To get 9 points and follow the rule
HIGH	145	They thought they would receive more money. So, they disregarded the rules in the hopes that doing it faster would increase their bonus.	They followed the instructions because they thought they would be disqualified. They also would want to follow the rules because it is the right thing to do.
HIGH	146	They want to earn more points and earn more money.	They want to be cooperative at the game and avoid being disqualified at the end of the task.
HIGH	147	They do it because sometimes people just want to do it to violate the rules set.	These people just want to follow the rules set.
HIGH	148	The test is timed. They believe that it is better to get a better score than to follow all the rules. Whichever one results in the greatest number of points.	They believe that it is better to follow the rules than to get points. If you get disqualified your score will not count.
HIGH	149	They wanted to maximize their pay-out, so they moved the circle earlier to maximize their endowment and therefore their bonuses.	These participants likely want to follow the rules and provide the best data possible, or they are concerned about being rejected, so they wanted to follow the rules as much as possible to reduce the likelihood of the task being rejected.
HIGH	150	Maybe they thought it was a fake rule trying to see if they would break it or not. Or maybe they just don't care.	Because they were told to do it that way.
HIGH	151	To make more of a bonus	They followed the rules
HIGH	152	They did that because they wanted to maximize their profit by not wasting any of the time allotted.	They thought they had.
HIGH	153	They received more points for going to the end before the cross disappeared. 15 points to be exact. If they waited, they would only get 9 points.	They were told what the rules were and felt if they broke the rule, they wouldn't get paid.
HIGH	154	They thought it would result in easier points and a better payment.	It's easier to just do the task as it's designed than try to make up new ways of doing a simple task when the outcome likely won't change.
HIGH	155	They broke the rule because they are rewarded for how quickly they complete the task and there is no penalty for breaking the rule.	These people believe that following rules important whether there are consequences or not.
HIGH	156	To earn more money.	They would feel guilty breaking the rule for more money.
HIGH	157	more points	follow the rules
HIGH	158	Because they did not follow the instructions or read them properly.	Because they read the instructions.
HIGH	159	In a rush to complete the task.	They wanted to give the task requester good representation of what was asked of them.
HIGH	160	because they paid no attention to the instructions or were a little distracted while performing the experiment	because they paid no attention to the instructions or were a little distracted while performing the experiment
HIGH	161	They did not understand the instructions clearly.	They read the instructions and understood it clearly.
HIGH	162	maybe they can't control themselves	they were so attentive
HIGH	163	They would do that in order to maximize their potential earnings.	They would do that because following the rules is important to them.
HIGH	164	Yes. I can imagine the situation. Somebody didn't read the instructions correctly. So, they couldn't follow the rule.	17 MTurkers read the instructions clearly. So, they moved their circle after the cross had disappeared according to the instructions.
HIGH	165	I would think that they either did not follow the directions, they did not understand the directions fully, or they were just in a rush to finish. Maybe they wanted to move the dot before the cross appeared to see if they could move the study along faster. I do not really know of any other reasons that they would move before they are supposed to. If they misunderstood the direction and are trying to understand as they go. Maybe they did not read the instructions to move the dot at a specific time.	I would guess that they did not understand the directions, or they just did not want to follow them. Maybe they are trying to hurry the survey along. If they felt that the pay did not match up to the time, they had felt they spent in the task. Maybe they are trying to understand the instructions as they go. I am not sure of any other reasons someone would not want to follow instructions. Maybe this information was not part of the instructions. Who knows it really could be many reasons?
HIGH	166	Either they are greedy, bots, or an ESL and didn't understand the directions.	They are real workers who don't want to get rejected, despite less pay.

HIGH	167	The reason behind the event is that the 83 MTurkers may be didn't read the instructions correctly. So, they moved their circle before the cross had disappeared.	17 MTurkers read all instructions correctly. So, they could do the actions according to the instructions
HIGH	168	I feel that the eighty- three MTurkers wanted to make more money by finishing the task faster to receive a higher bonus. The eighty-three MTurkers that broke the rule, also may not have read the directions completely.	They followed directions and wanted to receive their bonuses.
HIGH	169	To earn more points	They wanted to follow the rules for fear of not doing the study correctly
HIGH	170	Higher chance of earning more points at the end of the study resulting in higher pay.	Either they believed they can somehow outsmart the task and earn more than moving before the task is finished or simply wanted to follow the rules as requested of them.
HIGH	171	Usually, all workers are different from one to one so many of them failed but only understand people can be successful in the process.	i think that people only attention to finish the subject to the study
HIGH	172	To attempt to get a higher pay-out by “beating” the system they agreed to.	They were simply following the rules.
HIGH	173	They either wanted to finish quickly or maybe the cross would reappear.	They had patience to wait and follow the rules.
HIGH	174	They did not understand the instructions.	Those were the ones that understood the instructions and wanted to maximize profit.
HIGH	175	They moved because they had a financial incentive to not wait. They waited if they could but decided against staying any longer because it was costing them money. People are not very patient anyhow, so this is no surprise to me. I would probably do the same thing.	I bet they are rule followers and generally follow what they are told even if it is costly to them.
HIGH	176	I think they understood the question involving the elapsed 5 seconds to mean that they could move while the cross was still visible. With that understanding they wanted to maximize their bonuses, so they cheated on the original instructions.	They were just following the rules that were laid out at the beginning of the study. They thought that it was important to follow instructions.
HIGH	177	They did it to score more points and get a better payoff. There was no penalty for not waiting on the X to go away so why not take advantage of it since it offers the most points	For a few reasons. One could be that they didn't read the directions and somehow got passed the screening questions. The other would be that they were being honest. This is how I would have played the game even if I would have lost money.
HIGH	178	They would have broken the rule because they would have gotten more points for finishing in a shorter time, and thus did not want to wait until the cross disappears to move to ensure the shortest time of completion of the task.	They may have thought that there would be some sort of penalty such as withholding of a bonus or a possible rejection of the task if they broke the rule. Alternately they were the sort of people who follow any rule or law they are given without questioning it.
HIGH	179	To get through the study quicker, they can complete more HITs without trying for the bonus	They were motivated by the chance of higher pay
HIGH	180	A small handful probably did not understand or read the instructions to begin with. I believe most of them chose to disregard the rule because it would mean a bigger payoff for them (15 cents over 9 cents) if they moved the circle into the black area in a lesser amount of time.	These people chose to follow the rule as it was explained to them probably out of fear of being rejected if they didn't stick to the instructions. They wanted to make sure they did the task as they were told.
HIGH	181	To finish the task faster and have less time therefore less money subtracted from their potential bonus.	In order to follow the rules and not be penalized.
HIGH	182	They wanted to save more points since waiting decreases their points every second.	They were honest and wanted to exactly follow the rules
HIGH	183	The workers probably did not read the instructions s fully and did not understand that they could not move until the cross is gone.	These workers read the instructions and understood to not move. They did not want to break the rules and forfeit a bonus.
HIGH	184	They either did not understand the rules or wanted a higher bonus.	They followed the rules.
HIGH	185	Could be either impatience of just not reading the included instruction material,	They wanted to adhere to the rules and follow them.
HIGH	186	Eighty-Three MTurkers of One Hundred MTurkers is very large, and I am still in shock that people don't follow rules. They may have reasons for not obeying the rules and the first i can think of is, Not understanding the rules or lack misunderstanding the rules. If people don't know what is expected of them, it is very easy for anyone to fail. Secondly, it is possible that the people in question did not pay	The seventeen percent did not follow the other multitude. But the eighty-three also may have reasons for not obeying the rules and the first i can think of is. Not understanding the rules or lack misunderstanding the rules. If people don't know what is expected of them, it is very easy for anyone to fail. Secondly, it is possible that the people in question did not pay attention to the information provided which makes them to misunderstand the information given as the rule guiding the experiment

		attention to the rules provided which makes them to misunderstand the information given....	
HIGH	187	They want to reduce the amount of time used to reduce the rate of decrease in endowment points.	They just believed in the rules.
HIGH	188	I suppose because you earn 6 more points (a total of 15) than if you move your circle after the cross has disappeared.	The only reason I could guess is because these people like to follow rules.
HIGH	189	They didn't read the rules carefully.	They read the rules carefully.
HIGH	190	The rules were set up to reward them for doing so. They earned more points by breaking the rule and finishing faster as opposed to waiting for the X to disappear.	They likely followed the rule because they felt obligated to conform to the rules of the study, even if it meant less points for them. They may have expected an additional reward for following the rules.

All responses from D-method. These were responses to the questions asked after showing subjects the graphical representations of possible outcomes from a hypothetical study.

CHAPTER 4 Cultural Orientation informs rule compliance

4.1 Abstract

Norms are the informal rules in society. Evidence suggests a systematic variation in a country's prevalence of rule violations (Gächter & Schulz, 2016) and its degree of individualism (IDV): high IDV countries follow rules (against corruption, tax evasion and political fraud) more frequently than low IDV countries. Rule compliance varies, but the internal mechanism of this difference in outcomes has not been explored much. Recent research demonstrates that people's preference to follow or violate a rule may be shaped by the underlying societal beliefs of such outcomes. The strength of these beliefs may be influenced by the nature of the social reference groups and formal institutions that play a role in the process of social learning. To evaluate how such groups shape aggregate outcomes through their influence on people's beliefs and conditional preferences, we experimentally test the Bicchieri (2016) norm framework in a HIGH IDV country (Sweden) and a LOW IDV country (Turkey). We measure compliance using a simple computerised rule following task and elicit beliefs using direct elicitation and Krupa-Weber (2013) norm-elicitation. We elicit preference conditionality using the variant of the strategy method (Gächter et al., 2021). We find a few interesting results. First, rule compliance is higher in Sweden. Second, Swedish subjects hold more accurate empirical beliefs than Turkish subjects. Third, both countries hold strong normative beliefs of compliance but vary in their perception of the normative standard of violation. In other words, Swedish subjects hold harsher beliefs against violation. Fourth, although subjects from both countries display preference conditionality as a function of increasing violation, this influence is slightly stronger in Turkey. Finally, increasing information of disapproval does not affect subjects' compliance preferences from both countries.

4.2 Introduction

Norms are the shared informal rules (e.g., Bicchieri, 2005, 2016; Elster, 1989; Young, 2015) based on the social beliefs that are established and dissipated by the mechanism of shared social learning through members of society. The content of these beliefs and the preference to conform, conditional on such beliefs, is shaped by the cultural orientation. Culture can be defined as "those customary beliefs and values that ethnic, religious and social groups transmit fairly unchanged from generation to generation" (Guiso et al., 2006, p. 23). Although there are many ways of defining national culture and the cultural orientation of a country, shaping its social structure, we focus on the orientation of individualism (IDV), which has been extensively used in cross-cultural research (e.g., Roland, 2015). This is of central focus since it shapes the role of reference groups and formal institutions, which influence people's beliefs and conformity attitudes.

Research in the social sciences has investigated how aspects of norms, both systematic behavioural patterns and the underlying normative standard of such behavioural outcomes, and culture interact and inform one another. For instance, empirical evidence has found a positive association between a country's level of rule or norm compliance and individualism (e.g., Gächter & Schulz, 2016; Licht et al., 2007). Moreover, people’s social reference groups influence the nature of beliefs shaping the norms and the norm-enforcement mechanism. For instance, countries that exhibit strong social ties between in-group members, a key characteristic feature of the cultural orientation of low individualism (LOW IDV henceforth), focus on group harmony-based values and enforce norms on group members only, and thus influence the normative standard of some behavioural outcomes differently from highly individualistic (HIGH IDV) countries (e.g., Alesina & Giuliano, 2015; Enke, 2019; Tabellini, 2008).

However, all these studies investigate norms and their underlying aspects of beliefs and preferences in different ways. For instance, the studies that have established the notion that countries reflecting HIGH IDV follow rules more efficiently than low IDV countries have not investigated the reasons as to why some countries follow rules more than others. Moreover, suppose beliefs have a role to play in complying with rules. In that case, there is not much evidence to see if LOW IDV countries, who have a high degree of rule violation, have weak normative standards of such an outcome. Specifically, there is not much information to say whether countries that violate rules more than others have weak normative considerations of rule-compliance or of violation or a combination of both. Additionally, a few studies in the honesty literature (e.g., Cohn et al., 2019; Hugh-Jones, 2016) have found that people have inaccurate or vague beliefs regarding other's honesty (an intrinsic form of rule-compliance). This raises the question of whether people are cognizant of what most other people around them would do, especially regarding rule violation, and if such cognizance differs between countries with HIGH and LOW IDV.

Motivation to conform may be attributed to the ideas of "rational herding" (Banerjee, 1992), "social proof" (Cialdini, 1993), or other related reputational concerns (Akerlof & Kranton, 2000), all of which are based on peer group influences. These influences may be via the beliefs of what their group members do (empirical beliefs) and approve of being done (normative beliefs). In a recent study, Gächter et al. (2021) found that people's preference to conform and subsequently comply with a rule is indeed conditional on such empirical and normative beliefs. However, if rule compliance varies across the globe, it seems intuitive to expect heterogeneity in preference conditionality with such beliefs across countries with varying degree of IDV.

There is not much evidence on how people' conformity attitude, conditional on such beliefs, may be shaped differently by IDV through their underlying societal differences.

Additionally, the limited research in culture and conformity is quite mixed and unexplored. On one end, research shows that people from LOW IDV tend to conform more than those exposed to HIGH IDV, simply due to the relevance of group thinking over individualistic thinking (Bond & Smith, 1996). However, conformity preferences may vary depending on the in-group/out-group relationships. The evidence of limited versus generalized values (Alesina & Giuliano, 2015) that differ between such societies shows that rules and normative standards do not apply equally to all members in LOW IDV. Infact, people care less if the violation happens outside the group by others in LOW IDV. Likewise, out-group members may not have a role in shaping conformity and compliance with a rule. In other words, the effect of social disapproval might vary, depending on whether the disapproving agent is a member of the in-group or out-group, and this might influence conformity preferences differently.

To summarize, reference groups and their underlying belief systems within countries have a crucial role in shaping people's preferences, contributing to countries' attitudes towards adherence to rules. Additionally, in LOW IDV, since these groups have a more significant influence in shaping beliefs, they might weaken the influence of formal institutions regulating norms, consequently affecting aggregate compliance outcomes.

The key research objectives of this study are as follows: first, we aim to demonstrate that rule compliance varies between countries with a HIGH and LOW IDV. Second, we investigate if people hold accurate beliefs about the behavioural outcome of rule violation and compare if similarities or differences exist between such countries. Third, we look into the normative

standard of compliance and violation. In other words, we explore whether significant differences in the degree of rule-violation between HIGH and LOW IDV countries can be attributed to different strengths of the normative content of such an action. Fourth, we investigate whether cross-cultural variation exists, in preferences to follow (or violate) a rule, conditional on an increasing function of others' violation around them as well as on others' disapproval of violation around them. Put differently, we aim to check if people's preferences are dependent on these two types of social information and if cross-cultural variation exists in the degree of conformity, to account for the differences in the level of rule violation between HIGH and LOW IDV.

Hence, this study aims to see how societal beliefs, preferences and the behavioural outcome of rule compliance interact and influence one another. This chapter involves three sets of experiments conducted in two culturally distinct countries with regards to their societal structure. They are Sweden (SWE), a country with a cultural orientation of HIGH IDV, and Turkey (TUR), which reflects a cultural orientation of LOW IDV.

Experiment 1 investigates rule-compliance using an individual-level behavioural task in these two culturally divergent countries. In Experiment 2, we then measure the societal beliefs of violation, i.e., the empirical beliefs of what most others do (in a rule following task) and the normative standard of compliance and violation. We capture this using direct belief elicitation for empirical beliefs and the Krupka-Weber (2013) norm elicitation method (henceforth KW) for normative beliefs. Experiment 3 investigates people's preference to comply with or violate a rule based on social information. More specifically, we measure people’s preference towards rule violation as a function of a) others' decisions to violate the rule and b) others’ disapproval

towards such violation. We use the variant of the strategy method for both preference elicitation experiments, adapted from Gächter et al. (2021).

We find that rule compliance is significantly lower in the LOW IDV, as seen from the rule following task in Experiment 1. With empirical belief elicitation in Experiment 2, we find that the HIGH IDV reflects (more) accurate beliefs of rule-violation compared to the LOW IDV. However, there is no difference in the average elicited beliefs between the two countries. Moreover, we find a clear consensus regarding the normative content of social appropriateness of rule compliance in both countries. On the other hand, we find evidence of no clear collective consensus of the normative content of rule violation in the LOW IDV only. Subjects in SWE evaluate violation more harshly than subjects from TUR. In Experiment 3, the preference elicitation experiment with increasing violation rates, we find the following: subjects in both LOW and HIGH IDV display a preference to conform and violate as a function of increasing violation around them. We also find a higher share of preference to violate in the LOW IDV. This aligns with the low compliance rate (Experiment 1) and weak normative strength of violation (Experiment 2) in TUR, the interplay of which might lead to higher conformity towards violation here. On the other hand, we find no effect on subjects' preferences to follow rules conditional on an increasing information of others' disapproval (towards violation) in both countries.

The organization of this chapter is as follows: Section 4.3 discusses the related literature, followed by Section 4.4, which describes the experimental design. Section 4.5 discusses the related hypotheses. Section 4.6 presents the results; and Section 4.7 discusses the results and concludes.

4.3 Related literature and our contribution

This section provides an in-depth review of literature. In Subsection 4.3.1, we discuss the fundamental aspects of culture, composed of societal belief systems and individual preferences, by reviewing some important studies in economics. In Subsection 4.3.2, we then proceed by illustrating how national culture is measured, especially the breakdown of national culture into core dimensions encompassing the key elements and the importance of the dimension of individualism (IDV). Since the investigation of norm compliance is the primary goal of this study, we then discuss the Bicchieri framework for social norms (2005; 2016), which demonstrates that accurate beliefs and preference conditionality are the two main prerequisites for norm compliance in Subsection 4.3.3. Next, in Subsection 4.3.4, we discuss how these prerequisites of beliefs and preferences may be influenced by the cultural orientation, specifically by the cultural dimension of individualism (IDV). In Subsection 4.3.5, we briefly discuss the two countries: Turkey and Sweden, which differ on the IDV dimension and are a part of this study. Finally, in Subsection 4.3.6, we review relevant literature to provide the foundation for testable hypotheses regarding compliance, beliefs, and conditional preference outcomes from the three experiments. We then briefly summarise our contribution to the current literature in Subsection 4.3.7.

4.3.1 Culture in economics: Introduction

Many studies in social psychology and economics involving inter-disciplinary research have tried to define culture and find linkages between various aspects of culture (beliefs and preferences) and their effects on economic and non-economic outcomes. For instance, Tabellini (2010) defined culture in terms of trust, beliefs in individual effort, generalised morality, and level of obedience. He also used measures of education and political institutions in 69 European countries. He found that countries with higher trust, belief in individual effort,

generalised morality, and low obedience reflected higher GDP per capita and economic growth. Guiso et al. (2006) took the time-invariant aspects of cultural traits such as ethnic background and religious affinity in a different study. They found that these traits influenced beliefs of trust and redistribution preferences that ultimately affected the subsequent economic outcomes. Alesina and Giuliano (2015) summarised the studies investigating the interactions of culture, defined both in terms of beliefs, preferences, and institutions and found some critical correlations between societal structure and general moral and trust standards. Specifically, they found that countries that exhibited individualistic structures with weak family ties correlated positively with generalised morality and trust than countries lower on the individualism scale, i.e., the collectivist countries.[43] For instance, northern Europe, the USA, Australia, and New Zealand had higher trust and morality levels than their Asian, South American, and African counterparts[44]. All these studies demonstrate how culture is usually viewed through societal beliefs and preferences and that it affects both economic and behavioural outcomes.

4.3.2 Culture through IDV

Research in the social sciences over the years has observed and established the notion that the cultural orientation of a group help defines the structure of societies, shape the underlying value system, and guide people's beliefs, and consequently, their preferences. This thought has been extrapolated to define and categorise countries based on their societal structures and approach to resolve collective issues. Social psychologist Geert Hofstede (1980; 2001) defined culture as "the collective programming of the mind that distinguishes the members of one group or

[43] Generalised morality and trust refer to the idea that morals on cooperation and trust values apply to all members of the society. In contrast, limited morality and trust encompass situations where these values hold for ingroup members only (Tabellini, 2010).

[44] In a different line of study, Falk et al. (2018) contributed to the preference literature by collecting survey data on time, risk, and social preference such as positive and negative reciprocity and trust from 76 countries with 80,000 individuals. They found both between and within-countries heterogeneity in the preferences mentioned above.

category of people from others." It is a collective phenomenon and can be applied to any group of individuals with shared knowledge, understanding and practice. Societal, national and gender cultures are deep-rooted cultural orientations learned early on in life, affecting people's beliefs and preferences that shape economic and non-economic outcomes[45].

Although there are numerous issues that a country can address through their values and norms, Hofstede and previous literature in cross-cultural research focus on a few core issues. Based on these issues, Hofstede (1980; 2001) defined national culture along six orthogonal dimensions, each measured on a scale of 0-100. A key dimension is individualism (vs collectivism), which looks into the looseness and tightness of social ties within group members in society[46].

Although there are many ways of looking at national culture[47], we primarily focus on the critical dimension of individualism[48] (IDV henceforth), from Hofstede's dimensional approach. IDV defines an individual's identity through oneself or a group and imparts the relevant values and beliefs that may shape their preference for conformity and behavioural outcome of compliance. The IDV societal norm "stands for a society in which the ties between individuals are loose: everyone is expected to look after himself and the immediate family.

[45] Additionally, Hofstede conceptualises culture and societal norms with varying layers. First, the outer layer comprises visible practices such as rituals and symbols, and the inner core encompasses the group's value system in focus. Both, in turn, lead to the creation and maintenance of family systems and formal structures such as education, legal and political systems.

[46] Second is the power distance that reflects the degree of social hierarchy defining the social structure. The third dimension is masculinity (vs femininity), which represents whether a society reflects a materialistic, competitive, assertive, or a more cooperative, and caring preference. The fourth dimension is the uncertainty avoidance index, which captures the country's mindset about uncertainty and ambiguity. The fifth dimension is long term versus short term normative orientation, which reflects how adaptive a country is towards change in values and norms. Finally, the sixth dimension is indulgence versus restraint, which shows countries attitudes towards enjoying hedonic pleasures in life.

[47] Social psychologist Shalom H Schwartz (1994) classified countries based on key dimensions based on specific fundamental core values used by people to solve societal problems. His theory is based on bipolar dimensions, which reflect the two possible ways that societies can confront problems; these dimensions are autonomy (vs embeddedness), egalitarianism (vs hierarchy), and harmony (vs mastery).

[48] This dimension is similar to Schwartz's value orientation of autonomy versus embeddedness.

Collectivism stands for a society in which people from birth onwards are integrated into strong, cohesive in-groups, which throughout people's lifetime continue to protect them in exchange for unquestioning loyalty" (cited from Hofstede, 2001; pp 225).

IDV is one of the most researched dimensions in inter-disciplinary research (Roland, 2015)[49]. Countries with a LOW IDV score shift towards a collectivist orientation, while countries with a higher IDV score reflect an individualistic social structure. Furthermore, this dimension of a country's societal system is associated with economic growth (Gorodnichenko & Roland, 2011); gender equality (Davis & Williamson, 2019), democratisation (Gorodnichenko & Roland, 2021), general trust and moral values and higher quality of institutions (Alesina & Giuliano, 2015). As a result of these associations, generic inferences can be made regarding the socio-economic scenario just by looking at its IDV score. Additionally, and more importantly, the societal structure's characteristics provide the basis for the elements of the norm's framework regarding attitudes of conformity, beliefs, and relation to reference networks.

4.3.3 Social norms: Bicchieri's Model

Although various theories exist for social norm compliance across multiple disciplines, we focus primarily on the theoretical framework given by Bicchieri (2005; 2016). This framework captures the key elements that play a role in norm compliance, namely a) people's reliance on societal beliefs regarding what others do and believe ought to be done and b) peoples' attitude towards conformity. The formal definition of social norm is the following: "A social norm is a

[49] Another relevant dimension is the power distance (PDI), which investigates power, social hierarchy, and inequality. Corruption is common in countries with high PDI while low in countries with low PDI. It should also be noted that the IDV and PDI dimension are inversely correlated: countries that reflect an individualistic structure also demonstrate low PDI, whereas countries that are LOW IDV (or collectivist) reflect high PDI. This correlation remains even after controlling for economic wealth.

rule of behaviour such that individuals prefer to conform to it on condition that they believe that a) most people in their reference network conform to it (empirical expectation) and b) most people in their reference network believe they ought to conform to it (normative expectation)" (cited from Bicchieri, 2016, p.35).

This framework provides the underlying mechanism of rule compliance which would help understand why some countries follow rules more efficiently than others. Moreover, the elements of beliefs and preference conditionality can be measured using experimental tools, which would then help make cross-cultural evaluations between countries of distinct cultural orientations.

4.3.4 Linkages between IDV and Bicchieri's Model

In individualistic countries, social groups are larger, with loose social ties between them. As a result, members of the society face less social pressure from their reference networks to adhere to values and conform to norms focusing on group harmony. Consequently, there is less dependence on social learning from group members. Additionally, there is more focus on the universal values and norms of honesty, hard work, discipline, fairness, equality etc. Moreover, these societies also tend to be egalitarian, which means that all members of the society are treated equally.

As a result of these social characteristics, in countries with a HIGH IDV orientation, there is a weak influence of social groups. On the other hand, social learning of universal values is through the reliance on formal institutions. Due to the egalitarian nature of these societies, and the presence of efficient formal institutions, the same rules and punishment from transgression apply to all, enforced through such institutions. The interplay of the focus on universal norms

and reliance on formal institutions for norm adherence lets people from individualistic countries witness more societal transparency. This helps form accurate beliefs and leads to better macro-level societal coordination.

On the other hand, in LOW IDV countries, there are more significant social groups, and the social ties between the group members are very tight. Since group harmony is the focus of such cultural orientation, people tend to emphasise the group values of loyalty, agreeableness, and obedience, etc., which may in turn contradict the universal norms. For instance, nepotism or ingroup favouritism may be acceptable and critical to maintaining group dynamics. However, it may interfere with and erode the normative standards of a universal rule of fairness towards everyone. Additionally, these societies tend to maintain social hierarchy between group members. In other words, not all group members are viewed as the same: some are the leaders, while the rest follow the values and morals given by these "informal leaders", which are primarily based on maintaining group harmony.

As a result of these factors, people in such countries tend to focus more on group-based values than universal values and social learning is through the informal social communities than formal institutions. People form beliefs about their group members' actions and opinions regarding the socially appropriate outcomes, leading to more micro-level group coordination. Thus, the effect of these reference groups is more substantial in the LOW IDV countries. Consequently, formal institutions have a weak role, and informal social groups regulate most social learning and norm-enforcement.

Thus, the key social differences between LOW and HIGH IDV orientation, lies in the strength of social groups, and consequently the learning of group (LOW) versus universal (HIGH)

values through informal leaders (LOW) versus formal (HIGH) institutions. This, in turn, might shape aggregate compliance outcomes as well as the beliefs regarding such actions differently. A summary of the key differences in characteristics of the cultural orientation of HIGH and LOW IDV is listed in Table 4.1 below.

Based on these differences in the role of reference groups between the HIGH and LOW IDV countries, we will review how such influences may affect people's beliefs and conformity preferences, thereby affecting their adherence to rules. We will focus on the key elements of the Bicchieri framework, namely a) preference conditionality and b) social influences, i.e., beliefs regarding what others do and expect ought to be done, through the role of reference groups.

First, we investigate the effect of the reference groups in the LOW and HIGH IDV cultural orientation on peoples' empirical beliefs, i.e., beliefs regarding others' actions. In HIGH IDV countries with egalitarian social structure, due to a) the absence of tight reference groups, b) the stronger reliance of formal institutions for social order, and c) the priority on learning of universal norms, people can frame accurate expectations or a collective belief of other's actions (pertaining to such norms) in the society.

On the other hand, in countries with LOW IDV orientation, forming a collective societal (empirical) belief is relatively more complex. People in these societies belong to many social groups and get influenced by each of these groups. Due to the strong inter-personal ties between group members, and frequent exposure to each other's actions, people may be able to form accurate beliefs of their immediate group members' behavioural patterns. However, all these fragmented social groups may have their own norms and values, which may not align to

provide a collective societal belief. For instance, regarding empirical beliefs of corruption, people might be exposed to corrupt practices in the work environment. Still, they might be part of families and social communities that practice ethical and honest values. As a result of such contradictory actions of others around them, people might have accurate empirical beliefs about a specific social group. However, they may be vague regarding the empirical beliefs of the collective society.

	HIGH IDV	**LOW IDV**
1.	Focus on *individual thinking*, identity, and uniqueness.	Focus on *group thinking*, identity and conformity.
2.	Focus on *universal moral values*: honesty, fairness, equality, *applicable to all* members of society equally.	Focus on moral values limited to maintaining *group harmony:* loyalty, obedience, respect for authority; *applicable to in-groups' members* only.
3.	Reliance on *formal institutions* for social order and normative messaging	Reliance on the (more) powerful *reference group members* for social order and normative messaging
4.	Presence of *strong societal beliefs* and macro-level *societal coordination*	Presence of strong but *fragmented reference groups' beliefs* and *micro-level group coordination*
5.	Reflect good formal institutions; *reinforce and strengthen normative standards.*	Reflect poor formal institutions; *weaken and erode normative standards*.
6.	Less(er) social groups, and looser ties leading to social identification with lesser reference groups; Less*(er) degree of social conformity*	More social groups with stronger ties; social identification with multiple reference group's stronger *social conformity with (more) socially similar groups.*

Table 4.1: Key Differences in high and low IDV.

Next, we examine how these reference groups may influence the normative content and beliefs in HIGH and LOW IDV. Normative content is extracted from moral systems that help maintain social order and resolve social dilemma problems. These moral systems comprise values and beliefs transmitted through members of society and diffused into the social environment. Countries with LOW IDV scores have exhibit the notion of "limited morality" (Tabellini, 2008). Here, trust and moral values are enforceable among in-group members only.

Additionally, in such countries, due to a fragmented social structure consisting of multiple social groups and unequal power dynamics between group members, the content of moral values also remains limited, aimed primarily at maintaining group harmony only. As a result, limited values such as loyalty, obedience, respect for authority, agreeableness are encouraged and transmitted through in-group social learning (Enke, 2019). People from such countries might thus have informed beliefs regarding group-specific norms but not necessarily have accurate knowledge of universal norms. Additionally, people might project their in-group beliefs of what group members do on the general behavioural expectations of the country. For instance, people from LOW IDV countries might form accurate beliefs regarding a certain rule of obeying authority but may be vague about a universal norm of rule compliance. In LOW IDV countries, the informal social groups, as in the case of empirical beliefs, have a more assertive role than formal institutions regarding the creation of normative judgments. Additionally, the more powerful in-group members regulate the norm-enforcement of such local norms (which is prioritised over universal norms). As a result of these factors combined, a shared strong normative consensus over such norms may not exist in these countries.

On the other hand, in a HIGH IDV, the universal norms are more relevant than in a LOW IDV society. Even the limited informal social groups in such societies prioritise the social learning of the normative values of such norms. Additionally, the normative messaging of such values is regulated through formal institutions that punish transgressions and reinforce the approval or disapproval of specific behavioural patterns.

Finally, none of the factors mentioned above have a role in decision-making if individuals do not prefer to follow (or violate) a norm based on the beliefs shaped by the reference groups or formal institutions. If people display conformity, i.e., the preference to align one's behavioural

response based on the social influences around, only the empirical and normative beliefs or social influences of relevant others matter. People's conformity preferences may depend on the nature of action, the size of the reference' group, and the social closeness of the group to be followed. For instance, negative social information is more influential than positive social information (e.g., Croson & Shang, 2008; Keizer et al., 2008) and group size matters for conformity (e.g., Bond, 2005).

Conformity attitudes vary in strength depending on the cultural orientation of individuals. Bond and Smith (1996) performed a meta-analysis to study the relationship between cultural orientation (particularly IDV) and conformity using studies based on Asch-type conformity tasks. They found that countries low on IDV displayed higher conformity than their counterparts from HIGH IDV countries. This could be because the social fabric of such cultures is based on group-thinking where individual unique thinking is not emphasised. As a result, thinking and learning processes become dependent and conditional on others.

4.3.5 LOW and HIGH IDV: Turkey and Sweden

We chose Turkey and Sweden for our cross-cultural experiment regarding testing compliance and various aspects of the Bicchieri (2005; 2016) framework of social norms. Both are industrialised countries, with different degrees of cultural orientation of individualism. Turkey, an upper-middle-income country, has an estimated population of 83 million with a GDP per capita (2021) of $9406 and an IDV score of 37 (out of 100). Henceforth, Turkey (TUR) is considered a LOW IDV country. Sweden (SWE), on the other hand, has a smaller population of 10.38 million, with a GDP per capita (2021) of $58100. It has an IDV score of 71 (out of 100), and thus subsequently, we consider it a HIGH IDV country. Additionally, the PDI score from Hofstede dimensions represents the degree of social hierarchy within the countries: TUR

has a higher score of 66, while SWE has a lower score of 31. In other words, TUR reflects unequal power between members of the society, while SWE has a relatively more egalitarian social structure[50].

4.3.6 Literature related to the hypotheses

We will now review some key research that would help us frame the hypotheses from the three set of experiments. We first discuss the literature that helps predict the hypothesis associated with the outcome in the simple rule-following task in the two countries. Second, we discuss the literature to help build the hypotheses related to people's empirical and normative beliefs in these countries. Third, we discuss studies that may shape the hypotheses related to people's attitudes of conformity regarding empirical and normative beliefs.

Licht et al. (2007) investigated the association between norms of governance defined by the rule of law, corruption and democratic accountability, and cultural orientation of countries as seen through individualism and egalitarian social structures and found a positive association between these. They provided the rationale that corruption cannot sustain in an egalitarian setup since it is based on the use of power and disrespect towards the less powerful, which contradicts the value orientation of equality in such countries. Likewise, since the cultural orientation of individualism emphasises individual thinking and independent actions, these countries demand transparency in society regarding how people behave and how institutions manage the rule of law to help form accurate beliefs of actions and opinions around them.[51] Gächter and Schulz (2016) investigated how the prevalence of rule violations (PRV) in a country influences the

[50] We tabulate the cultural dimensional scores and some of the key socio-economic differences between these countries in Table A4.1 in Appendix.

[51] Some other studies that have investigated these associations: Klasing (2013) found individualism and social hierarchy as predictors of the quality of formal institutions; Hugh-Jones (2016) found that dishonesty was negatively linked with GDP per capita in 2011 and had a positive relationship with collectivist countries. He reasoned this behaviour by suggesting that individuals from developing or under-developed economies have a greater incentive to lie if they benefit from an extra monetary gain than people from more prosperous economies.

intrinsic honesty of individuals. They measured individual-level intrinsic honesty by conducting a die-rolling experiment in selected countries and then found an inverse relationship between honesty and IDV. Countries that exhibited a low prevalence of rule violation and consequently seemed to have an individualistic orientation, had a higher level of intrinsic honesty than collectivist high PRV countries. They found that an increased incidence of such violations led to the normalisation of such actions, making them acceptable standards of behaviour. Additionally, they also observed that frequent violations and exposure to such violations in high PRV countries made people more cognizant and accepting of the unfairness in society over time.

Another line of literature points to the systematic relationship between corruption – another major form of rule violation and the underlying societal structures or cultural orientation (see Banuri & Eckel, 2012 for a review). In fact, there are various channels through which culture influences people's attitudes, perceptions, and preferences towards various forms of corrupt activities. Treisman (2000) investigated various reasons for the origin of corruption in societies, and one of the key observations was that countries following a protestant culture had lower levels of corruption.[52] This could be attributed to the fact that protestant cultures teach values such as self-reliance or individualism and less importance to family ties and in-group favouritism or nepotism. Serra (2006) ran a similar study, and the empirical evidence reinforced earlier suggestions: richer countries following a protestant mindset had lower levels of corruption. Observing the Inglehart and Welzel (2010) cultural map (WVS, Wave 6), we can say that SWE follows a Protestant culture.

[52] An important rationale for corruption in LOW IDV is the lack of societal transparency: Lopez & Santos (2014) observed that countries that were inclined to have hierarchical social structures had higher corruption levels; they attributed this occurrence to the unequal distribution of power, with more power to authoritative figures that led to a lack of transparency. This incentivised such figures to cheat and get involved in corruption.

All these studies show two important findings. First, there is a certain variation in the degree of adherence to rules across countries. Second, countries with relatively higher income levels, HIGH IDV orientation with equal power amongst all members, and generalised moral values and norms tend to follow rules more than countries with collectivist ideologies.

Next, we review the literature to understand how both empirical and normative beliefs may be affected by IDV. There are still limited studies that look into people's beliefs regarding rule following and norms in their society. Hugh-Jones (2016) investigated honesty, beliefs of honesty and economic growth in 15 countries. He found that honest behaviour was correlated with economic growth as measured by GDP per capita, but beliefs of honesty did not match with actual behavioural patterns[53]. Such beliefs could be shaped by personal experiences and prevalence of honest or dishonest behaviour around, or other cognitive biases such as social projection, whereby people project an expectation on others based on their own preference and behaviour. Cohn et al. (2019) study investigating civic honesty elicited people's empirical beliefs about such an action[54]. While behavioural results showed that people tended to return the lost wallet with more money, the respondents' elicited empirical beliefs regarding civic honesty were in odds with the actual behavioural outcome. While the chances of reporting a "lost wallet" were the highest when the wallet contained a substantial amount of money,

[53] The 15 countries were a mix of both western and eastern countries and thus contained both HIGH and LOW IDV(s).

[54] They conducted large scale field experiments in 40 countries to highlight civic honesty worldwide. In their experiment, the confederate reported and left an experimentally manipulated "lost wallet" with no, low, or high amount of money at various public institutions such as museums, theatres, banks, post offices, hotels, and other public offices. They predicted that there would be no (or few) reporting and returning the wallets, thereby catering to self-interest. However, they witnessed the opposite behavioural outcomes with the highest rate of reporting in high, followed by low and no money treatments, which was an interesting result. However, in a recent study by Du et al. (2020), the authors re-analysed the Cohn et al. (2019) data in the context of inequality and GDP measures. Specifically, they regressed the rate of return of a lost wallet on countries' level of income inequality and wealth measured by Gini index and GDP variable. They found that the rate of return was inversely associated with the level of income inequality: countries with high-income inequality showed a lower return rate. This result remained robust after controlling for various other socio-demographic and cultural co-variates.

respondents believed that reporting a wallet would be the highest when the wallet contained no money, depicting a low expectation of civic honesty around. Interestingly, people from the US (HIGH IDV) expected others to return a wallet less when it contained more money, reflecting low expectations of honesty. Although USA has a HIGH IDV orientation, it has many heterogeneous reference groups. Therefore, forming an accurate empirical belief of a meta-rule (in this case of honesty) may be difficult in such a country. These studies show that individuals generally have inaccurate beliefs regarding other's action, especially with regards to an intrinsic rule of honesty. Therefore, we expect to find high variation or vagueness in the stated empirical belief in a country with many heterogeneous groups or low IDV.

Next, we review studies that reflect variation in normative standards of certain behavioural outcomes across countries. Countries with LOW IDV, with weak governance norms (Licht et al., 2007), also exhibit a higher prevalence of rule violation (Gächter & Schulz, 2016). As a result of such a high incidence of many kinds of violations in everyday life, people from such countries may eventually update their normative considerations of rule violation. More specifically, due to the prevalence of such violations, the strength of social disapproval towards violations might erode over time. For instance, Cameron et al. (2009) found that Indian subjects (LOW IDV) not only engaged more in corrupt decisions in a bribery game but also displayed a higher tolerance towards corrupt behaviour compared to their Australian counterparts (HIGH IDV)[55].

[55] Cross-cultural experiments in cooperation (e.g., Gächter & Herrmann, 2009; Gächter et al., 2010) show that societies reflect heterogeneity in attitudes towards cooperation and disapproval (through punishment mechanisms) towards selfish self-interest. Henrich et al. (2006) carried a set of field experiments with the ultimatum, public good and dictator games in various societies around the world. They found that people punished the less than equal amounts, implying similar disapproval towards such offers everywhere. However, societies displayed variation in the degree of disapproval.

Next, we review the literature related to research in conformity. Conformity can be due to multiple reasons: one of the earliest rationales for conformity was through a social learning strategy known as “rational herding”, whereby people imitate and follow other members of society in the face of uncertainty. This is based on the assumption that others are better-informed (Banerjee, 1992; Bikhchandani et al., 1998). A similar explanation of learning strategy is given through the “social proof theory” (Cialdini, 1993) in social psychology, which posits that people learn through social comparisons with others. A different rationale for conformity comes through the notion of “group or peer pressure”, which considers other inherent values such as reputation concerns (Akerlof, 1980) or social status and acceptance (Bernheim, 1994).

Certain factors such as the nature or content of the action, the size of the group to be followed and the social ties or proximity to the reference group, both by themselves and through their interplay, have influenced people’s preference of conformity. For instance, a few studies show that between positive and negative social information, people tend to act following the negative information, which has a more substantial effect on people’s decisions (Bicchieri & Dimant, 2019; Croson & Shang, 2008). Next, many studies show how peer groups influence bad behaviour through such conformist attitudes. Oliveira et al. (2014) investigated the impact of such peer influences in group dynamics in public goods provisions. They found that the selfish group members contribute less themselves and influence the group's contribution level and lead to erosion of cooperation. Gino et al. (2009) found such peer influence in the context of dishonesty. Falk and Fischbacher (2002) found similar peer effects in influencing criminal behaviour in the lab.

Additionally, group size also matters for conformity. Bond (2005) conducted a meta-analysis of 125 Asch-type experiments that confirmed the effect of group size on conformity, although the nature of association may differ[56]. Sasaki (2019), in a field experiment involving charitable donations in Japan, manipulated the size of the reference group and subsequently observed a positive effect on people's donation decisions. Egebark and Ekström (2017) conducted a field experiment using Facebook, where they investigated the impact of group size and social ties on people's conformity decision to "like" content. They found that subjects were more likely to conform and like the content when the reference group's size (of strangers) increased from 1 to 3 and when the reference was composed of a friend instead of a stranger.

Bond and Smith (1996) conducted a meta-analysis using 133 Asch-type conformity studies in 17 countries. They found that collectivist countries reflecting a LOW IDV score depicted higher conformity than countries with a HIGH IDV orientation. Cialdini et al. (1999) investigated the willingness to take part in a survey without pay in Poland (LOW IDV) and the USA (HIGH IDV) to test theories related to social proof and consistency and found a stronger effect of social proof in Poland than the USA. Mesoudi et al. (2015) conducted a social learning-based experiment with Chinese individuals living in Mainland China, Hong Kong, and the UK. They found that the Chinese individuals in Hong Kong and the UK displayed more asocial learning like the West due to its presence and influence than the Chinese who resided in Mainland China.

We use the evidence from all these studies to build testable hypotheses in Section 4.5.

[56] Muthukrishna et al. (2016) also presented a formal model that showed that the strength of conformity depends on the group size, based on the number of traits in the environment.

4.3.7 Our Contribution

Looking at the vast majority of the relevant literature, we find that there is also a systematic pattern to variation in rule violation: countries with HIGH IDV have a lower incidence of rule violation than countries with LOW IDV. We establish this result and contribute to the robustness of this behavioural pattern through our simple rule following task which is devoid of any potential framing effects in a LOW and HIGH IDV. Additionally, we dig deeper into the possible mechanisms of such outcomes. There is not much research into investigating the normative perceptions of rule-following in LOW and HIGH IDV(s). More specifically, there is not much to see if high rule violations in LOW IDV can be attributed to weak norms towards compliance or weak norms against violations.

We add to this line of thought by checking beliefs from a HIGH and LOW IDV after establishing that rule violation differs between these countries. Societal beliefs are shaped by the social structures involving formal institutions and informal communities that differ between a LOW and a HIGH IDV. While in HIGH IDV, regulation of rules is through formal institutions, in LOW IDV, this is influenced by the informal structures of social groups. It is interesting to see how such reference groups may shape beliefs. This, in turn, might help understand the variation in rule following.

Additionally, current literature has shown that people's conformity is based on information content (bad behaviour is easier to imitate), group size, reputation concerns, etc. However, the evidence on culture and conformity is limited and mixed (e.g., Bond & Smith, 1996; Tabellini, 2009). Gächter et al. (2021) employed a variant of the strategy method in the context of an abstract rule following task to investigate conformity with rule compliance based on social information. They found evidence of preference conditionality in compliance outcomes based

on both what others do (empirical information) and also approve of doing (normative information).

Since behavioural outcome (of compliance) and the underlying belief systems varies between LOW and HIGH IDV, it seems intuitive to expect that conformity attitudes with regards to empirical and normative information, may vary between such societies. We contribute to the literature in conformity in two main ways: first we replicate the Gächter et al (2021) preference conditionality experiment in a cross-cultural setup to review the robustness of their result. Second, we investigate if and how conformity attitude might vary between a LOW and a HIGH IDV. More specifically, we investigate how subjects from a LOW and HIGH IDV country perceive the information of others' violation, which may or may not affect their own compliance preference. Moreover, we also check if others' (strangers from their country) disapproval matters for such a preference, and if this influence varies between the two countries. For instance, in Turkey, which is our LOW IDV country with tight-knit groups, people might not be influenced to follow rules, with disapproval from strangers. On the other hand, such a social influence may work in Sweden, our HIGH IDV country since rule following is expected to be more internalised.

4.4 Experimental Design

4.4.1 General Setup

To see how the cultural orientation of individualism (HIGH IDV) or collectivism (LOW IDV) informs the belief system (involving empirical and normative beliefs) and people's preference to conform, we carried out experiments that elicit these elements that are key to the social norm framework given by Bicchieri (2005; 2016), in Turkey (LOW IDV) and Sweden (HIGH IDV). To identify and observe if these components vary between two culturally different countries in

terms of their IDV, which in turn may affect their societal beliefs and attitude of conformity, our experiments comprised of a three-part study (see Figure 4.1) in Turkey (TUR) and Sweden (SWE), using online subjects, recruited through Qualtrics. In summary, we used a 4 x 2 between-subject design, with four sets of experiments in two countries, to draw meaningful comparisons between TUR and SWE.

Experiment 1 is comprised of a simple rule following task, the abstract task. We measure the compliance rate (or violation) in a LOW IDV and a HIGH IDV country and see if similarities or differences exist between these two culturally distinct countries. This serves as a benchmark to see if people from a specific cultural orientation follow rules more efficiently than others.

Experiment 2 aimed to elicit people's empirical and normative beliefs of rule violation in TUR and SWE. In other words, here, the objective was to check if people from a LOW IDV and HIGH IDV a) hold accurate beliefs of the incidence of rule violation around them; and b) display a similar or dissimilar normative standard of compliance and violation outcomes. We used the direct elicitation method to extract people's beliefs of the frequency of rule violations from their country. For the normative belief elicitation, we used the Krupka-Weber (2013) norm elicitation methods in the context of an abstract rule following task. The details of both these measures are discussed below.

Experiment 3 was aimed at looking into people's attitudes of conformity, i.e., if people changed their preference to follow or violate a rule, conditional on the actions and opinions of others from their countries. We ran two sets of treatments, each in TUR and SWE. One of the treatments aimed to investigate people's conditional preference to violate a rule based on varying degrees of rule violation around them. This served as our empirical information

manipulation treatment. In the other treatment, we investigated people's preference to follow or violate a rule based on the varying degree of social disapproval of violation around them. This served as our normative information manipulation treatment. Both treatment manipulations were carried out in TUR and SWE, using a variant of the strategy method, adapted from Gächter et al. (2021).

In total, we had four sets of experiments for both SWE and TUR. We had between 100-200 subjects from each country for each set of experiment, and all studies took roughly 7-8 minutes to complete. The subjects were paid in their local currency. To keep the base rate equal, we converted the incentive in terms of the purchasing power based on the Big Mac Index (2018). Based on this, for each Turkish Lira, the equivalent amount was roughly 5 Swedish Krona. We programmed the experiment using LIONESS (Giamattei et al., 2020) and recruited the subjects using a Qualtrics online panel.

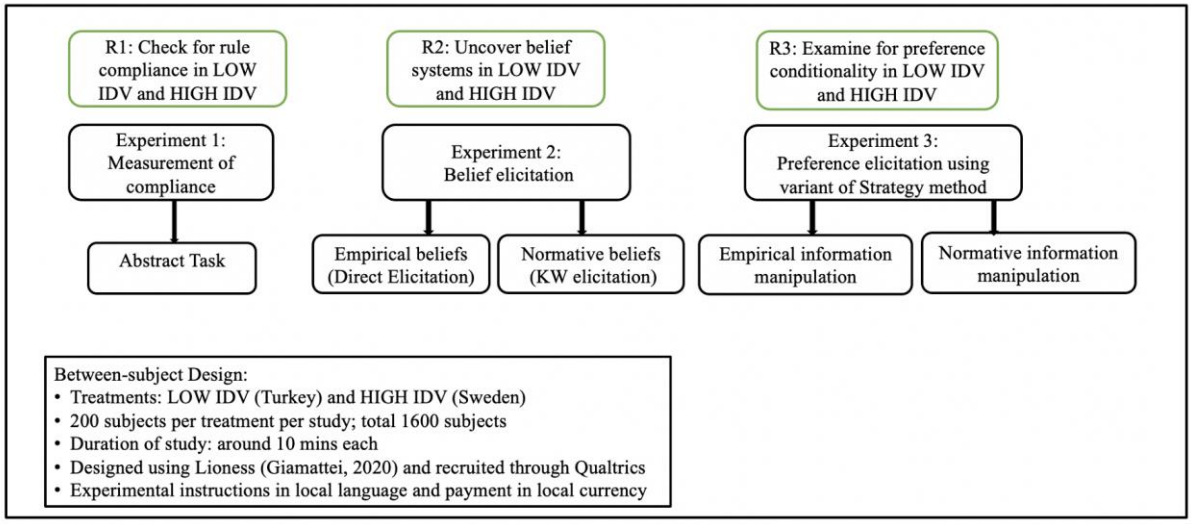


Figure 4.1: Overview of the experimental design

In all three sets of experiments, subjects were first given an online link to enter the Qualtrics interface and asked about their country of birth. This was to ensure that we collected responses only from these two countries. On answering this, they were recruited into the LIONESS

interface, where they completed the experiment. At the end of the experiment, subjects were redirected back into Qualtrics to get registered to be paid the show-up and bonus incentives.

We describe the decision environment of the rule-following task, the abstract task below. This task is used in Experiment 1 to investigate people's rule-following behaviour in SWE and TUR. Following this, we use the same decision environment to elicit people's beliefs in Experiment 2 and elicit conformity attitudes in both informational (empirical and normative) manipulation treatments in Experiment 3. At the beginning of all the three experiments, subjects were given general instructions regarding the duration of the experiment (approximately 10 minutes), the show-up fee, and the opportunity to earn a bonus fee based on decisions in the experiment.

4.4.2 Experiment 1: Abstract rule following task

In Experiment 1, we use the abstract simplified version of the rule-following task, adapted from Gächter et al (2021)[57]. The abstract task is incorporated into the experiment to observe people's intrinsic propensity to follow rules in SWE (HIGH IDV) and TUR (LOW IDV).

The description of the computerized task is as follows: A subject needs to move a grey circle from the left side of the screen to the right side (see Figure 4.2 below). At the beginning of the task, the subject has 20 endowment points. As soon as s/he clicks on the move button, the grey circle the subject controls moves into a grey area in the middle of the screen and stops there. At the same time as the first click, the endowment starts reducing - with each passing second, the endowment decreases by 1 point. There is a cross mark at the top of the grey area (X). The

[57] The behavioural logic of the rule-following task is as follows: a simple task is shown, and then a rule is communicated to the subject – e.g. "The rule is to do X". Here, if anyone follows the rule, s/he loses monetary payoff. Thus, the task is devised to incentivize violation (indirectly), so if anyone follows the rule, despite the incentive to violate, it suggests that they have an internalized motivation to follow a rule. Although a traffic light task was initially developed by Kimbrough and Vostroknutov (2016), Gächter et al. (2021) suggested modifying the task into an abstract setup to avoid potential framing effects.

rule (which is mentioned in the instructions) is to "wait in the grey area until the cross disappears". Once the circle reaches the grey area, the subject can make a choice between two options – either wait in the grey area until the X disappears and follow the rule or move before the X disappears and violate the rule. The cross disappears after 12 seconds. So, if some subject clicks on the move button after 12 seconds, then s/he has followed the rule. Any click before 12 seconds is considered a rule violation. On the 2nd click, the circle moves from the grey area into the black area situated on the right side of the screen, and this is the end of the task.

Following the rule in this task is costly. The longer one waits in the grey area, the more endowment points the subject loses. The payoff in this task is determined by the number of endowment points remaining at the end of the task, multiplied by some positive factor. For example, if someone had completed the task with 16 points remaining, the monetary payoff is 16*0.05= 0.80. Thus, the fewer points one has, the lower is the monetary payoff. Before subjects decide to wait or move in the task, they are reminded of the instructions and the rule that applies in the task (to wait until the cross disappears).

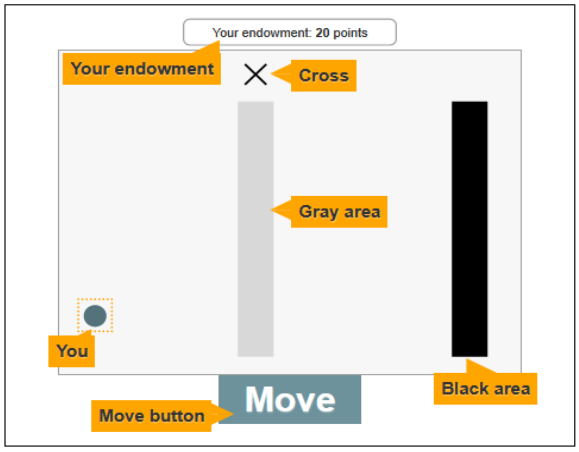


Figure 4.2: Screenshot of the abstract task. On the first click on the Move button, the grey circle moves from the left pane to the centre and stops there. At the same time, the timer is activated and with each passing second, the endowment decreases by 1 point. On the second click, the grey circle moves from the centre into the black area, and this is the end of the task.

After reading the instructions for the task, subjects had to answer a few comprehension questions. This was to make sure that all subjects who did participate in the task had indeed understood the mechanism and payoffs of the task. Subjects were allowed to attempt the questions multiple times, i.e., no subject was eliminated for incorrect attempts. Moreover, the quiz was also not timed, so subjects could take as much time as was necessary to fill in their responses. They were also allowed to go back to the previous screen to review the instructions of the task. The control questions were as follows:

i) How many additional points will you earn if you move after the cross has disappeared and you finish the task in 13 seconds (answer: 7 points)
ii) How many additional points will you earn if you move before the cross has disappeared and you finish the task in 4 seconds (answer: 16 points)

If the subjects provided an incorrect response, the software said "incorrect answer" in the translated text, so jumping ahead to complete the task without clearing these questions was impossible.

After clearing the control questions, subjects took part in the rule following task. Upon completing the task, they completed a short questionnaire stating their age, gender, monthly income, and level of education[58]. They were then shown the earnings page. Depending on the points remaining at the end of the task, subjects were paid an additional bonus, where each point was equivalent to 0.5 Lira or 2.75 Krona.

[58] These same questions were asked in all the studies to maintain uniformity in the experimental instructions.

4.4.3 Experiment 2: Belief Elicitation

The next step was to investigate the beliefs involving rule violation in the LOW IDV (TUR) and HIGH IDV (SWE), i.e., what people expected others (from their country) to do in a situation involving violation, and the underlying normative standard of possible actions from the decision environment of the abstract rule following task. More specifically, in Experiment 2, we aimed to elicit a) the empirical societal belief of rule violation in both countries and b) the normative societal belief of both a rule violation and compliance. We employed the direct elicitation method (for empirical beliefs) and the KW norm-elicitation (2013) for normative beliefs.

Subjects were told that 100 other subjects from their respective country had taken part in a simple computerized decision-making task. This decision task was the abstract task, and subjects were familiarised by showing them the instructions and illustrations of the task[59].

Subjects were then told that they had to evaluate the possible actions of these 100 subjects based on the social appropriateness of the action. By "socially appropriate", we stated what "most people from their country would agree to be the right thing to do". Subjects had to provide their normative judgements for the two possible actions from the decision environment of the abstract task, i.e., follow the rule or break the rule. For the violation scenario, they were given the following information: they had to imagine someone from their country who had previously participated in this task. He or she decided to break the rule and moved before the cross disappeared, completing the task in 5 seconds, and earning 15 points. Subjects in this task now had to judge how socially appropriate the action was. The possible responses available

[59] Subjects were also shown the results of the control questions from Experiment 1 to help them comprehend the rules of the task. More specifically, they were shown that if someone moved the circle figure after the cross disappeared and completed the task in 13 seconds, they would get 7 points. Likewise, if they moved the circle before the cross disappeared and completed the task in 4 seconds, they would get 16 points.

were "very socially appropriate", "socially appropriate", "socially inappropriate", and "very socially inappropriate". For the compliance scenario, they were given the following information: they again had to imagine another participant from their country who had also participated in the task. He or she decided to follow the rule and moved after the cross disappeared, completing the task in 9 seconds, and earning 11 points. Subjects again had to judge this action as "very socially appropriate", "socially appropriate", "socially inappropriate", and "very socially inappropriate".

On each screen, they were reminded that they would be paid for their response if their answer matched the modal response provided by most other subjects from their country. In other words, the KW belief-elicitation task was incentivized such that subjects should be motivated to reveal a socially coordinated focal response if it existed. For example, if someone had rated the action as socially inappropriate, and most people from their country had provided the same response of socially inappropriate, then the subject would be paid an additional bonus of 20 points, where each point was either 0.5 Turkish Lira or 2.75 Swedish Krona. Otherwise, they would earn 0 additional points. Hence, subjects were incentivized to provide the normative judgement of compliance and violation based on their perception of a shared social approval for compliance or disapproval for violation.

After providing their normative responses for the two possible scenarios with violation and compliance outcomes, subjects then had to report their empirical beliefs regarding rule violation in the abstract task. More specifically, they were reminded that 100 other subjects had previously taken part in the computerised abstract task. Subjects were asked to state their belief regarding how many they thought out of 100 subjects moved before the cross disappeared in the task and broke the rule. It was mentioned that if their answer were accurate

with an error of at most +/- 3 percentage points, they would be paid 20 additional points at the end of the study.

After both belief elicitation tasks were over, subjects filled in a brief questionnaire. They were then shown the earnings page, which was also the last screen of the experiment. We mentioned that one of the two norm-elicitation questions would be chosen at random, and each subjects' response would be matched with the most common rating. If the response matched, then subjects would be given 20 additional points. If the rating did not match, subjects would get 0 additional points. Moreover, we also stated that we would compare their empirical belief (regarding how many out of 100 they thought followed the rule in the abstract task) with the actual fraction of subjects who followed the rule in the task. If their response were within +/- 3 percentage points, they would be paid 20 additional points on top of what they were paid for the norm-elicitation if their response matched.

4.4.4 Experiment 3a: Conditional preferences with empirical information

The aim of Experiment 3 was to investigate subjects' conformity attitudes concerning increasing violation rates (empirical information manipulation) and increasing rates of disapproval (normative information manipulation) in a 2 x 2 between-subject setup with subjects from TUR and SWE. We first explain the empirical information treatment manipulation in Experiment 3(a) and then the normative information treatment manipulation in Experiment 3(b).

Our decision environment was the abstract task, where the two possible choices or strategies were to follow or to break the rule. Subjects were told that they would have to elicit their preference between following or breaking a rule in four different scenarios of increasing

information of rule violation. The four outcomes differed in the fraction of the previous participants who broke the rule and moved before the cross disappeared. More specifically, the four possible scenarios of the percentage of people breaking the rule and moving before the cross disappeared were i) 0-25%; ii) 26-50%; iii) 51-75%; and iv) 76-100%. The content of the scenarios involved a rule-violation outcome, and they only varied in the degree of rule violation. Each subject thus was required to reveal a preference for either rule compliance or violation in the four hypothetical scenarios of increasing rule violation. This was adapted from Gächter et al. (2021). This provides us with four compliance or violation preference responses per subject.

The potential points to be earned from the elicited preference to follow or violate a rule was based on the modal compliance and modal violation from Experiment 1 in the abstract task. In Experiment 1, most subjects who followed the rule completed the task in 15 seconds and earned 5 points; most subjects who violated the rule completed the task in 4 seconds and earned 16 points. We used these modal responses to compute the earnings for this task. If subjects chose to follow the rule, their potential earning would be 5 points, and if they chose to violate a rule, their earning would be 16 points.

While subjects provided four responses for the four possible scenarios of increasing violation, only one of the four scenarios was chosen for payment consideration. This selection of scenario for bonus payment was based on the actual rate of violation from Experiment 1 from the respective countries. In other words, we checked the actual percentage of previous participants who moved before the cross disappeared in study 1 from TUR and SWE. We then used that

bracket or decision scenario and checked the elicited preference of the subjects to pay the additional points.[60]

Subjects were provided with the illustration and working of the abstract task to familiarise them with the decision environment. Prior to the preference elicitation, they were also given the feedback from Experiment 2 regarding the normative evaluation of violation from their fellow citizens who took part in a previous study. We told the subjects that we had run a similar study with a different group of subjects from their country and had asked them how socially appropriate they thought it was to move before the cross disappeared. Based on the results from Experiment 2, for the SWE subjects, we informed them that 72% of SWE subjects from a previous study had said that moving before the cross (X) disappeared was socially inappropriate. Likewise, for the TUR subjects, we informed them that 46% of TUR subjects from a previous study considered a violation as socially inappropriate. We did this because we wanted to hold the normative belief constant. We did not want the normative belief to vary with the four scenarios (of increasing violation) and confound the identification of the scenario effect.

After completing the task, subjects were asked to complete a brief questionnaire. They were then shown the actual violation information from Experiment 1 (45% and 61% for SWE and TUR, respectively), based on which their bonus earnings were computed.

[60] In Experiment 1, the actual violation rate was 45% in SWE and 61% in TUR (details in Results). As a result, scenario ii) (26-50% broke the rule) was selected for SWE, and scenario iii) (51-75% broke the rule) was selected for TUR participants.

4.4.5 Experiment 3b: Conditional preferences with normative information

In this experiment, we incorporated normative information manipulation with an increasing disapproval rate towards violation in both SWE and TUR. This investigates people's preference conditionality regarding the normative influence of (increasing) disapproval.

Just as in Experiment 3(a), subjects were first made familiar with the decision environment of the abstract task. They were to reveal their preference between moving before the cross disappeared and waiting until the cross disappeared in four possible decision scenarios. The four decision scenarios varied in the extent to which previous participants considered breaking the rule socially inappropriate. The four possible scenarios of the percentage of people breaking the rule and moving before the cross disappeared were as follows: i) 0-25%; said that breaking the rule and moving before the cross disappeared is socially inappropriate, ii) 26-50%; iii) 51-75%; and iv) 76-100% . For each scenario, subjects could either choose to wait until the cross disappeared and follow the rule or move before the cross disappeared and violate the rule.

Following the incentive protocol of Experiment 3(a), the computation of potential earnings was based on the modal compliance and violation points of Experiment 1. For choosing to break the rule, subjects could earn 15 points, and for choosing to follow the rule, they could earn 4 points. One out of the four possible scenarios were selected for payment computation at the end of the task. This was based on the result from the norm-elicitation task of Experiment 2. [61]

[61] 46% of the TUR from the previous experiment had perceived violation as socially inappropriate. So, for TUR, scenario B (26-50% said that breaking the rule and moving before the cross disappeared is socially inappropriate) was considered for payment. Likewise, 72% of SWE subjects from the earlier study considered rule-violation as a socially inappropriate action, so scenario C (51-75% said that breaking the rule and moving before the cross disappeared is socially inappropriate) was chosen for payment for SWE subjects.

Before beginning the task, subjects were given empirical feedback on the actual rate of rule violation from their countries. We did this because we wanted to hold constant the empirical belief and not let it interfere with the identification of the scenario effect.[62]

Subjects then elicited their preferences for the four possible scenarios of increasing social disapproval of violation, following which they completed the questionnaire. At the end of the task, they were informed of the result from study 2 regarding the actual rate of social disapproval for violation from their countries, based on which scenarios ii) and iii) were chosen for payment in TUR and SWE, respectively.

4.5 Hypotheses

We now stipulate a few testable hypotheses based on the previous literature (reviewed in Subsection 4.3.6) for Experiments 1, 2, and 3. Based on the literature on institutional quality and its effect on rule violation (e.g., Gächter & Schulz, 2016; Licht et al., 2007) as well as the studies looking into a specific form of rule violation, namely, corruption (e.g., López & Santos, 2014; Serra, 2006), we propose the following hypothesis for the behavioural outcome of following a rule, in the abstract task:

H1: Proportion of compliant decisions $_{SWE}$ > Proportion of compliant decisions $_{TUR}$

Due to the limited moral scope in LOW IDV countries, both in terms of the content (e.g., Enke, 2019), as well as the applicability of these rules and values to smaller in-group members only (e.g., Tabellini, 2008), we conjecture that subjects from TUR have more dispersed empirical

[62] For SWE, subjects were informed that 45% of previous participants from another study had moved before the cross(X) disappeared and broken the rule. Likewise, for TUR, subjects were informed that 61% of previous participants from another study had decided to move before the cross (X) disappeared and broke the rule.

beliefs regarding a universal rule of other's actions. We elicit subjects' beliefs of how many others from their country will violate a rule, and so subjects need to state their beliefs of "strangers" from their county. This represents a significantly large reference group, and the accuracy of beliefs regarding others' actions may be influenced by the behavioural patterns of the group members of the smaller local reference groups that subjects might belong to. Additionally, norm enforcement is regulated more through social groups than formal institutions. The interplay of these factors may lead to uncertainty regarding the actions of out-group members or strangers in a LOW IDV. As a result, we expect the following:

H2: Accuracy of empirical belief of violation $_{SWE}$ > Accuracy of empirical belief of violation $_{TUR}$

Based on the evidence of cross-cultural experiments and other studies on cooperation, altruism, and fairness (e.g., Gächter et al., 2010; Henrich et al., 2006; Kimbrough & Vostroknutov, 2016), people everywhere display a shared agreement regarding the social appropriateness of rules of fairness, cooperation, reciprocity which are considered to be universal rules. This leads us to the following hypothesis regarding the normative standard of compliance:

H3: Similar normative perception of rule compliance in both SWE and TUR

The prevalence of a higher incidence of rule violation (e.g., Gächter & Schulz, 2016; Licht et al., 2007) leads to the acceptability of such violations over time (Cameron et al., 2009) in LOW IDV societies. Additionally, the dominance of local norms aimed at maintaining group harmony, such as obedience, respect for the more powerful members, may contradict and weaken the universal norm of rule-following. Moreover, formal institutions have a weak role

in enforcing these norms through punishment mechanisms. In LOW IDV societies, disapproval is shown towards in-group members only, and there is less concern towards violations outside the group. On the other hand, due to the relevance of such norms in HIGH IDV societies, enforced through formal institutions (e.g., Alesina & Giuliano, 2015; Licht et al., 2007), we can speculate a stronger normative evaluation of violation in such countries. Based on these studies, we formalise the following:

H4: Low or no collective consensus of social inappropriateness of violation in TUR

H5: Harsher normative evaluation of violation in SWE than TUR

Based on the evidence that a) people put more weight on negative information, such as bad behaviour of others (e.g., Croson & Shang, 2008; Gino et al., 2009; Oliveira et al., 2014) to serve as a guiding factor in one's one behaviour; and b) group size matters in conformity attitude (e.g., Bond, 2005; Egebark & Ekström, 2017; Sasaki, 2019), we propose the following:

H6: People from both TUR and SWE display preference to violate functional on an increasing rate of others' violation

Studies show higher conformity in LOW IDV countries (e.g., Bond & Smith, 1996; Cialdini et al., 1999; Mesoudi et al., 2015). Additionally, from the related literature regarding rule violations (e.g., Gächter & Schulz, 2016; Hugh-Jones, 2016; Licht et al., 2007; López & Santos, 2014), as well as our hypothesis from Experiment 1, we can expect H7:

H7: Conditional preference for violation $_{TUR}$ > Conditional preference for violation $_{SWE}$

Due to more violation than compliance in a LOW IDV, both based on past literature (Gächter & Schulz, 2016) and our result from Experiment 1, where TUR individuals showed a low compliance rate of 39% compared to 54% compliance in SWE, we can expect the following from the normative manipulation treatment:

H8: Conditional preference for compliance $_{TUR}$ < Conditional preference for compliance $_{SWE}$

Additionally, based on inference from previous literature of (Tabellini, 2008) that normative messaging in LOW IDV countries may be inferred and learnt through the more valued and trusted in-group members, and disapproval coming from peers (strangers) may not necessarily induce conformity towards a certain action, leads us to the following:

H9: No effect on subjects' preference to comply in TUR, as a function of others (strangers) increasing disapproval towards violation

4.6 Results

The overall attrition rate was 10-20% across all the experiments in both countries. We had 192, 202, 241 and 226 subjects for Experiment 1 (abstract task), Experiment 2 (belief elicitation) and Experiments 3(a): empirical manipulation and 3(b): normative manipulation from SWE. Likewise for TUR, we had 220, 273, 224 and 231 subjects from the four experiments[63]. The average age was between 31-33 years. There was a slightly higher proportion of male

[63] We had aimed for around 800 subjects (200 per treatment) per country. The figures mentioned are after attrition and other recruitment issues were accounted for.

respondents from the Turkish pool compared to the Swedish respondents[64]. We compared the proportion of respondents with a bachelor's degree from both the countries. This decision was based on the fact that most respondents were in their 30s and were thus expected to have completed an undergraduate degree. 96-99% of the respondents were born and brought up in their country of origin. As expected, the Swedish subjects reflected higher monthly income compared to the Turkish counterparts. All these demographic information may be indicative of the country-specific characteristics of the panels used by Qualtrics, shown in Table 4.2.

	Treatment	Exp 1	Exp 2	Exp 3a	Exp 3b
Average age	Sweden	32	33	31	31
	Turkey	31	32	31	32
Male %	Sweden	34%	30%	49%	46%
	Turkey	34%	42%	56%	63%
Bachelor's degree +	Sweden	58%	43%	14%	17%
	Turkey	88%	26%	22%	24%
Country at birth	Sweden	98%	98%	98%	97%
	Turkey	96%	96%	98%	99%
Monthly Income ($USD)	Sweden	$4189	$5545	$21552	$9390
	Turkey	$2864	$1536	$ 1174	$824

Table 4.2: Demographics of all subjects. We had asked for the monthly income in terms of Krona/Lira. Here we report them after converting the stated responses in US dollars, to make meaningful comparisons between the two countries' income levels.

We report the results from Experiments 1, 2 and 3 in the following section. Experiment 1 compares the behavioural outcome in our rule following task. Results from Experiment 2 compare people's empirical belief predictions (of violations) with actual violation outcome from Experiment 1. It also checks for subjects' normative beliefs of compliance and violation

[64] For all the regressions, gender is never a significant predictor.

outcomes. Results from Experiment 3 reflect if subjects display preference conditionality with empirical and normative beliefs. For all studies, we make comparisons between a HIGH IDV, namely SWE, and a LOW IDV, namely TUR, to see if similarities or differences exist between the two countries with regards to these elements of the norm's framework.

4.6.1 Abstract task

To check for compliance, we converted the responses as follows: We created a dummy variable for compliance with responses > 12 seconds, getting a value of 1 and responses <12 seconds, getting a value of 0. This is because 12 seconds was the benchmark, when the cross disappeared, and the subject could move the circle figure as per the stated rule. Hence, all responses were converted into a binary response of comply (1) or violate (0) in the abstract task. Based on this, out of 220 subjects, only 86 TUR subjects followed the rule, or waited more than 12 seconds to complete the task. In other words, 39.1% Turkish subjects waited in the grey area until the cross (X) disappeared. By contrast, 107 out of 192 (55.2%) SWE subjects followed the rule. We performed a proportion test and found a statistically significant difference in the compliance outcomes between the subject pools at 1% (Z =-3.0622; p = 0.0022), with higher rule compliance in SWE

The faster subjects completed the task, the more points they earned.[65] Since the cross disappeared only after 12 seconds, and the rule was to wait in the grey area until the cross disappeared, waiting, and following the rule would reduce the points earned in the task. However, if subjects would move the circle figure as soon as possible, without considering the

[65] Each endowment point was equal to 0.5 Turkish Lira or 2.75 Swedish Krona. Based on this, 20 points were either 10 Turkish Lira or 55 Swedish Krona, which had similar purchasing capacity in their respective countries. Based on our data, the average bonus earnings were 23.80 Krona from SWE and 4.91 Turkish Lira from TUR. In both cases, the minimum earning was 0 when some subjects took 20 seconds to complete the task. However, the maximum bonus earnings in SWE were 44 Krona, while in TUR, it was 8.5 Lira. This was equivalent to completing the task in 4 seconds with 16 points remaining in both countries.

rule, they would complete the task much faster and earn more points. So, points earned and following the rule were inversely associated in this task. Based on the result that a bigger proportion of SWE subjects followed the rule compared to the TUR subjects, we can expect the SWE subjects to earn less points. We ran another Mann-Whitney test and confirmed that SWE subjects earned less than TUR subjects and this difference in distributions in points earned was statistically significant at 5%: Rank sum (Z) = 2.281; $p = 0.0226$. Both these results are depicted in Figure 4.3.

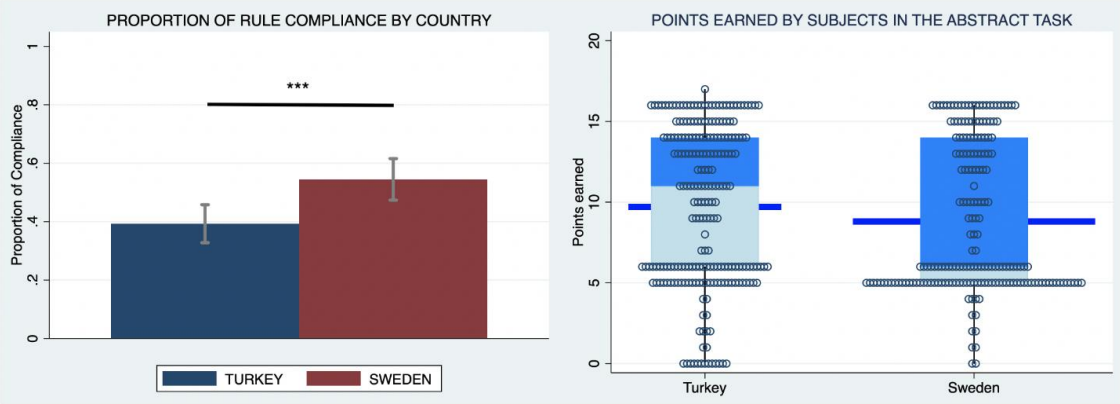


Figure 4.3: Results from the abstract task. (L): Compliance rate in Turkey is 39% and in Sweden is 55% and the difference is statistically significant: Results supported by proportion test; error bars represent the 95% CI. (R): Each circle represents a subject's response. The average points earned in SWE (8.7), and TUR (9.7) is reflected by the blue line. The dark and the light blue shade in the boxplot are demarcated by the median; these values for SWE and TUR are 6 and 11 points respectively.

Since the outcome variable in this study is a binary decision, between following a rule and violating a rule, we use a non-linear model to investigate if cultural orientation of individualism, i.e., HIGH, or LOW IDV has any effect on people's propensity to follow a rule. This is reported in Table 4.3 below. In Col (1) we report the log odds coefficient, which is positive and significant. We report the odds ratio in Col (2): the odds of complying with a rule are almost 82.3% higher for SWE subjects than TUR subjects, and this is statistically significant at 1%. In Col (3), we report coefficients from a probit model, which shows a similar positive significant effect on compliance outcome from a HIGH IDV. We then report average marginal effect (AME) in Col (4) from the probit model. This represents the average change in

probability when we compare responses from TUR and SWE. The probability of following the rule increases by 14.7 percentage points in SWE, compared to TUR.

VARIABLE: Follow Rule	(1) Logit	(2) Odd Ratio	(3) Probit	(4) AME
SWE (HIGH IDV)	0.603***	1.823***	0.378***	0.147***
	(0.221)	(0.403)	(0.138)	(0.053)
Constant	-1.295	0.229	-0.917	
	(0.842)	(0.195)	(0.530)	
Other controls	Yes	Yes	Yes	Yes
Observations	408	408	408	408

Table 4.3: Logit and Probit model from abstract task. This was to check if IDV has any effect on people's decision to follow the rule in the abstract task. Robust standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1. We control for other socio-demographic characteristics such as age, gender, income (converted to $USD), and education level. None of the control variables are significant.

To summarise, based on our results from the first study, we find the following:

R1: Rule compliance is 14.7% higher in SWE than TUR.

4.6.2 Belief Elicitation

With empirical belief elicitation, we had asked the following "how many subjects out of 100 would break the rule (from their country)". We found that the average expected empirical belief was 48% in TUR and 45% in SWE[66]. This showed that both countries depicted similar average beliefs of violation. In fact, there was no statistically significant difference in the average elicited empirical beliefs (Rank sum Z = 1.165; p = 0.2442) between the two countries.

However, SWE subjects were able to predict the actual rate of rule violation in the abstract task from Experiment 1 more accurately than their TUR counterparts. In Experiment 1, the rate of violation was 61% for TUR and 45% for SWE. Using a one sample t-test, we compared these

[66] We calculated the average elicited empirical belief based on the total number of responses. For SWE, this was based on 202 responses and in TUR, this was based on 273 responses.

averages with the elicited average empirical beliefs of violation from Experiment 2: there was a significant difference between the actual violation rate and the average elicited empirical belief in TUR at 1% (Mean: 48.5; SD = 29.7; t = -6.3913; p = 0.0000). On the other hand, there was no difference between the two averages for SWE, (Mean: 45.23; SD = 24.15; t = 0.1399; p = 0.3889), suggesting that SWE subjects had more accurate beliefs about the incidence of rule-violation in their country. These differences are also illustrated in Figure 4.4.

Additionally, we also looked into the variation in the stated empirical beliefs between countries, as is demonstrated in the right panel in Figure 4.4, using a histogram plot. Higher variation reflects more heterogeneity in the stated beliefs. We find higher variation in the stated beliefs in TUR compared to SWE, and this is also verified using a variance ratio test SD (TUR): = 29.7, SD (SWE) = 24.14; f = 1.5133; p = 0.0020[67]

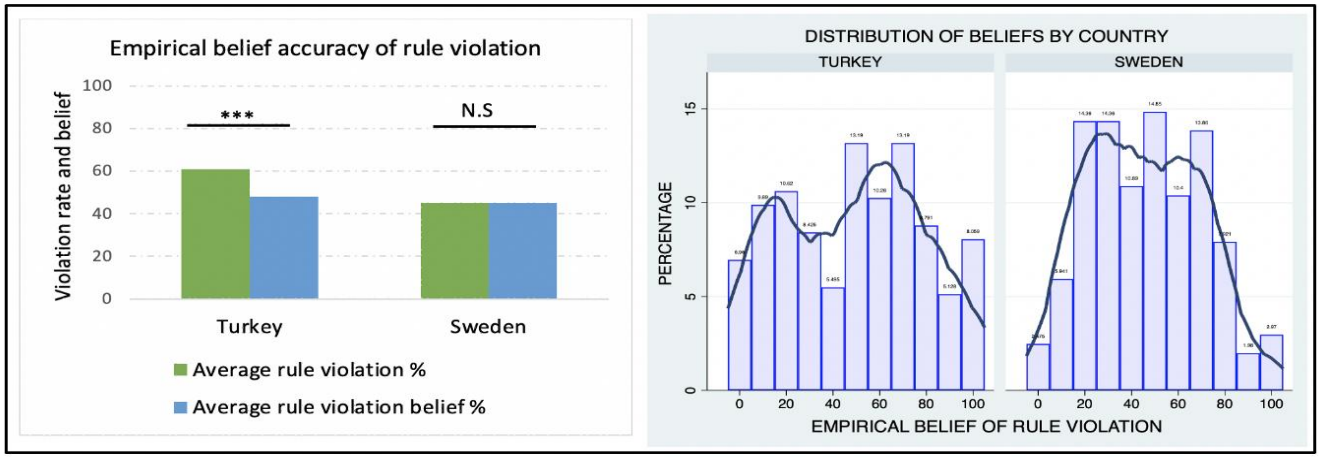


Figure 4.4: Accuracy and distribution of empirical beliefs. (L): Accuracy of empirical beliefs by country. The blue bar represents the rule violation rate from Experiment 1 and the green bar shows the average belief (rate) of rule-violation. In TUR, subjects underestimated the actual level of rule violation whereas in SWE, the prediction was exactly the same as the actual rate of violation from Experiment 1. (R): Distribution of beliefs within countries. TUR reflects slightly higher variability in elicited empirical beliefs than SWE; results supported by variance ratio test. The blue line is based on kernel density plot. It is steeper in SWE, reflecting a lower variation in the stated beliefs. Since we had asked subjects to state how many out of 100, they thought would violate the rule, the range for x axis is from 0-100.

[67] This test compared the standard deviation of the 2 groups of distributions. If the sd(s) are the same, then the null is verified where the ratio is sd(TUR)/sd(SWE)=1. In the alternate hypotheses, the ratio is not equal to 1.

Based on these results, we can conclude the following:

R2: Subjects elicited (more) accurate empirical beliefs of violation in SWE than TUR.

Next, we discuss the normative beliefs in SWE and TUR. Based on the decision environment of the abstract task, we had asked subjects to evaluate the two possible actions of following the rule or breaking the rule as socially appropriate or socially inappropriate behaviour. By socially appropriate, we defined what most people from their country would agree is the right thing to do.

Following the KW literature, we assign the following values to the possible responses in the KW task: -1 to "very socially inappropriate", -0.33 to "socially inappropriate", +0.33 to "socially appropriate" and +1 to "very socially appropriate". In SWE, the average rating value was 0.45, with the modal response being "socially appropriate" for the action of following the rule, while for TUR the average was 0.55 and the action was seen as "very socially appropriate" by majority subjects. In other words, compliance with a rule, was perceived as a normatively appropriate action in both countries. On the other hand, for violation outcome, the average rating was -0.17 with the modal response being "socially inappropriate" in SWE, while for TUR, the average rating was 0.07, with no clear majority response. A breakdown of all the responses is given in Table 4.4 below.

We wanted to investigate if a shared consensus, or a majority opinion, existed regarding the social appropriateness of compliance or social inappropriateness of violation within the countries. For this we created a dummy of social appropriateness. The responses of "very socially appropriate" and "socially appropriate" were coded 1, while the responses of "socially inappropriate" and "very socially inappropriate" were coded as 0. We ran a one sample test of proportion to check if a statistically significant difference existed between the responses of

"appropriate" and "inappropriate" evaluations. Under this test, if there is no statistically significant difference in the proportion of responses indicating between appropriate and inappropriate, the null proportion should be 0.5. However, if there is a higher responses of "appropriateness" rating over the "inappropriateness" rating, then the proportion should not be 0.5, rather it should be a value > 0.5.

Case: Follow rule	Mean	+1	+0.33	-0.33	-1
SWE	0.45	37.1%	46.5%	14.4%	1.9%
TUR	0.55	50.2%	36.3%	10.6%	2.9%
Case: Violate rule					
SWE	-0.17	5.9%	20.8%	65.8%	8.4%
TUR	0.07	27.5%	26%	26.7%	19.8%

Table 4.4: Normative evaluations for compliance and violation by country. For the case of rule-compliance, the modal response was "socially appropriate" in SWE, and "very socially appropriate" for TUR. Majority of the subjects in both countries considered the action as appropriate behaviour. On the other hand, for violation, while most SWE subjects deemed the action as "socially inappropriate", TUR subjects elicited no such modal response.

In the case of compliance, we find evidence of a statistically significant difference between the proportion of socially "appropriate" and "inappropriate" responses for both TUR and SWE, indicating a higher share of responses indicating the action as "socially appropriate" (Z_{TUR} = 12.04; p = 0.0000; Z_{SWE} = 9.5689; p = 0.0000). In other words, in both these countries, majority responses indicated the action of following the rule as a (very) socially appropriate or appropriate action. There was a clear shared consensus regarding the normative standard of this action.

In the case of violation, there is a statistically significant difference in the proportion of responses indicating "appropriateness" and "inappropriateness" in SWE only: Z_{SWE} = -6.6138; p = 0.0000. In fact, there is a higher proportion of responses indicating the action of breaking

the rule as socially inappropriate. However, we find no evidence of a statistically significant difference in the proportion of responses (of appropriateness and inappropriateness) in TUR (Z_{TUR} = 1.1499; p = 0.2502). In other words, we find evidence for no clear consensus within TUR subjects regarding the normative standard of violation. On average, half the responses considered violation as appropriate, while the other half of the respondents considered it inappropriate.

Additionally, we also compared the elicited KW distributions between SWE and TUR for both the actions. For the action of compliance or following the rule, we find a statistically significant difference between the two countries at 5%: Rank Sum Z = 2.532; p = 0.0114. For the action of violation or breaking a rule, we find a statistically significant difference at 1%: Rank Sum Z = 4.001; p = 0.0001. The normative evaluations for both compliance and violation outcomes are illustrated in Figure 4.5 below. Thus, for both violation and compliance with a rule, there is evidence of significant difference in the distribution of responses, between the two countries.

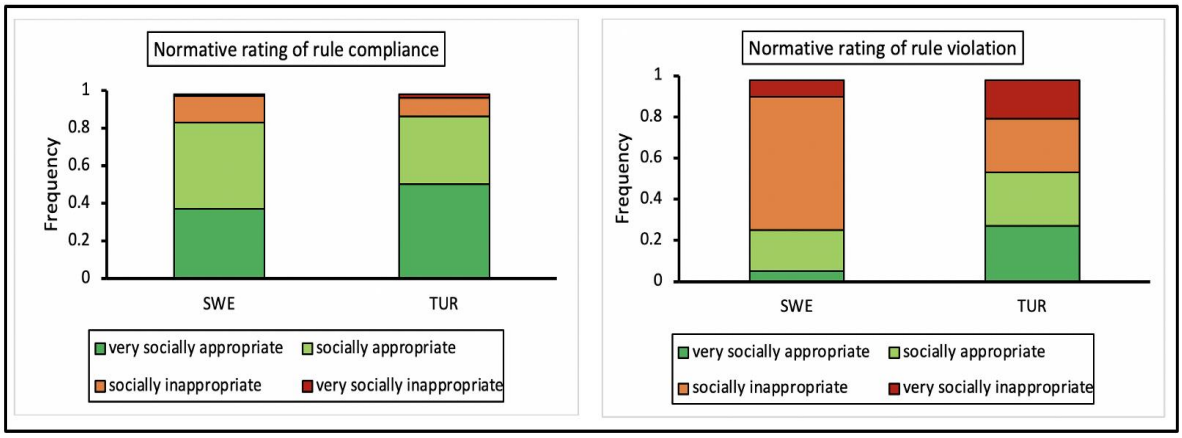


Figure 4.5: Norm elicitations by country. (L) In both, the majority agree that following the rule is the socially right thing to do. There was also a significant difference in the distribution of these evaluations between the two countries at 5%, results supported by the Mann Whitney test. (R): Norm elicitation for rule violation in the 2 countries. In SWE, a majority of the respondents consider the action as socially inappropriate, thus the normative message of this action seems clear. On the other hand, in TUR, there is no modal response in the elicited evaluations in TUR, reflecting the absence of a shared consensus regarding this action's normative standard. There is significant difference in the distribution of these evaluations at 1% between the 2 countries: results supported by Mann Whitney test.

Next, we investigate if subjects' normative evaluations towards compliance and violation varied between the two countries. We ran OLS regression model as shown in Table 4.5 below. In Col (1) we investigate the effect of HIGH IDV (SWE dummy variable), on normative evaluation of violation. We find that subjects provided a harsher average rating of -0.238 in SWE than in TUR. In other words, subjects from SWE, perceived rule violation much more strongly (negatively) compared to subjects from TUR. In Col (2), we report the effect of HIGH IDV (SWE) of subjects' normative evaluations of compliance. We find that subjects from SWE rate the action of following a rule, harshly by a lower rating of 0.150 compared to TUR. This reflects the fact the majority of TUR subjects considered rule compliance as "very socially appropriate", while for SWE, the majority opinion was that compliance is "socially appropriate" (as is reflected in the left panel in Figure 4.5).

VARIABLE: Average Norm rating	(1) Violate rule	(2) Follow rule
SWE	-0.238***	-0.150***
	(0.0606)	(0.0493)
Constant	-0.0274	0.706***
	(0.238)	(0.190)
Controls	Yes	Yes
Observations	475	475
R-squared	0.048	0.023

Table 4.5: OLS regression for KW elicitation of violation. We control for age, gender, income, and education levels in both countries. Robust standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1. We control for other socio-demographic characteristics such as age, gender, income (converted to $USD), and education level.

Based on the above-mentioned evidence, we can formulate the following results:

R3: There is evidence of a shared normative standard of rule compliance in both SWE and TUR.

R4: There is no consensus of social inappropriateness of rule violation in TUR.

R5: Subjects perceive rule-violations more harshly in SWE.

4.6.3 Conditional Preference elicitation

The aim of this experiment was to investigate subjects' conformity attitudes with respect to increasing violation rates (empirical information manipulation) and increasing rates of disapproval (normative information manipulation) using the incentive-compatible method of preference elicitation proposed by Gächter et al. (2021), where subjects had to state their preference to either follow or violate a rule in scenarios of increasing majority size of violation and disapproval towards violation.

First, we investigate the results from our treatment manipulation with increasing empirical information of violation. Since this preference elicitation was based on the variant of the strategy method, we have four responses per subject. The four scenarios of increasing empirical information of rule violation were: a) 0-25% broke the rule; b) 26-50% broke the rule; c) 51-75% broke the rule and d) 76-100% broke the rule. For each scenario, subject had to state a preference between choosing to follow a rule or break a rule.

We first summarize the responses from both the countries in Table 4.6. Since the preference was based on a binary decision between compliance and violation, we tabulated the proportion of responses that elicited a preference to break a rule, conditional on increasing empirical information of violation. We find that for both SWE and TUR, on average, there was an increasing proportion of responses preferring to violate, with increasing empirical information of violation. We show the average elicited preferences within the countries in Figure 4.6 below.

SWEDEN	Proportion of breaking the rule	95% CI	
0-25% broke the rule	0.32	0.26	0.38
26-50% broke the rule	0.36	0.30	0.42
51-75% broke the rule	0.39	0.33	0.45
76-100% broke the rule	0.40	0.34	0.46
TURKEY			
0-25% broke the rule	0.42	0.35	0.48
26-50% broke the rule	0.39	0.33	0.46
51-75% broke the rule	0.51	0.44	0.57
76-100% broke the rule	0.55	0.49	0.62

Table 4.6: Proportion of violation preference by country. We show the proportion of subjects that elicited the preference to violate (and conform to the empirical information) in both SWE and TUR, along with the 95 % confidence intervals. In both, the highest proportion of preference to violation was in the HIGH majority (76-100%) of violation.

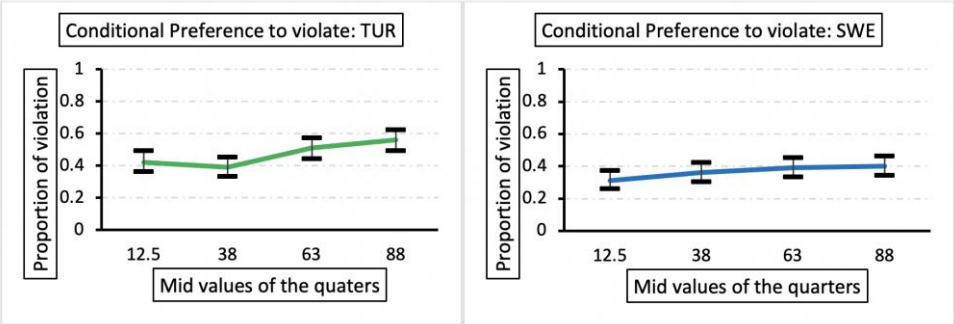


Figure 4.6: Average preference to violate as a function of increasing violation within the countries, along with 95% CI. For both the countries, there is a small degree of preference conditionality, with slightly more conditionality in TUR.

We compared the proportion of subjects' preferences to violate a rule, between SWE and TUR to investigate if cross-cultural differences existed in such elicited preferences. It should be noted that we renamed the four scenarios of increasing empirical information, in the following manner: 0-25% broke the rule as LOW, 26-50% broke the rule as MID LOW, 51-75% broke the rule as MID HIGH and 76-100% broke the rule as HIGH. We found statistically significant

differences between the 2 countries, with higher elicited violation preferences in TUR in LOW at 5% (Rank sum Z = -2.237; p = 0.034); MID HIGH at 5% (Rank sum Z = -2.578; p = 0.0198) and HIGH at 1% (Rank sum Z = -3.401; p = 0.0028). There was no differences between the distributions of responses for MID LOW between the 2 countries (Rank sum Z = -0.713; p = 0.476).

To investigate the causal effect of increasing rate of violation on preference to violate, within the countries and the effect of IDV, we regressed the model on 2 main predictors namely % of others' violation and TUR. The results of the regression are given in Table 4.7 below. In Col (1), we look into the effect of % of others violation on subjects' preference to violate a rule in TUR[68]. There is an increased odds of violating a rule by 0.8% with 1 percent increase in violation rate after controlling for age, gender, education, and income levels. In Col (2), we look into the subjects' decisions to violate in SWE. Again, we find evidence for increased odds of violating a rule by 0.4% with a 1% increase in violation rate and this is statistically significant at 5%.

In Col (3), with pooled data, we again find significant evidence of increasing odds of a preference of rule violation, by 0.6% with a 1% increase in other's violation rate and this is significant at 1%. However, with TUR subjects, there is an increased odds of 50.4% preferring to violate based on empirical information of violation compared to their SWE counterparts.

[68] For this, we converted the independent variable of others' violation around as follows: for each of the bracket of violation, we considered the midpoint of the bracket. For instance, for 0-25% broke the rule, we took the mid-value of 12.5, for 26-50% broke the rule, we used 38, for 51-75%, we used 63 and for 76-100%, we used 88. This helped us to modify the independent variable of increasing violation rate, which was based on 4 scenarios that differed in the fraction of violation, in terms of a continuous variable with 4 values of 12.5, 38, 63 and 88.

This provides evidence of higher degree of conformity (with rule violation) in a LOW IDV TUR compared to a HIGH IDV country such as SWE[69].

	(1)	(2)	(3)
VARIABLE: **Violate Rule**	Odds Ratio TUR	Odds Ratio SWE	Odds Ratio Pooled
% Of others violation	1.008***	1.004**	1.006***
	(.002)	(0.002)	(.0016)
TUR			1.504***
			(0.191)
Constant	0.386	1.620	0.644
	(0.276)	(1.292)	(0.336)
Other Controls	Yes	Yes	Yes
Observations	900	964	1864

Table 4. 7: Logit Regression with empirical manipulation. Robust standard errors in parentheses, clustered at subject level to account for intra-subject correlations in responses; *** p<0.01, ** p<0.05, * p<0.1. We control for other socio-demographic characteristics such as age, gender, income (converted to $USD), and education level. We have 225 subject level clusters for TUR, 241 for SWE and 466 for pooled data.

Based on these results, we can summarize the following results:

R6: Subjects from both SWE and TUR display slight conditional violation with increasing information of others' violation.

R7: Subjects from TUR display slightly higher preference conditionality with empirical information of violation.

We now look into the normative manipulation treatment. In this case, we again reviewed people's conditional preference, but with respect to increasing rate of disapproval towards violation. In other words, the aim of this treatment was to check if people, both from a LOW IDV and HIGH IDV, displayed conformity attitude as a function of increasing normative disapproval, towards the action of breaking a rule.

[69] We report results from OLS regressions in Table A4.2 and Table A4.3 (with dummies for the fractions of information) in Appendix.

We first summarize the proportion of compliance preference, across the four scenarios of increasing disapproval in both SWE and TUR in Table 4.8. It was interesting to find that irrespective of the information of increasing disapproval, the overall proportion of compliance preference in both countries remained within 0.50- 0.60. In other words, 50-60% of the subjects out of a total pool of 226 and 231, in SWE and TUR, chose to comply with the rule across the four scenarios of increasing normative information of disapproval. We illustrate within-countries' elicited average preferences in Figure 4.7.[70]

SWEDEN	Proportion of elicited preference to comply with the rule	95% CI	
0-25% disapproved	0.58	0.36	0.48
26-50% disapproved	0.57	0.36	0.49
51-75% disapproved	0.58	0.35	0.48
76-100% disapproved	0.60	0.34	0.47
TURKEY			
0-25% disapproved	0.52	0.42	0.55
26-50% disapproved	0.58	0.36	0.49
51-75% disapproved	0.56	0.37	0.50
76-100% disapproved	0.51	0.43	0.56

Table 4.8: Proportion of compliance preference by country. We show the proportion of subjects that elicited the preference to comply (and conform to the normative information) in both SWE and TUR, along with the 95 % confidence intervals.

For all the four cases, we found similar rates of compliance preference between the two countries. As in the case of empirical information treatment, we call the 0-25% bracket as LOW, 26-50% as MID LOW, 51-75% as MID HIGH, and 76-100% as HIGH (disapproval).

[70] Between countries' differences from both information manipulation is shown in Figure A4.2 in Appendix.

We compared the distribution of responses between SWE and TUR, for the 4 scenarios and found no statistically significant differences in the distribution of the responses: LOW: Rank Sum Z = 1.337; p = 0.3628; MID LOW: Rank Sum Z = -0.054; p = 0.9573; MID HIGH: Rank Sum Z = 0.417; p = 0.902; HIGH: Rank Sum Z = 1.903; p = 0.2284.

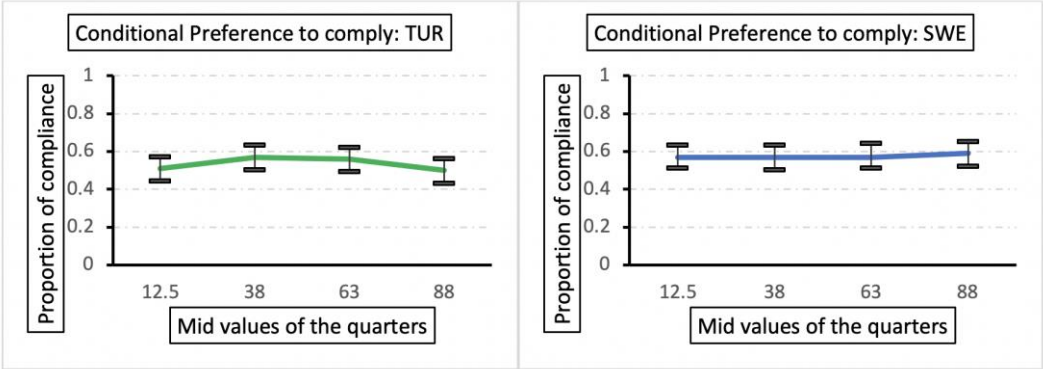


Figure 4.7: Average (elicited) preference to comply within countries along with 95% CI.

From the statistical tests and illustrations above, we gauge that there may not exist preference conditionality with normative information manipulation in the two countries. To verify this, we again ran a logit model to see if subjects' preference to comply with a rule was conditional on increasing rate of disapproval and IDV. As before, we converted the independent factor of other's disapproval, into a continuous predictor variable, by taking the midpoint of the bracket of increasing fraction of disapproval.

The dependent variable here is a dummy for compliance preference. From the Col (1) and Col (2) in Table 4.9, we find that rate of others disapproval towards violation has no significant effect on subject's preference to follow a rule in TUR and SWE respectively. Additionally, we pool the data and look into the effect of IDV on subject's elicited preference to follow a rule in

Col (3). Neither predictor have any significant effect in influencing people's preference to follow a rule[71].

	(1)	(2)	(3)
VARIABLE: Follow Rule	Odds Ratio TUR	Odds Ratio SWE	Odds Ratio Pooled
% Of others disapproval	0.984	1.02	1.004
	(0.051)	(0.049)	(0.036)
TUR			0.843
			(0.115)
Constant	1.209	1.294	1.364
	(0.051)	(0.192) *	(0.174)**
Other Controls	Yes	Yes	Yes
Observations	928	904	1832

Table 4.9: Logit Regression with normative manipulation. Robust standard errors in parentheses, clustered at subject level to account for intra-subject correlations in responses; *** p<0.01, ** p<0.05, * p<0.1. We control for other socio-demographic characteristics such as age, gender, income (converted to $USD), and education level. We have 232 subject level clusters for TUR, 226 for SWE and 458 for pooled data.

This confirms the following result:

R8: Subjects do not display conditional preferences with rule following based on social information of disapproval (of violation).

4.7 Discussion and Conclusion

This study comprises of a three-part experiment investigating the effect of cultural orientation on rule compliance. More specifically, we look into the effect of the cultural orientation of individualism (IDV) on three main aspects of the Bicchieri framework of social norms (2005; 2016). First, we investigate if the level of rule compliance varies between LOW and HIGH IDV countries, using a simple computerized task, the abstract task. Second, we elicit people's beliefs in these countries, both empirical beliefs regarding what others around them do and

[71] We report results from OLS regressions in Table A4.4 and Table A4.5 (with dummies for the fractions of information) in Appendix.

normative beliefs that reflect what they think others around them approve of being done. Third, we elicit both the normative evaluation of rule-compliance and rule-violation since these are the two possible outcomes in the decision environment of the abstract task. We use the Krupka-Weber (2013) norm elicitation protocol to elicit these beliefs. Finally, we test if people from a LOW and HIGH IDV prefer to follow a rule based on the actions and opinions (disapproval) of others around them. For this, we elicit people's preference to follow a rule as a function of a) increasing rate of rule violation and b) increasing rate of disapproval towards violation. We incorporate the strategy method in the context of the abstract task, where people could choose between a binary preference of rule compliance or rule violation based on an increasing fraction of individuals that violated the rule or displayed disapproval towards violation.

Our findings can be listed as follows: firstly, we found a statistically significant difference in the distribution of compliance outcome from the abstract task between SWE and TUR. A higher proportion of SWE subjects followed the rule (55%) compared to their TUR counterparts, where the compliance rate was only 39%. We also found a causal effect of higher rule-following in a cultural orientation of HIGH IDV. This aligns with the previous evidence investigating the relationship between cultural orientation and rule following (Gächter & Schulz, 2016; Licht et al., 2007).

Next, we found some interesting results regarding the elicited empirical and normative beliefs. We found comparable average empirical belief of violation at 45% in SWE and 48% in TUR. This reflected that although these HIGH and LOW IDV countries differed in the actual rate of violation (from Experiment 1), they did not differ in their expectation of the rate of violation around them. Additionally, we found that SWE subjects displayed a higher accuracy with the actual behavioural outcome of violation. In other words, results from the SWE showed that, on

average, they expected 45% of their people to violate a rule. In comparison with the behavioural data from Experiment 1, we found that this was within three percentage points of the actual violation rate. This led us to infer that people in a HIGH IDV country are better informed about what most others do around them. This may be because they have macro-level social structures involving looser social groups who play a minor role in belief projection. Additionally, rules are regulated through formal institutions, whereby the same rules apply to all members equally. As a result of such transparency, people from such countries might find it easier to form beliefs regarding others' actions.

On the other hand, TUR subjects predicted a violation rate of 48%, which was much below the actual violation of 60%. In LOW IDV structures, with many more social groups with stronger ties, people might be more aware of localized groups' behaviours. Still, they may not be necessarily well-informed of what strangers within the country do. Additionally, these countries have inefficient formal institutions that have less role to play in regulating rules. Therefore, the under-estimated beliefs of violation in TUR, may be based on the projected beliefs of the subjects' reference group's behavioural patterns.

Thirdly, we compared the normative evaluations regarding compliance and violations between these countries. We found a clear consensus regarding the social appropriateness of following a rule in both countries. However, regarding the normative judgement of violation, while SWE displayed a clear modal response of social inappropriateness, there was no clear consensus within the subjects in TUR. Additionally, subjects from SWE provided a harsher judgement for a violation than TUR subjects. All these observations show that the normative message for rule compliance is similar irrespective of the degree of individualism. But countries differ in their perceptions of the normative content of violations. In other words, LOW and HIGH IDV

don't affect the norms underlying compliance, but they do shape the norm of violation differently.

This may be due to ineffective formal institutions with weak norm-enforcement via punishment mechanisms. As a result, there is a high incidence of rule violation, which erodes the normative strength of the inappropriateness of violations over time. Moreover, informal reference groups are more salient in shaping such normative messages in such societies. These groups have unequal power dynamics between group members, and the group leaders regulate rule-violations pertaining to the members only (Alesina & Giuliano, 2015). The subjective nature of such norm-enforcement measures makes the norm harder to internalize.

Fourth, we investigated if individuals' preference to follow a rule is conditional on an increasing rate of rule violation and increasing rate of disapproval towards violation. We found that subjects from both SWE and TUR preferred to violate a rule as a function of increasing violations. For subjects from both countries, increasing violation rates may indicate congruent normative consideration of violation (Bicchieri et al., 2020). In TUR, subjects are exposed to a higher incidence of rule violation (from Experiment 1; PRV score of 0.044) than SWE. Additionally, they have weak norms against violation (from Experiment 2). These two factors combined may help explain a higher preference to violate a rule conditional on the information of increasing violation in TUR than SWE.

On the other hand, we found that an increasing rate of normative message through disapproval towards violation does not influence people's preference to comply with a rule in both SWE and TUR. This may be explained on the grounds that normative messaging through disapproval may not necessarily indicate high rule compliance (Bicchieri et al., 2020) and so doesn't act as

a sufficient condition for conformity. Additionally, in a LOW IDV, people's actions are motivated by their reference groups, so disapproval from known group members might exert pressure to conform and comply (Alesina & Giuliano, 2015). On the other hand, disapproval from strangers (even from the same country) might not exert enough pressure to follow a rule.

In summary, rule violation is indeed higher in a LOW IDV country such as TUR. On investigating the reasons behind this low compliance, using the Bicchieri framework, we find that although the normative consideration of following a rule is the same everywhere, societies differ in the extent of establishing the normative message against violation. This may be due to the role of informal reference groups having a more significant role in norm-enforcement than formal institutions. In addition, relational dynamics between ingroup and out-group members leads to uncertainty in punishment against rule violators. Ingroup members may be punished or shown disapproval but not necessarily out-groups. This leads to vagueness in the normative message against violation. As a result, a norm against violation is not well established. Additionally, the incidence of high violations, motivated by such weak normative beliefs, could also explain the stronger conformity attitude in individuals towards violation.

4.8 Appendix

4.8.1 Supplementary Analysis and Figures

	Dimension	Indicator	Turkey	Sweden
1.a	Hofstede Cultural dimension	Individualism	37	71
		Power distance	66	31
1.b	Inglehart-Welzel dimensions	Traditional vs Secular	-1.2	1.86
		Survival vs. self- expression	-0.33	2.35
2.a	Institutional Quality	PRV (2016)	0.044	-2.764
2.b		WGI (2019)	25-50th Per	90-100 Per
		Voice and accountability	1.59	-0.81
		Political stability	1.05	-1.34
		Government effectiveness	1.83	0.05
		Regulatory control	1.80	-0.01
		Rule of law	1.91	-0.28
		Control of corruption	2.12	-0.29
3.	Global Preference	Patience	-0.04	1.07
		Positive reciprocity	-0.44	-0.02
		Negative reciprocity	0.51	-0.09
		Trust	0.12	0.30
		Altruism	-0.28	-0.17
		Risk	0.02	0.05
4.	Economic Prosperity	GDP per capita (2020)	$9402	$58,100

Table A4.1: Cultural, institutional, and economic indicators of Sweden and Turkey

Other than the Hofstede dimensions, we also mention the Inglehart-Welzel dimensions based on the WVS data analysis. This database posits that cross-cultural variation is based on two primary dimensions in the world. These dimensions are i) traditional values versus secular-rational values and ii) survival versus self-expression values. Societies that prioritise traditional values give importance to religion and social ties and have a nationalistic view. They reject notions of divorce, abortion, and suicide and score high in this dimension. On the other hand,

societies with opposite preferences towards the situations mentioned above score less on this dimension and sit on the opposite end of the spectrum. These are considered secular-rational countries. From Table A4.1 we find that TUR is more traditional, while SWE reflects a secular-rational approach. The other dimension is survival versus self-expression. Countries that value economic and physical security have a survival-based orientation. In contrast, countries focused on factors other than these, such as environment, openness to foreigners etc., are the ones that have a self-expression-based orientation. We find that TUR has a survival-based approach while SWE has a self-expression based cultural orientation.

Additionally, we also summarise the differences between the countries based on the quality of institutions that have a role to play in norm enforcement. For this, we chose to show the prevalence of rule violation index (PRV scores), developed by Gächter and Schulz (2016). They created the prevalence of rule violation (PRV) index using a principal component analysis by taking three primary forms of violation, namely political fraud, tax evasion and corruption, from 23 countries, and the PRV index ranged from -3.1 (low PRV) to 2.0 (high PRV). We find that TUR has a higher (positive) score than SWE, which shows that TUR society has a higher incidence of rule-violation in daily life than SWE. This is also reflected through the six World Governance Indicators, whereby SWE has a higher score for all indicators, showing a more elevated and efficient system of formal institutions. As a result, SWE has a higher rank (90-100th percentile) compared to TUR (25-50th percentile). The percentile rank (0-100) indicates rank of country among all countries in the world in the Gächter and Schulz (2016) dataset; 0 corresponds to lowest rank and 100 corresponds to highest rank.

Next, we also list the differences between the social, risk and time preferences from the data collected by Falk et al. (2018) through the Global Preference Survey. SWE is more patient, has

higher positive reciprocity, trust, and risk preference than TUR. On the other hand, TUR has a higher preference for negative reciprocity (depicting outcomes such as revenge) than SWE.

We also show how IDV cultural orientation is related to other indicators of interest and the distance between SWE, and TUR based on these relationships in Figure A4.1.

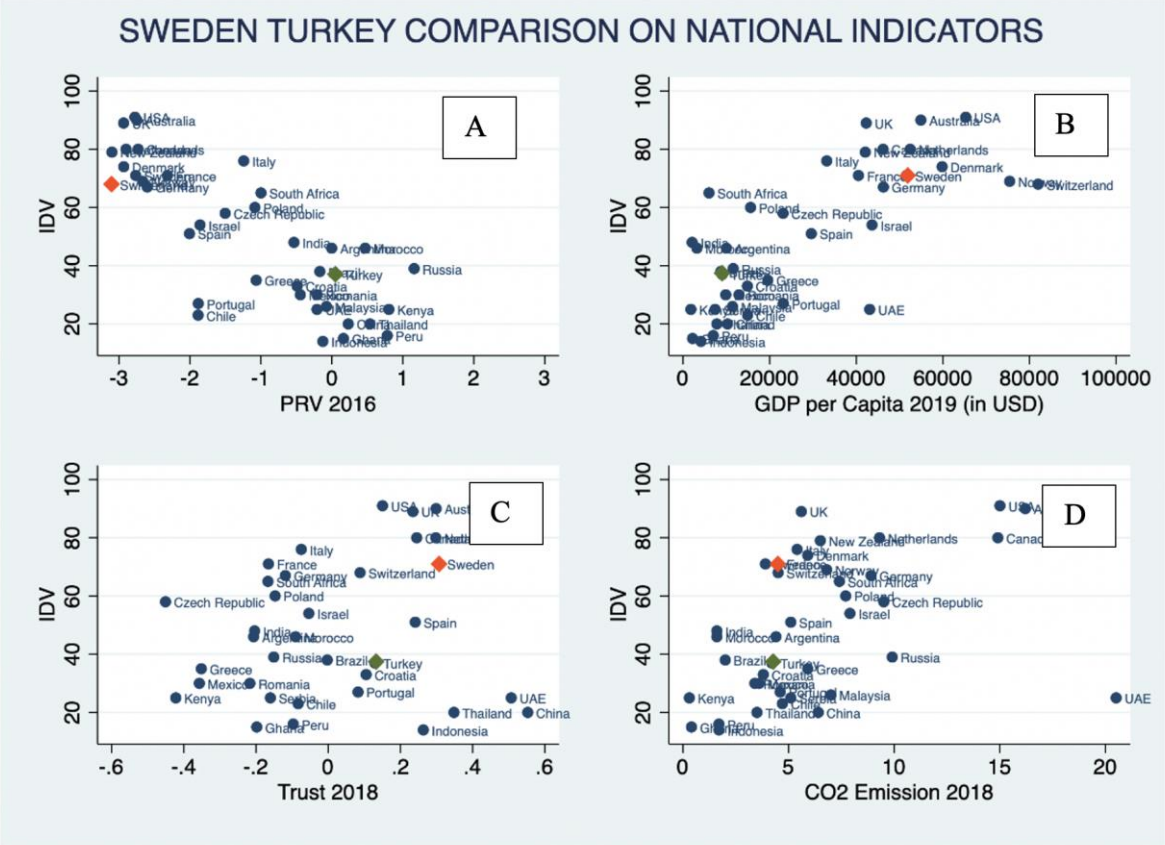


Figure A4.1: Comparative view between SWE and TUR on various indicators such as social capital (A, C); income (B) and environmental indicator (D). We have IDV on the vertical axis, which is why SWE is higher than TUR in all figures. From a small sample of countries, we find a negative association between IDV and PRV and a positive relation between IDV and GDP per capita, trust and CO2 emission.

Based on a small sample of countries, we first find a negative relationship between IDV and PRV: more individualistic countries also have a low prevalence of rule violation. Based on a sample of 40 countries[72], we found a strong negative correlation between IDV and PRV r (39) = -0.8112; $p < 0.0000$.

[72] This is taken from the dataset of Gächter and Schulz (2016) study.

SWE has a HIGH IDV and low PRV score, while TUR has a LOW IDV and high PRV score. On the other hand, we find a positive relationship between GDP per capita and IDV: Positive association, and statistically significant; r (39) = 0.7509; p<0.0000. Likewise, SWE lies much higher than TUR in this graph. We also see a positive association between trust and IDV level: SWE, being high on IDV, has higher trust than TUR (reflected in the Global Preference Survey). Finally, we also examine the association between environment emissions and IDV and find a positive relationship again: these countries, having a higher growth outcome, also produce more CO2 emissions than countries low on the IDV scale: Positive association, and statistically significant; r (39) = 0.4432; p = 0.0047.

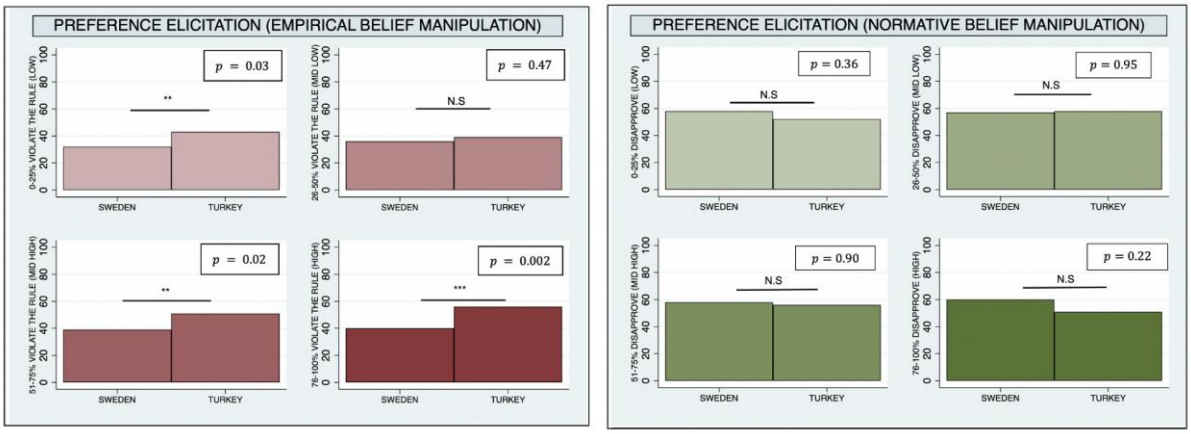


Figure A4.2: Differences in elicited preference between countries. (L): For all the four cases of empirical belief manipulation (of increasing violation rate), we find higher rate of elicited preference to violate in TUR than SWE; the differences are statistically significant for LOW, MID LOW and HIGH; results supported by Mann-Whitney corrected for multiple comparisons using FDR method. (R): For all the 4 cases of normative belief manipulation (of increasing disapproval), we find slightly higher rate of elicited preference to comply in SWE than TUR; but the differences are not statistically significant any of the cases; results supported by Mann-Whitney corrected for multiple comparisons using FDR method.

VARIABLE: Violate Rule	(1) TUR	(2) SWE	(3) Pooled
% Of others violation	0.00205***	0.00112**	0.00156***
	(0.000553)	(0.000489)	(0.000368)
TUR			0.0979***
			(0.0304)
Constant	0.269	0.600***	0.395***
	(0.174)	(0.183)	(0.125)
Other controls	Yes	Yes	Yes
Observations	900	964	1,864
R-squared	0.028	0.019	0.024

Table A4.2: LPM regression with empirical information manipulation. Robust standard errors in parentheses, clustered at subject level to account for intra-subject correlations in responses; *** p<0.01, ** p<0.05, * p<0.1. We control for other socio-demographic characteristics such as age, gender, income (converted to $USD), and education level. We have 225 subject level clusters for TUR, 241 for SWE and 466 for pooled data.

VARIABLE: Violate Rule	(1) TUR	(2) SWE	(3) Pooled
MID LOW	-0.0311	0.0456	0.00858
	(0.0392)	(0.0379)	(0.0272)
MID HIGH	0.0844**	0.0747**	0.0794***
	(0.0388)	(0.0361)	(0.0264)
HIGH	0.133***	0.0830**	0.107***
	(0.0424)	(0.0388)	(0.0286)
TUR			0.0979***
			(0.0304)
Constant	0.325*	0.608***	0.425***
	(0.174)	(0.184)	(0.125)
Other controls	Yes	Yes	Yes
Observations	900	964	1,864
R-squared	0.032	0.020	0.025

Table A4.3: LPM regression 2 with empirical information. Robust standard errors in parentheses, clustered at subject level to account for intra-subject correlations in responses; *** p<0.01, ** p<0.05, * p<0.1. We control for other socio-demographic characteristics such as age, gender, income (converted to $USD), and education level. We have 225 subject level clusters for TUR, 241 for SWE and 466 for pooled data.

VARIABLE: Follow Rule	(1) TUR	(2) SWE	(3) Pooled
% Of others disapproval	-0.000149	0.000276	4.55e-05
	(0.000516)	(0.000467)	(0.000348)
TUR			-0.0419
			(0.0336)
Constant	0.545***	0.565***	0.577***
	(0.0372)	(0.0321)	(0.0286)
Other controls	Yes	Yes	Yes
Observations	928	905	1,832
R-squared	0.000	0.000	0.002

Table A4.4: LPM regression with normative information manipulation. Robust standard errors in parentheses, clustered at subject level to account for intra-subject correlations in responses; *** p<0.01, ** p<0.05, * p<0.1. We control for other socio-demographic characteristics such as age, gender, income (converted to $USD), and education level. We have 232 subject level clusters for TUR, 226 for SWE and 458 for pooled data.

VARIABLE: Follow rule	(1) TUR	(2) SWE	(3) Pooled
MID LOW	0.0603**	-0.00189	0.0284
	(0.0297)	(0.0409)	(0.0252)
MID HIGH	0.0474	0.00696	0.0262
	(0.0342)	(0.0347)	(0.0243)
HIGH	-0.00862	0.0202	0.00437
	(0.0397)	(0.0382)	(0.0275)
TUR			-0.0419
			(0.0336)
Constant	0.513***	0.573***	0.565***
	(0.0329)	(0.0330)	(0.0280)
Other controls	Yes	Yes	Yes
Observations	928	904	1,832
R-squared	0.004	0.000	0.002

Table A4.5: LPM regression 2 with normative information. Robust standard errors in parentheses, clustered at subject level to account for intra-subject correlations in responses; *** p<0.01, ** p<0.05, * p<0.1. We control for other socio-demographic characteristics such as age, gender, income (converted to $USD), and education level. We have 232 subject level clusters for TUR, 226 for SWE and 458 for pooled data.

CHAPTER 5 Conclusion

This thesis explored some pertinent questions related to norm compliance. To be more specific, the main research questions of the thesis are listed as follows: First, we studied the influence of the saliency of normative considerations in a social dilemma problem. Second, we aimed to find an efficient method to elicit preference conditionality or conformity attitudes in a simple rule-following task. Finally, we investigated ways in which social structures inform people's beliefs and conformity attitudes that shape aggregate rule-following behaviour.

In Chapter 2, we were interested in looking at the effect of the saliency of normative considerations on cooperation outcomes in a social dilemma problem. We exogenously manipulated the saliency of a cooperative norm from none to private thoughts to shared beliefs and observed the effects on both short term and long-term beliefs and cooperation outcomes. We also looked into people's post play normative evaluations of cooperation after their local experience in the repeated public good game, with partner matching and belief elicitation.

We obtained a few noteworthy results: first, normative considerations need to be made salient at a group-level, where all members are informed of a shared belief of cooperation. This helps improve contribution at least in the first round of a finitely repeated public good game, through an immediate improved expectation of each other's contribution choices. This result demonstrates that people update their empirical beliefs after norm-saliency, and both information helps make people behave in a norm-congruent manner immediately. However, with heterogeneous preferences, some group members act selfishly and pull down the average group beliefs, which consequently lowers cooperation in the long run. In other words, such saliency of normative standards of cooperation has a transient effect on contribution choices.

Moreover, we found that groups that performed better in terms in average cooperation, irrespective of the initial saliency, also provided a harsher evaluation for no and low contribution choices than groups that performed worse. Moreover, they also provided a more positive normative evaluation for full cooperation. However, all groups irrespective of the degree of saliency agreed on the content of the cooperation norm, which reflected the stability of a norm of cooperation.

In Chapter 3, we tested two methods in preference elicitation in the context of a rule-following task, using Amazon MTurkers. We compared the strategy method (Gächter et al., 2021) with a direct method inspired by Abeler et al. (2019) after evaluating the key advantages and disadvantages of both. People's preferences to conform are conditional on the social expectations of other's actions and opinions (of disapproval) and group size (Gächter et al., 2021). However, there exist limited tools in experimental methods to elicit such preferences in rule-following situations.

For both methods, we elicited subjects' preference to follow or violate a rule across four situations that differed in the size of previous MTurkers' violation outcomes. With the S method, we found that subjects elicited an increase in the preference to violate conditional on an increasing rate of others' violation. This reflected proof of conditionality with empirical information of violation. Moreover, this method was also useful in representing people's heterogeneous preferences with varying degrees of sensitivity with rule-following. We found three important results with regards to heterogeneity in preferences in rule-compliance outcomes. First, the majority of the subjects were conditional rule-followers, i.e., subjects who chose to follow a rule initially, but then shifted to rule violation with higher rates of others' violation. Second, among these conditional followers, subjects reflected varying degree of

sensitivity towards the information of other's violation, with most of them requiring more than 50% of others violating (or a majority others violating) to shift from the socially appropriate to the selfish action. Third, very few of the subjects were either unconditional rule followers or unconditional rule violators, who didn't shift their preferences irrespective of the actions of others.

In the D method, subjects elicited a lowered preference to violate a rule with increasing information of others' violation, a contradiction to the notion of preference conditionality (with empirical information). The method involved graphical representation of varying fraction of rule followers (in green circles) and rule violators (in red circles) prior to the decision-making in the rule-following task. Subjects were also asked to think and provide possible reasons as to why people follow and violate rules. All these stages prior to the actual task, may have made normative messaging of the inappropriateness of violation (or appropriateness of compliance) salient in subjects' minds that may have impacted their preferences.

Between the two methods, we recommend using the S-method for preference elicitation in rule-following. This method is economical and can be used in a wide array of decision-environments. It is extremely useful in testing out proof of concept experiments, as it provides a "clean setup" where one can isolate the role of a social factor (empirical information in this case) from other confounding influences, on subjects' preferences. Additionally, this is beneficial is finding the "tipping point" within individuals with conditional preferences to jump from compliance to violation, based on the proportion of others violating the rule.

In Chapter 4, we investigated how societal background informs and interacts with rule compliance. Norms are group-level behavioural patterns, and thus are based on social learning

of beliefs and values through relevant reference groups. These groups can be informal social groups or efficient formal institutions, enforcing the rules. We tested the Bicchieri (2005; 2016) framework of norms involving social expectations and preferences for conformity in two culturally distinct countries. These were Sweden: a country reflecting an egalitarian and individualistic outlook, and Turkey, a relatively less individualistic country with unequal social structures.

We used a simple abstract rule following task adapted from Gächter et al. (2021) to first evaluate the aggregate level of rule-compliance in these countries. We found that rule-violations were higher in Turkey than in Sweden. We then elicited their social beliefs using direct belief elicitation for empirical beliefs and Krupka and Weber's (2013) method for eliciting normative beliefs. Looking at the effectiveness of the strategy method in preference elicitation in rule-following (from Chapter 3), we used this to elicit people's preference for conformity with empirical information (of others' violation) and normative information (of others disapproval towards violation).

We found a few interesting results: first, people were more aware of others' actions (towards violation) in highly individualistic Sweden. Second, in both countries, there was a clear consensus of the social appropriateness of rule-following, but they differed in their perceptions towards the (in) appropriateness of a violation. While there was a clear majority response of considering rule violation as a socially inappropriate action in Sweden, there was no clear consensus towards the inappropriateness of this action in Turkey. Third, people from both countries demonstrated conformity with increasing (negative) information of others violation, with a slightly stronger preference in Turkey. However, the normative information of

(increasing) others' disapproval towards violation had no effect on people's conformity attitudes from both countries.

This investigation helped in understanding how societal background may inform people's beliefs and individual preferences for conformity through social learning. In low individualism societies, there is more emphasis on group learning through informal institutions involving families and communities. On the other hand, in individualistic countries such as Sweden, there is a stronger dependence of formal institutions that inform people's beliefs through efficient norm-enforcement mechanisms (e.g., Alesina & Giuliano, 2015). This helps them form accurate beliefs of others' rule compliance outcomes, as well as clear strong normative message regarding the social inappropriateness of such an action. On the other hand, such messaging is not well-established in low individualism countries, where there in uncertainty in punishing transgressions. The authoritative group members (or leaders) may punish the rule-violator if he belongs to the group but may stay indifferent towards such actions in the society at large. The interplay of a high prevalence of rule-violation as a result of weak normative standards of behaviour and a stronger reliance on others for decision-making, leads to further erosion of rules in low IDV. This is a first step in understanding what underlying social factors contributes to the systematic variation on rule compliance outcomes across the globe.

Based on the results, from our studies, we now make a few suggestions regarding future research. Following the findings from Chapter 2, we recommend checking norm-saliency using the Krupka Weber (2013) norm elicitation method in the context of a finitely repeated game to see how group-level experience affects norms of conditional cooperation. Next, we suggest using the strategy method for eliciting such conformity attitudes in experiments investigating social norms. The strategy method in the context of the rule-following task can be explored

further with between-subject designs to overcome the potential issues of the experimental demand effects. Moreover, based on the results from Chapter 4, we suggest testing the elements of the Bicchieri (2005; 2016) norm framework in several other countries that offer a wide variation in the degree of their individualism. This would help review and understand the pathways that shape rule-following in societies that differ in the extent of their individualism, an important feature defining social structures that shape belief systems.

We conclude with a few methodological and technical lessons we have learned during this research. We start with the software we used. The experimental designs were implemented using LIONESS (Giamattei et al., 2020). This is an extremely useful software that provides the flexibility of conducting experiments both online and in the laboratory. We recruited subjects from Qualtrics panels in Sweden and Turkey and Amazon MTurkers, from the US. We made some observation during data-collection regarding these online crowdsourcing platforms. While MTurk is useful to collect data from the US, Qualtrics on the other hand has representative samples from many countries. Moreover, the attrition rate in Qualtrics stays very low, which is useful for online experiments, especially the ones involving interactive games. However, we observed that the experimenter's control over data-collection is limited here, as the traffic is regulated through the Qualtrics panel. We also collected online student samples from the University of Nottingham during November - December 2020 due to the pandemic. Subjects were recruited using ORSEE and entered into a "virtual lab" using Microsoft Teams, where they provided their identity to the experimenter, and were then directed into the LIONESS interface to complete the experiment online. Since this was part 2 of the saliency experiments, the protocol was followed from part 1 laboratory experiments. Subjects were paid using PayPal transfers. All these studies were beneficial in expanding ideas of data-collection

with online experimentation, a fast and cost-effective method to carry out experiments with different pools of subjects.

Bibliography

Abeler, J., Nosenzo, D., & Raymond, C. (2019). Preferences for truth-telling. *Econometrica, 87*(4), 1115-1153.

Akerlof, G. A. (1980). A theory of social custom, of which unemployment may be one consequence. . *The Quarterly Journal of Economics, 94*(4), 749-775.

Akerlof, G. A., & Kranton, R. E. (2000). Economics and identity. *The Quarterly Journal of Economics, 115*(3), 715-753.

Alesina, A., & Giuliano, P. (2015). Culture and Institutions. *Journal of Economic Literature, 53*(4), 898-944.

Allcott, H., & Rogers, T. (2014). The Short-Run and Long-Run Effects of Behavioral Interventions: Experimental Evidence from Energy Conservation. *American Economic Review, 104*(10), 3003-3037.

Arechar, A. A., Gächter, S., & Molleman, L. (2018). Conducting interactive experiments online. *Experimental Economics, 21*(1), 99-131.

Balliet, D. (2009). Communication and Cooperation in Social Dilemmas: A Meta-Analytic Review. *Journal of Conflict Resolution, 54*(1), 39-57.

Banerjee, A. V. (1992). A simple model of herd behavior. *The Quarterly Journal of Economics, 107*(3), 797-817.

Banuri, S., & Eckel, C. (2012). *Experiments in culture and corruption: A review*.

Barcelo, H., & Capraro, V. (2015). Group size effect on cooperation in one-shot social dilemmas. *Scientific Reports, 5*(1), 1-8.

Baumeister, R. F., Bratslavsky, E., Finkenauer, C., & Vohs, K. D. (2002). Bad is stronger than good. *Review of General Psychology, 5*(4), 323-370.

Bazerman, M. H., Moore, D. A., Tenbrunsel, A. E., Wade-Benzoni, K. A., & Blount, S. (1999). Explaining how preferences change across joint versus separate evaluation. *Journal of Economic Behavior & Organization, 39*(1), 41-58.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal statistical society: series B (Methodological), 57*(1), 289-300.

Bernheim, B. D. (1994). A theory of conformity. *Journal of Political Economy, 102*(5), 841-877.

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

Bicchieri, C. (2016). *Norms in the wild: How to diagnose, measure, and change social norms*. Oxford University Press.

Bicchieri, C., & Chavez, A. (2010). Behaving as expected: Public information and fairness norms. *Journal of Behavioral Decision Making, 23*(2), 161-178.

Bicchieri, C., & Dimant, E. (2019). Nudging with care: The risks and benefits of social information. *Public Choice*, 1-22.

Bicchieri, C., & Funcke, A. (2018). Norm change trendsetters and social structure. *Social Research: An International Quarterly, 85*(1), 1-21.

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

Bicchieri, C., Dimant, E., & Sonderegger, S. (2020). It's Not A Lie if You Believe the Norm Does Not Apply: Conditional Norm-Following with Strategic Beliefs. *Available at SSRN 3326146*.

Bicchieri, C., Dimant, E., Gächter, S., & Nosenzo, D. (2022). Social proximity and the erosion of norm compliance. *Games and Economic Behavior, 132*, 59-72.

Bikhchandani, S., Hirshleifer, D., & Welch, I. (1998). Learning from the behavior of others: Conformity, fads, and informational cascades. *Journal of Economic Perspectives, 12*(3), 151-170.

Bond, R. (2005). Group Size and Conformity. *Group Processes & Intergroup Relations, 8*(4), 331-354.

Bond, R., & Smith, B. P. (1996). Culture and conformity: A meta-analysis of studies using Asch's (1952b, 1956) line judgment task. *Psychological Bulletin, 119*(1), 111-137.

Bott, K. M., Cappelen, A. W., Sørensen, E. Ø., & Tungodden, B. (2020). You've got mail: A randomized field experiment on tax evasion. *Management Science, 66*(7), 2801-2819.

Brandts, J., & Charness, G. (2011). The strategy versus the direct-response method: a first survey of experimental comparisons. *Experimental Economics, 14*(3), 375-398.

Brosig, J., Weimann, J., & Yang, C.-L. (2003). The hot versus cold effect in a simple bargaining experiment. *Experimental Economics, 6*(1), 75-90.

Burlando, R. M., & Guala, F. (2005). Heterogeneous agents in public goods experiments. *Experimental Economics, 8*(1), 35-54.

Bursztyn, L., Fiorin, S., Gottlieb, D., & Kanz, M. (2019). Moral incentives in credit card debt repayment: evidence from a field experiment. *Journal of Political Economy, 127*(4), 1641-1683.

Cameron, L., Chaudhuri, A., Erkal, N., & Gangadharan, L. (2009). Propensities to engage in and punish corrupt behavior: Experimental evidence from Australia, India, Indonesia and Singapore. *Journal of Public Economics, 93*(7-8), 843-851.

Capraro, V., Jagfeld, G., Klein, R., Mul, M., & Van De Pol, I. (2019). Increasing altruistic and cooperative behaviour with simple moral nudges. *Scientific Reports, 9*(1), 1-11.

Cason, T. N., & Mui, V. L. (1998). Social influence in the sequential dictator game. *Journal of Mathematical Psychology, 42*(2-3), 248-265.

Cialdini, R. B. (1993). The Psychology of Persuasion. *New York*.

Cialdini, R. B., & Goldstein, N. J. (2004). Social influence: compliance and conformity. *Annu Rev Psychol, 55*, 591-621.

Cialdini, R. B., Kallgren, C. A., & Reno, R. R. (1991). A focus theory of normative conduct: A theoretical refinement and reevaluation of the role of norms in human behavior. *Advances in Experimental Social Psychology 24*, 201-234.

Cialdini, R. B., Reno, R. R., & 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), 1015.

Cialdini, R. B., Wosinska, W., Barrett, D. W., Butner, J., & Gornik-Durose, M. (1999). Compliance with a request in two cultures: The differential influence of social proof and commitment/consistency on collectivists and individualists. *Personality and Social Psychology Bulletin, 25*(10), 1242-1253.

Cinyabuguma, M., Page, T., & Putterman, L. (2005). Cooperation under the threat of expulsion in a public goods experiment. *Journal of Public Economics, 89*(8), 1421-1435.

Cohn, A., Maréchal, M. A., Tannenbaum, D., & Zünd, C. L. (2019). Civic honesty around the globe. *Science, 365*(6448), 70-73.

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

Dal Bó, E., & Dal Bó, P. (2014). "Do the right thing:" The effects of moral suasion on cooperation. *Journal of Public Economics, 117*, 28-38.

Dandurand, F., Shultz, T. R., & Onishi, K. H. (2008). Comparing online and lab methods in a problem-solving experiment. *Behaviour Research Methods, 40*(2), 428-434.

Davis, L. S., & Williamson, C. R. (2019). Does individualism promote gender equality? *World Development, 123*, 1-16.

De Oliveira, A. C. M., Croson, R. T. A., & Eckel, C. (2014). One bad apple? Heterogeneity and information in public good provision. *Experimental Economics, 18*(1), 116-135.

Demarque, C., Charalambides, L., Hilton, D. J., & Waroquier, L. (2015). Nudging sustainable consumption: The use of descriptive norms to promote a minority behavior in a realistic online shopping environment. *Journal of Environmental Psychology, 43*, 166-174.

Dimant, E. (2019). Contagion of pro- and anti-social behavior among peers and the role of social proximity. *Journal of Economic Psychology, 73*, 66-88.

Dong, B., Dulleck, U., & Torgler, B. (2012). Conditional corruption. *Journal of Economic Psychology, 33*(3), 609-627.

Duffy, J., & Xie, H. (2016). Group size and cooperation among strangers. *Journal of Economic Behavior & Organization, 126*, 55-74.

Egebark, J., & Ekström, M. (2017). Liking what others "Like": using Facebook to identify determinants of conformity. *Experimental Economics, 21*(4), 793-814.

Ellingsen, T., & Johannesson, M. (2008). Anticipated verbal feedback induces altruistic behavior. *Evolution and Human Behavior, 29*(2), 100-105.

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

Enke, B. (2019). Kinship, Cooperation, and the Evolution of Moral Systems. *The Quarterly Journal of Economics, 134*(2), 953-1019.

Falk, A., & Fischbacher, U. (2002). "Crime" in the lab-detecting social interaction. *European Economic Review, 46*(4-5), 859-869.

Falk, A., Becker, A., Dohmen, T., Enke, B., Huffman, D., & Sunde, U. (2018). Global Evidence on Economic Preferences. *The Quarterly Journal of Economics, 133*(4), 1645-1692.

Fehr, E., & Fischbacher, U. (2004). Social norms and human cooperation. *Trends in Cognitive Sciences, 8*(4), 185-190.

Fehr, E., & Schurtenberger, I. (2018). Normative foundations of human cooperation. *Nature Human Behaviour, 2*(7), 458-468.

Fischbacher, U., & Gächter, S. (2010). Social Preferences, Beliefs, and the Dynamics of Free Riding in Public Goods Experiments. *American Economic Review, 100*(1), 541-556.

Fischbacher, U., Gächter, S., & Quercia, S. (2012). The behavioral validity of the strategy method in public good experiments. *Journal of Economic Psychology, 33*(4), 897-913.

Fornara, F., Pattitoni, P., Mura, M., & Strazzera, E. (2016). Predicting intention to improve household energy efficiency: The role of value-belief-norm theory, normative and informational influence, and specific attitude. *Journal of Environmental Psychology, 45*, 1-10.

Gächter, S., Herrmann, B., & Thoni, C. (2010). Culture and cooperation. *Philosophical Transactions of the Royal Society, B: Biological Sciences, 365*(1553), 2651-2661.

Gächter, S., Molleman, L., & Nosenzo, D. (2021). *Why people follow rules*, Working Paper.

Gächter, S., & Schulz, F. J. (2016). Intrinsic honesty and the prevalence of rule violations across societies. *Nature, 531*(7595), 496-499.

Giamattei, M., Yahosseini, K. S., Gächter, S., & Molleman, L. (2020). LIONESS Lab: a free web-based platform for conducting interactive experiments online. *Journal of the Economic Science Association, 6*(1), 95-111.

Gino, F., Ayal, S., & Ariely, D. (2009). Contagion and differentiation in unethical behavior: The effect of one bad apple on the barrel. *Psychological Science, 20*(3), 393-398.

Goldstein, N. J., Cialdini, R. B., & Griskevicius, V. (2008). A room with a viewpoint: Using social norms to motivate environmental conservation in hotels. *Journal of Consumer Research, 35*(3), 472-482.

Gorodnichenko, Y., & Roland, G. (2011). Which Dimensions of Culture Matter for Long-Run Growth. *American Economic Review, 101*(3), 492-498.

Gorodnichenko, Y., & Roland, G. (2021). Culture, institutions and democratization. *Public Choice, 187*(1), 1-31.

Greiner, B. (2015). Subject pool recruitment procedures: organizing experiments with ORSEE. *Journal of the Economic Science Association, 1*(1), 114-125.

Guiso, L., Paola, S., & Luigi, Z. (2006). Does culture affect economic outcomes? *Journal of Economic Perspectives, 20*(2), 23-48.

Güth, W., Huck, S., & Müller, W. (2001). The Relevance of Equal Splits in Ultimatum Games. *Games and Economic Behavior, 37*(1), 161-169.

Güth, W., Huck, S., & Rapoport, A. (1998). The limitations of the positional order effect: Can it support silent threats and non-equilibrium behavior? *Journal of Economic Behavior & Organization, 34*(2), 313-325.

Hallsworth, M., List, J. A., Metcalfe, R. D., & Vlaev, I. (2017). The behavioralist as tax collector: Using natural field experiments to enhance tax compliance. *Journal of Public Economics, 148*, 14-31.

Hamann, K. R., Reese, G., Seewald, D., & Loeschinger, D. C. (2015). Affixing the theory of normative conduct (to your mailbox): Injunctive and descriptive norms as predictors of anti-ads sticker use. *Journal of Environmental Psychology, 44*, 1-9.

Henrich, J., McElreath, R., Barr, A., Ensminger, J., Barrett, C., Bolyanatz, A., Cardenas, J. C., Gurven, M., Gwako, E., & Henrich, N. (2006). Costly punishment across human societies. *Science, 312*(5781), 1767-1770.

Hofstede, G. (1980). Culture and organizations. *International Studies of Management & Organization, 10*(4), 15-41.

Hofstede, G. (2001). *Culture's consequences: Comparing values, behaviors, institutions and organizations across nations.* Sage publications.

Horton, J. J., Rand, D. G., & Zeckhauser, R. J. (2011). The online laboratory: Conducting experiments in a real labour market. *Experimental Economics, 14*(3), 399-425.

Hugh-Jones, D. (2016). Honesty, beliefs about honesty, and economic growth in 15 countries. *Journal of Economic Behavior & Organization, 127*, 99-114.

Inglehart, R., & Welzel, C. (2010). The WVS cultural map of the world. *World Values Survey*.

Innes, R., & Mitra, A. (2013). Is Dishonesty Contagious? *Economic Inquiry, 51*(1), 722-734.

Isler, O., & Gächter, S. (2022). Conforming with peers in honesty and cooperation. *Journal of Economic Behavior & Organization, 195*, 75-86.

Ito, K., Ida, T., & Tanaka, M. (2018). Moral suasion and economic incentives: Field experimental evidence from energy demand. *American Economic Journal: Economic Policy, 10*(1), 240-267.

Keizer, K., Lindenberg, S., & Steg, L. (2008). The spreading of disorder. *Science, 322*(5908), 1681-1685.

Keynes, J. (1930). *A treatise on money*. London, UK: Macmillan.

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

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

Klasing, M. J. (2013). Cultural dimensions, collective values and their importance for institutions. *Journal of Comparative Economics*, 41(2), 447-467.

Kölle, F., & Quercia, S. (2021). The influence of empirical and normative expectations on cooperation. *Journal of Economic Behavior & Organization, 190*, 691-703.

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

Krupka, E., & Weber, R. A. (2009). The focusing and informational effects of norms on pro-social behavior. *Journal of Economic Psychology, 30*(3), 307-320.

Licht, A. N., Goldschmidt, C., & Schwartz, S. H. (2007). Culture rules: The foundations of the rule of law and other norms of governance. *Journal of Comparative Economics, 35*(4), 659-688.

Lindenberg, S., & Steg, L. (2007). Normative, gain and hedonic goal frames guiding environmental behavior. *Journal of Social Issues, 63*(1), 117.

López, J. A. P., & Santos, J. M. S. (2014). Does corruption have social roots? The role of culture and social capital. *Journal of Business Ethics, 122*(4), 697-708.

López-Pérez, R., & Vorsatz, M. (2010). On approval and disapproval: Theory and experiments. *Journal of Economic Psychology, 31*(4), 527-541.

Maier-Rigaud, F. P., Martinsson, P., & Staffiero, G. (2010). Ostracism and the provision of a public good: experimental evidence. *Journal of Economic Behavior & Organization, 73*(3), 387-395.

Masclet, D., Noussair, C., Tucker, S., & Villeval, M. C. (2003). Monetary and nonmonetary punishment in the voluntary contributions mechanism. *American Economic Review, 93*(1), 366-380.

Mesoudi, A., Chang, L., Murray, K., & Lu, H. J. (2015). Higher frequency of social learning in China than in the West shows cultural variation in the dynamics of cultural evolution. *Proceedings of the Royal Society B: Biological Sciences, 282*(1798), 1-7.

Muthukrishna, M., Morgan, T. J. H., & Henrich, J. (2016). The when and who of social learning and conformist transmission. *Evolution and Human Behavior, 37*(1), 10-20.

Nosenzo, D., Quercia, S., & Sefton, M. (2015). Cooperation in small groups: the effect of group size. *Experimental Economics, 18*(1), 4-14.

Noussair, C., & Tucker, S. (2005). Combining monetary and social sanctions to promote cooperation. *Economic Inquiry, 43*(3), 649-660.

Ones, U., & Putterman, L. (2007). The ecology of collective action: A public goods and sanctions experiment with controlled group formation. *Journal of Economic Behavior & Organization, 62*(4), 495-521.

Ostrom, E. (2000). Collective action and the evolution of social norms. *Journal of Economic Perspectives, 14*(3), 137-158.

Peeters, R., & Vorsatz, M. (2013). Immaterial rewards and sanctions in a voluntary contribution experiment. *Economic Inquiry, 51*(2), 1442-1456.

Pellerano, J. A., Price, M. K., Puller, S. L., & Sánchez, G. E. (2017). Do extrinsic incentives undermine social norms? Evidence from a field experiment in energy conservation. *Environmental and Resource Economics, 67*(3), 413-428.

Peysakhovich, A., & Rand, D. G. (2016). Habits of Virtue: Creating Norms of Cooperation and Defection in the Laboratory. *Management Science, 62*(3), 631-647.

Rapoport, A., & Fuller, M. A. (1998). Coordination in noncooperative three-person games under different information structures. *Group Decision and Negotiation, 7*(4), 363-382.

Rege, M., & Telle, K. (2004). The impact of social approval and framing on cooperation in public good situations. *Journal of Public Economics, 88*(7-8), 1625-1644.

Reno, R. R., Cialdini, R. B., & Kallgren, C. A. (1993). The transsituational influence of social norms. *Journal of Personality and Social Psychology, 64(1)*(1), 104.

Roland, G. (2015). Economics and Culture. *Emerging Trends in the Social and Behavioral Sciences: An Interdisciplinary, Searchable, and Linkable Resource*, 1-18.

Roth, A. (1995). Bargaining Experiments. In *In J.Kagel and A. Roth, (eds.), Handbook of Experimental Economics* (pp. 253-348.).

Sasaki, S. (2019). Majority size and conformity behavior in charitable giving: Field evidence from a donation-based crowdfunding platform in Japan. *Journal of Economic Psychology, 70*, 36-51.

Schotter, A., Weigelt, K., & Wilson, C. (1994). A laboratory investigation of multiperson rationality and presentation effects. *Games and Economic Behavior, 6*(3), 445-468.

Schram, A., & Charness, G. (2015). Inducing Social Norms in Laboratory Allocation Choices. *Management Science, 61*(7), 1531-1546.

Schram, A., Di Zheng, J., & Zhuravleva, T. (2022). Corruption: A cross-country comparison of contagion and conformism. *Journal of Economic Behavior & Organization, 193*, 497-518.

Serra, D. (2006). Empirical determinants of corruption: A sensitivity analysis. *Public Choice, 126*(1), 225-256.

Shaffer, L. S. (1983). Toward Pepitone's vision of a normative social psychology: What is a social norm?. *The Journal of Mind and Behavior*, 275-293.

Shank, D. B., Kashima, Y., Peters, K., Li, Y., Robins, G., & Kirley, M. (2019). Norm talk and human cooperation: Can we talk ourselves into cooperation? *Journal of Personality and Social Psychology, 117*(1), 99-123.

Simpson, B., Willer, R., & Harrell, A. (2017). The Enforcement of Moral Boundaries Promotes Cooperation and Prosocial Behavior in Groups. *Scientific Reports*, 7(1), 42844.

Smith, J. R., Louis, W. R., Terry, D. J., Greenaway, K. H., Clarke, M. R., & Cheng, X. (2012). Congruent or conflicted? The impact of injunctive and descriptive norms on environmental intentions. *Journal of Environmental Psychology, 32*(4), 353-361.

Tabellini, G. (2008). Institutions and culture. *Journal of the European Economic Association, 6*(2-3), 255-294.

Tabellini, G. (2010). Culture and institutions: economic development in the regions of Europe. *Journal of the European Economic Association, 8*(4), 677-716.

Thöni, C., & Gächter, S. (2015). Peer effects and social preferences in voluntary cooperation: A theoretical and experimental analysis. *Journal of Economic Psychology, 48*, 72-88.

Treisman, D. (2000). The causes of corruption: a cross-national study. *Journal of Public Economics, 76*(3), 399-457.

Tversky, A., & Kahneman, D. (1974). Judgment under Uncertainty: Heuristics and Biases: Biases in judgments reveal some heuristics of thinking under uncertainty. *Science*, 185(4157), 1124-1131.

Weber, R. A., Camerer, C. F., & Knez, M. (2004). Timing and virtual observability in ultimatum bargaining and "weak link" coordination games. *Experimental Economics, 7*(1), 25-48.

Xiao, E., & Houser, D. (2009). Avoiding the sharp tongue: Anticipated written messages promote fair economic exchange. *Journal of Economic Psychology, 30*(3), 393-404.

Young, H. P. (2015). The Evolution of Social Norms. *Annual Review of Economics, 7*(1), 359-387.

Zizzo, D. J. (2009). Experimenter demand effects in economic experiments. *Experimental Economics, 13*(1), 75-98.

Experimental Instructions

Saliency Instructions

Here, we show the instructions from High condition. For NO, there were only 2 parts: PART 1 was the PGG and PART 2 was the Stage 2 KW task, and each part had a 50 % chance of being selected for payment. For LOW, there were 3 parts just as HIGH, except that those participants here, didn't get any feedback from the Stage 1 KW task before PGG.

Welcome!

Welcome to our decision- making experiment. This experiment consists of 3 parts: PART 1, PART 2 and PART 3.

In each part of the experiment, you will be asked to make some decisions and you can earn money. The amount of money you earn in each part will depend on your decisions and the decisions of other participants in the experiment here today.

During the experiment we will compute your earnings in "points". At the end of the experiment your point earnings will be converted into cash at a rate of **£0.08 per point.**

Only one part of the experiment will be taken into account in determining your final earnings from today's experiment. At the end of the experiment, the computer will randomly select one of the Parts for payment**. Part 1 will be selected with 25% chance, Part 2 will be selected with 50% chance, and Part 3 will be selected with 25% chance**. You can then collect your payment, in cash and in private. It is important that you do not talk to any other participant until the experiment is over.

Please refrain from using mobile phones during the length of the experiment. If you have a question, raise your hand, and the experimenter will come to your desk and answer it. Please note that your identity will not be revealed to anyone during the experiment. Thus, your decisions will remain anonymous.

At the end of the session, you will be asked to fill in a questionnaire. The answers you provide in this questionnaire are completely anonymous. Your responses to the questionnaire will not affect your earnings from the experiment.

In the next page you will receive instructions about PART 1 of the experiment.

General Instructions for Part 1

In this part of the experiment, you will be asked to evaluate different possible choices an individual might make.

Specifically, we will describe you a decision situation and a choice that an individual (we will call this person "Individual A") has already made in that situation, and you will decide whether making that choice is "socially appropriate" and "consistent with moral or proper social behaviour" or "socially inappropriate" and "inconsistent with moral or proper social behaviour".

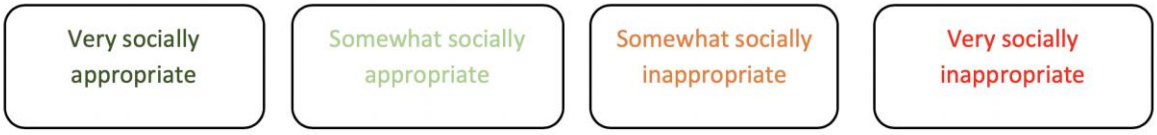
By socially appropriate, we mean behaviour that <u>most people in this room today</u> agree is the "correct" or "ethical thing to do".

Another way to think about what we mean is that, if someone were to make a socially inappropriate choice, then someone observing this behaviour might get angry at the person who made the choice for acting in that manner. In each of your response, we would like you to evaluate what constitutes socially appropriate or inappropriate behaviour.

To give you an idea of how the task in Part 1 works, consider the example on the next page.

Example Situation

Individual A is at a café. Individual A notices that someone has left a wallet at one of the tables. How socially appropriate would it be for Individual A to take the wallet for himself?



If this were the situation we asked you in the experiment, you would indicate the extent to which you believe taking the wallet would be "socially appropriate" and "consistent with moral or proper social behaviour" or "socially inappropriate" and "inconsistent with moral or proper social behaviour". Recall that by socially appropriate, we mean behaviour that most people agree is the "correct" or "ethical thing to do".

For instance, if you thought that taking the wallet was very socially inappropriate, then you would indicate your response by selecting the last box.

If you have any questions at any time, please raise your hand and wait for the experimenter to come to you. Otherwise, click on the next page to continue.

Situation for Part 1

The situation you will be asked to consider in Part 1 of the experiment is the following.

Individual A has been invited to an experiment and placed in a group with 2 other randomly selected participants. Individual A is given an endowment of 5 tokens and has to decide how many of these 5 tokens (between 0 and 5, included) to contribute to a "group account". Any token Individual A does not contribute to the group account will be kept in their "private account". The other members of Individual A's group have to make the same decision. They

also receive 5 tokens each and decide how many to contribute to the group account or how many to keep in their private accounts.

Individual A's earnings from this situation are computed as follows.

For each token Individual A keeps in their private account, Individual A receives 1 point.

Individual A also receives points for the tokens contributed to the group account by Individual A, and the other group members. These tokens are added up and multiplied by 1.5. The resulting number of points is then divided equally between Individual A and the other two group members (irrespective of how much they contributed to the group account).

The points Individual A earns from the situation are equal to the points from their private account + the points from the group account.

The earnings of the other two group members are calculated in the same way.

Your task

In the next screen, we will ask you to evaluate the 6 possible contribution choices that Individual A can make in the decision situation we have described above. For each contribution choice, you will have to indicate whether that choice is socially appropriate or socially inappropriate.

Your earnings from Part 1

Your earnings from Part 1 will depend on whether your evaluation of Individual A's choices matches the evaluation of one other randomly selected participant in this session.

At the end of the experiment, if Part 1 is selected for payment, we will randomly select one of the 6 possible contribution choices by Individual A. We will compare you evaluation of that contribution choice with the evaluation of the other randomly selected participant. If your

evaluation is exactly the same as theirs you will receive 125 points, otherwise you will receive 62.5 points.

For instance, if the example situation above was part of this study and was selected for payment, and if your response had been "Very socially inappropriate", then you would receive 125 points if this was also the response provided by a randomly selected participant in this session. If your response did not match, you would receive 62.5 points.

If you have any questions, please raise your hand and wait for the experimenter to come to you. Otherwise, click on the next page to begin the to continue

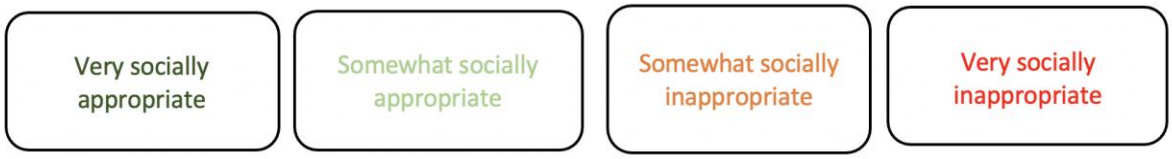
Your task for Part 1

Please provide your evaluation of each possible contribution choice that Individual A can make. Remember: You must consider each possible choice and evaluate how socially appropriate the action is:

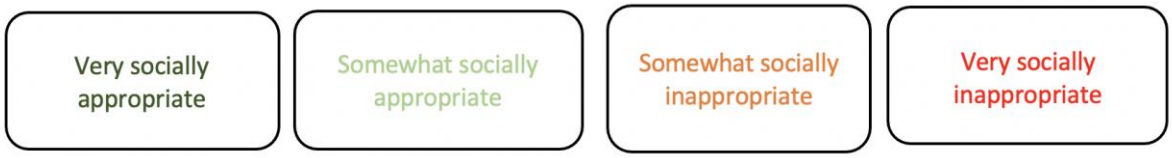
By socially appropriate, we mean behaviour that <u>most people in this room today</u> agree is the "correct" or "ethical thing to do".

A contributes ***5 tokens*** to the group account. Is this

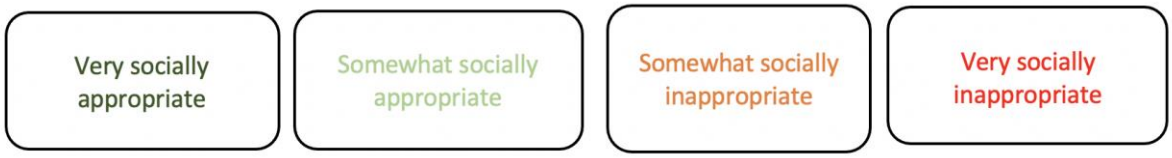
A contributes ***4 tokens*** to the group account. Is this



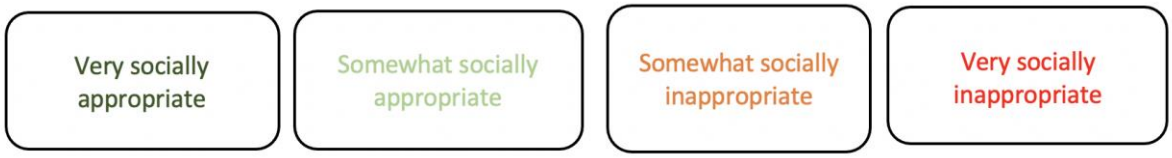
A contributes ***3 tokens*** to the group account. Is this



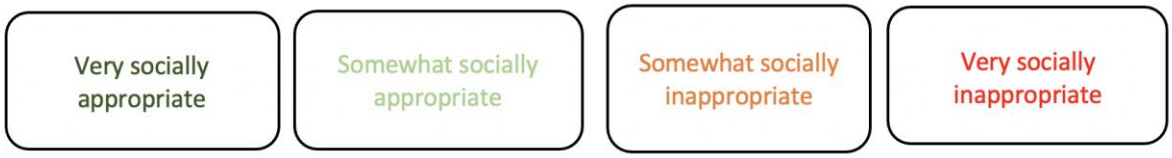
A contributes ***2 tokens*** to the group account. Is this



A contributes ***1 token*** to the group account. Is this



A contributes ***0 token*** to the group account. Is this



You have now finished Part 1. Please wait until all participants have completed this Part of the experiment.

General Instructions for Part 2

Groups

At the beginning of the experiment, all participants will be randomly divided into groups of three. Together with you, there will be 2 other participants in your group. You will not learn who the other participants in your group are, at any point in time.

The decision situation

Part 1 consists of 20 decision-making periods. In each period, you will be given an endowment of 5 tokens. Your task is to decide how many of these 5 tokens (between 0 and 5, included) to contribute to a "group account". Any token you do not contribute to the group account will be kept in your "private account". The other members of your group will have to make the same decision. They will also receive 5 tokens each, and decide how many to contribute to the group account or how many to keep in their private accounts.

Earnings

At the end of each period, after all group members have made their decisions, the computer will compute your earnings as follows. For each token kept in your private account, you will receive 1 point.

You will also receive points for the tokens contributed to the group account by you and the other group members. These tokens will be added up and multiplied by 1.5. The resulting number of points will then be divided equally between you and the other two group members (irrespective of how much they contributed to the group account).

The points you earn in a period are equal to the points from your private account + the points from the group account.

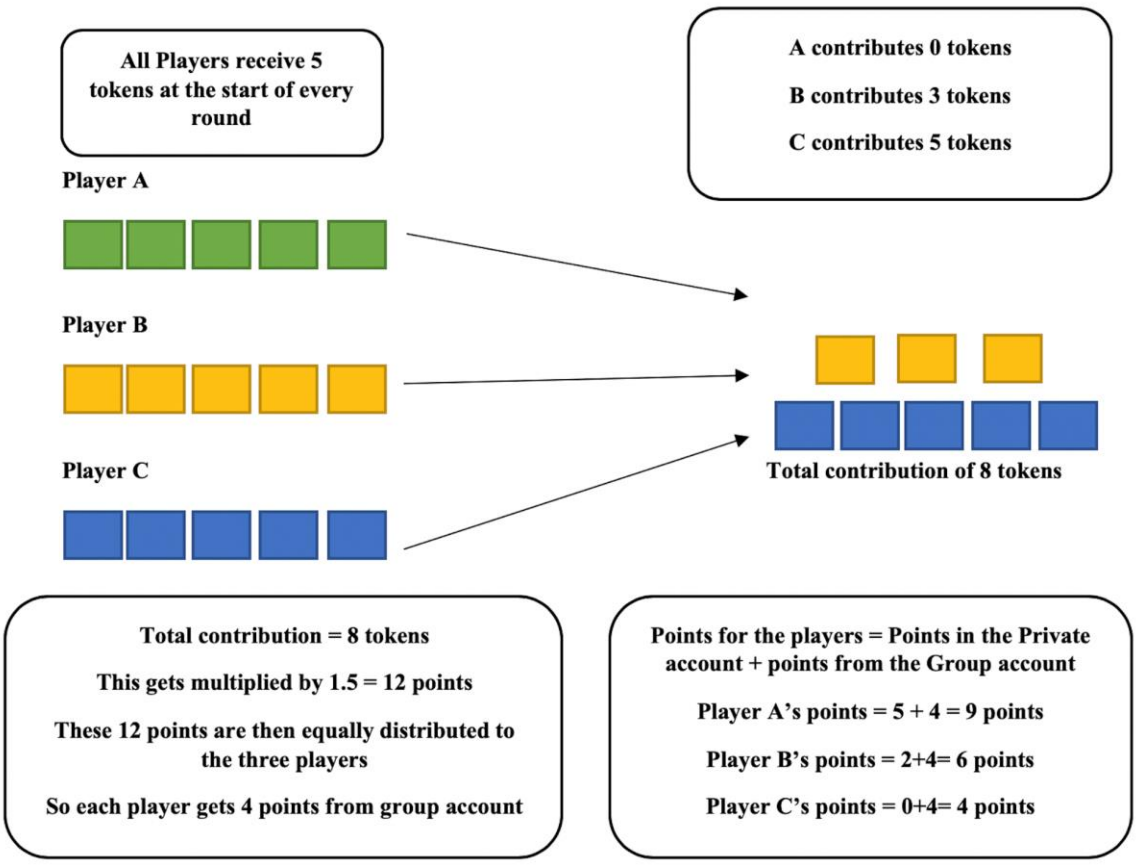
The earnings of the other two group members will be calculated in the same way. Your total earnings from Part 1 will be equal to the sum of your earnings in each of the 20 periods of Part 1.

Example

Suppose A decides to contribute 0 to the group account, and the other 2 members of the group contribute 3 tokens and 5 tokens respectively, giving a total of 8 tokens to the group account.

A's points from their private account = 5

To compute A's points from the group account we will proceed as follows: There are 8 tokens in total in the group account. These 8 tokens are multiplied by 1.5, yielding a total of 12 points. Each participant receives 12/3 = 4 points from the group account. A's points in total = Points from the private account + points from group account =5 + 4 = 9 points. Points of the other 2 group members = Points from their private accounts + the points from the group account (as shown below)



In the next page, we will go through a brief quiz to check your understanding of the task.

Questions

Please answer all questions below. They serve to ensure that everyone understands this part of the experiment.

1. At the start of a round, each group member receives 5 tokens. Suppose nobody (including you) contributes any tokens to the group account.

How many points will you earn in total?

How many points will each of the other two group members earn in total?

2. At the start of a round, each group member receives 5 tokens. Suppose you contribute 5 tokens to the group account. The other 2 group members contribute a total of 10 tokens (5 tokens each) to the group account.

How many points will you earn in total?

How many points will each of the other two group members earn in total?

3. At the start of a round, each group member receives 5 tokens. Suppose the other two group members each contribute a total of 10 tokens (5 tokens each) to the group account. Now suppose you contribute 0 tokens to the group account.

How many points will you earn in total?

How many points will each of the other two group members earn in total?

4. At the start of a round, each group member receives 5 tokens. Suppose the other two group members contribute a total of 6 tokens (3 tokens each) to the group account. Now suppose you contribute 2 tokens to the group account.

How many points will you earn in total?

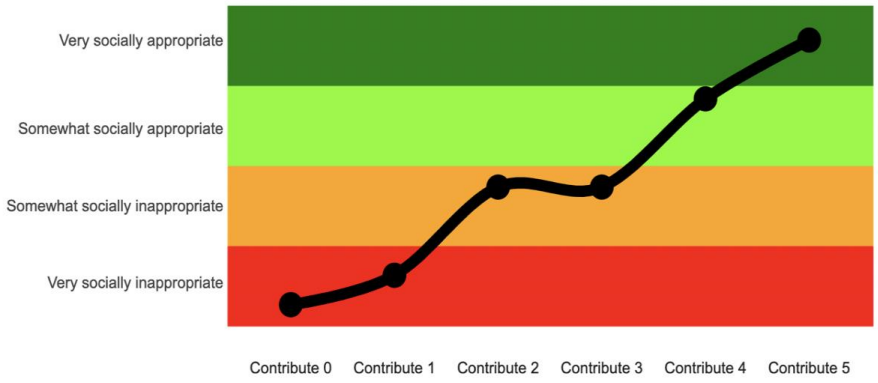
How many points will each of the other two group members earn in total?

In Part 1 we asked you to evaluate a decision situation that is exactly the same as the one you are about to face.

Before starting with round 1, we want to show you how you and the other two members of your group have evaluated this situation in Part 1 of the experiment.

We will show the same to the other two members of your group.

This graph shows what YOUR GROUP thinks about the social appropriateness of each level of contribution to the group account.



Screenshot of the group feedback from Stage 1 KW elicitation

A new round has started. You received an endowment of 5 tokens

Your contribution to the group project

What do you think will be the average contribution of the other two group members?

Results

Your contribution to the group account: **() tokens**.

Tokens kept in your private account: **() tokens**

Sum of contributions to your group account: **() tokens**.

This amount is multiplied by $multiplier$, yielding **() points**.

Each group member receives an equal share: **() points.**

Your earnings (in Points)

Points kept for yourself**: () points.**

Your share of the points from the group account**: () points.**

Your total earnings in this round**: () points.**

Your total earnings including this round**: () points.**

This is the end of Part 2

You have now finished Part 2. Please wait until all participants have completed this Part of the experiment.

General Instructions for Part 3

In this part of the experiment, you will be asked again to evaluate different possible choices an individual might make, in the same way as you did in Part 1.

We will ask you again to consider a decision situation that is identical to the situation you evaluated in Part 1 and that you actually faced in Part 2 (choice to allocate tokens between a private account and a group account).

We will show you 6 possible contribution choices that a person in such a situation ("Individual A") could make, and you will have to decide whether making each of those choices is “socially

appropriate" and "consistent with moral or proper social behaviour" or "socially inappropriate" and "inconsistent with moral or proper social behaviour".

Remember: by socially appropriate, we mean behaviour that most people in this room today agree is the "correct" or "ethical thing to do".

As in Part 1, your earnings for Part 3 will depend on whether your evaluation of Individual A's choices matches the evaluation of one other randomly selected participant in this session. **NOTE: this participant will be selected among those participants who were NOT in your group in Part 2**.

At the end of the experiment, if Part 3 is selected for payment, we will randomly select one of the 6 possible contribution choices by Individual A. We will compare you evaluation of that contribution choice with the evaluation of the other randomly selected participant. If your evaluation is exactly the same as theirs you will receive 125 points, otherwise you will receive 62.5 points. This amount will be paid to you in cash at the end of the experiment.

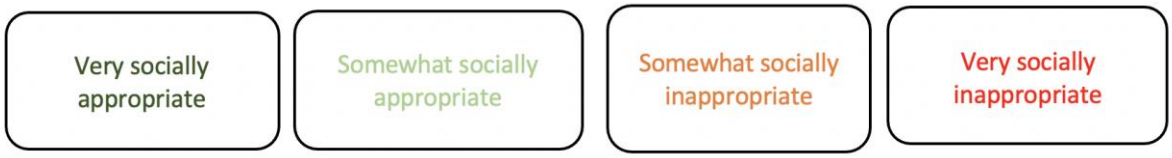
If you have any questions, please raise your hand and wait for the experimenter to come to you. Otherwise, click on the next page to begin the task.

Your task for Part 3

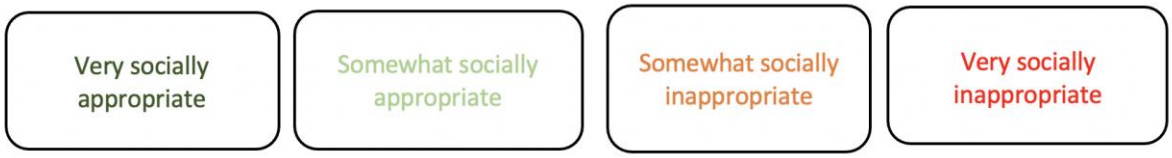
Please provide your evaluation of each possible contribution choice that Individual A can make. Remember: You must consider each possible choice and evaluate how socially appropriate the action is.

By socially appropriate, we mean behaviour that most people in this room today agree is the "correct" or "ethical thing to do".

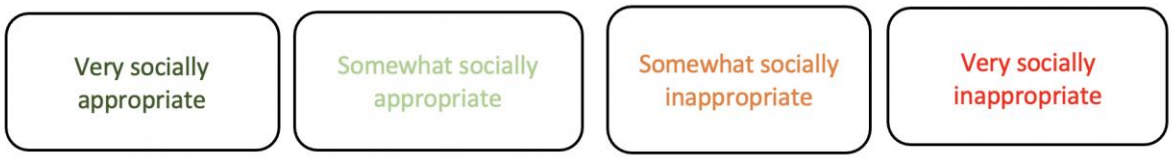
A contributes ***5 tokens*** to the group account. Is this



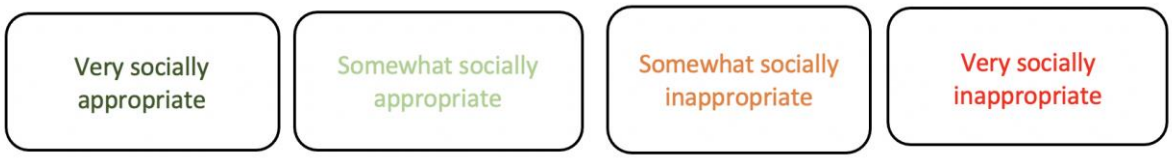
A contributes ***4 tokens*** to the group account. Is this



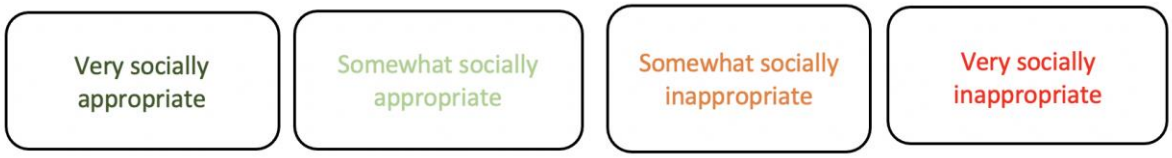
A contributes ***3 tokens*** to the group account. Is this



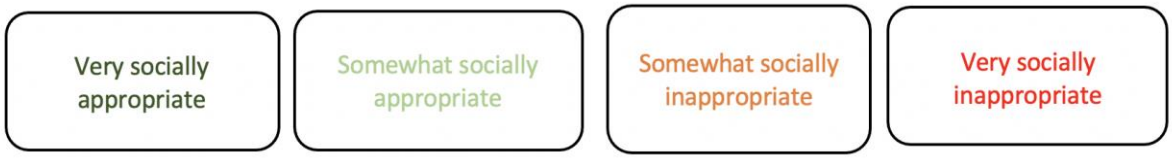
A contributes ***2 tokens*** to the group account. Is this



A contributes ***1 token*** to the group account. Is this



A contributes ***0 token*** to the group account. Is this



This is the end of Part 3

You have now finished Part 3. Please wait until all participants have completed this Part of the experiment

Questionnaire

What is your age?

What is your gender?

How much is your annual household income (approximately)?

What is your ethnicity?

What is your level of education?

Your earnings

This is the end of the experiment. You have completed all the parts of the experiment. Only one part of the experiment will be taken into account in determining your final earnings from today's experiment. If Part 1 is selected, and if one of your randomly drawn evaluations matches the evaluation of another randomly selected participant in the session, you will get 125 points. If it doesn't match, then you will get 62.5 points. If Part 2 is selected, then your earning will be based on the total amount of points you earned in the 20 periods of Part 2. If Part 3 is selected, and if one your randomly drawn evaluations matches the evaluation of another randomly selected participant in the session (but NOT in your group of Part 2), you will get 125 points. If it doesn't match, then you will get 62.5 points. Your point earnings will be converted into cash at a rate of £0.08 per point.

*For the online version, the instructions were the same except the addition of an extra stage before the start of the experiment (shown below). This was collected during the Covid-19 pandemic, during the months of November-December 2020. We recruited students from the University of Nottingham, using ORSEE. Subjects were paid using PayPal transfers. Subjects first entered a virtual setup through Microsoft Teams. Here they conversed with the

experimenter using the “Chat box”, through a 1-1 interaction. This was to make sure that the subjects couldn’t interact with one another.

On entering a session, they first typed in their student ID. Upon verification, there were then given the online link to the experiment. They then entered the LIONESS interface and completed the experiment within 60 minutes. Then then returned to the Chat box to inform the experiment of their submission. This was the end of a session. *****

Hello! Thank you for taking part in this study.

You will receive the link for today's experiment in the Chat box in a few minutes. **Please copy paste the link in GOOGLE CHROME.** If you default browser is GOOGLE CHROME, you can just click on the link. You will be automatically directed into the experiment, which will open as a new webpage on your browser. **Note that you will be allowed to use this link only once**. You can read through the instructions and proceed with the stages.

Do not close the Teams Chat-box at any point in time. This is the only way the experimenter can communicate with you, if you have any queries during the experiment.

The instructions in the initial screens will explain how you can earn money in today’s experiment. It is therefore important, that you read the instructions carefully.

You may have to wait for other players in some stages of the experiment. You will get a waiting message in some screens. Do not worry much, and please wait patiently for others, when asked.

If you have any questions, please let me know by typing your query in the chat box in the Teams meeting. I will respond to your questions directly through chat. No Video or audio needs to be open for this.

It is important that you complete the experiment without any interruptions. During the experiment, please do not close the experiment window at any point in time. If you leave the experiment midway, we will not be able to pay you and the other participants in today's study. If you close the experiment window accidentally, message me in the Chat-box immediately, and I will help you out.

Once you complete the experiment, you can let me know in the Chat-box, and I will confirm your participation.

Are there any questions at this stage? If not, Type OK and I will share the experiment link in a few minutes. Please note that mobile phones should not be used during the experiment.

D-method Instructions

Welcome

Thank you for participating in our study. Including the time for reading these instructions, the study will take about 10 minutes to complete. During the study, please do not close this window or leave the HIT's web pages in any other way. If you do close your browser or leave the HIT, you will not be able to re-enter, and we will not be able to pay you.

During this HIT you can earn points. These points will determine your bonus payment for this HIT. The number of points you earn depends on your decisions. At the end of the HIT, your points will be converted into real money **(20 points are worth $1).**

In addition to this bonus, you will receive a guaranteed participation fee of $0.50 upon completion. You will receive a code to collect your payment via MTurk at the end of this HIT.

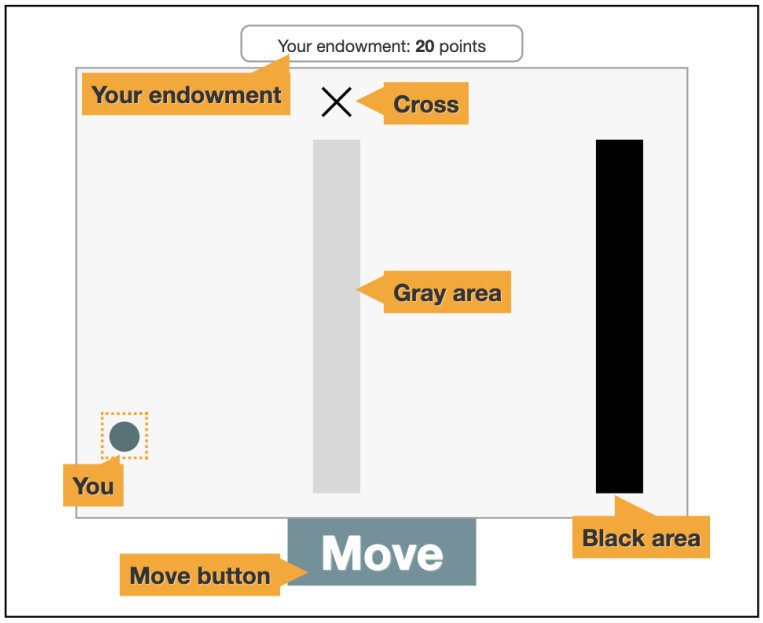
Note: If you are using a desktop/laptop to complete this HIT, we recommend that you maximise your browser screen before you start

Instructions 1/2

In this HIT, we ask you to read a short description of a potential study, to imagine two possible outcomes of this potential study, and then to answer several questions about these outcomes.

Imagine that we would run the following study with 100 participants who, like yourself, are recruited via Amazon Mechanical Turk from the USA.

Participants are asked to complete a computerized task in which they control a circle figure that they have to move across the screen into the black area (the vertical bar on the right-hand side of the screen). The figure below shows a screenshot of the task.



When a participant clicks the "Move" button at the bottom of the screen, his or her circle approaches the grey area marked with the cross (X) and stops to wait. To make the circle move again, the participant has to click again the "Move" button.

Participants are told that **the rule is to wait in the grey area until the cross disappears.**

Participants' earnings are determined by the amount of time it takes them to move their circle into the black area. Specifically, they begin with an initial endowment of 20 points. Each second spent on the task, the endowment decreases by 1 point. The endowment stops decreasing when the circle reaches the black area or when the task ends (after 20 seconds).

Instructions 2/2

Before playing the computerized task, participants answer two questions to ensure they comprehend the task.

1) They are asked to calculate their earnings if they move **after** the cross has disappeared and finish the task in 11 seconds [correct answer: 9 points].

2) They are also asked to calculate their earnings if they move **before** the cross has disappeared and finish the task in 5 seconds [correct answer: 15 points].

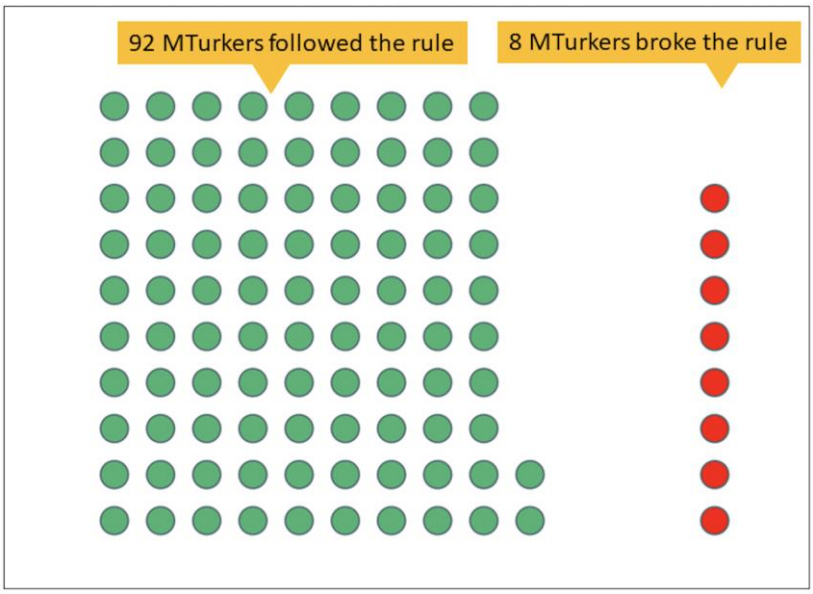
After answering these questions, participants are reminded of the rule and then play the task.

After the task, they are paid according to the time it takes them to complete it.

Possible outcomes of the study

Imagine that the results of the study we described to you would be as follows.

Of the 100 MTurkers who play the task, 92 follow the rule and move after the cross disappears, while 8 break the rule and move before the cross disappears. This outcome is shown in the figure below.



We now ask you to answer a number of questions about this imagined outcome of the study. Please consider each question carefully before answering it.

8 MTurkers broke the rule and moved their circle before the cross had disappeared. Can you imagine why they would do that?[73]

92 MTurkers followed the rule and moved their circle after the cross had disappeared. Can you imagine why they would do that?

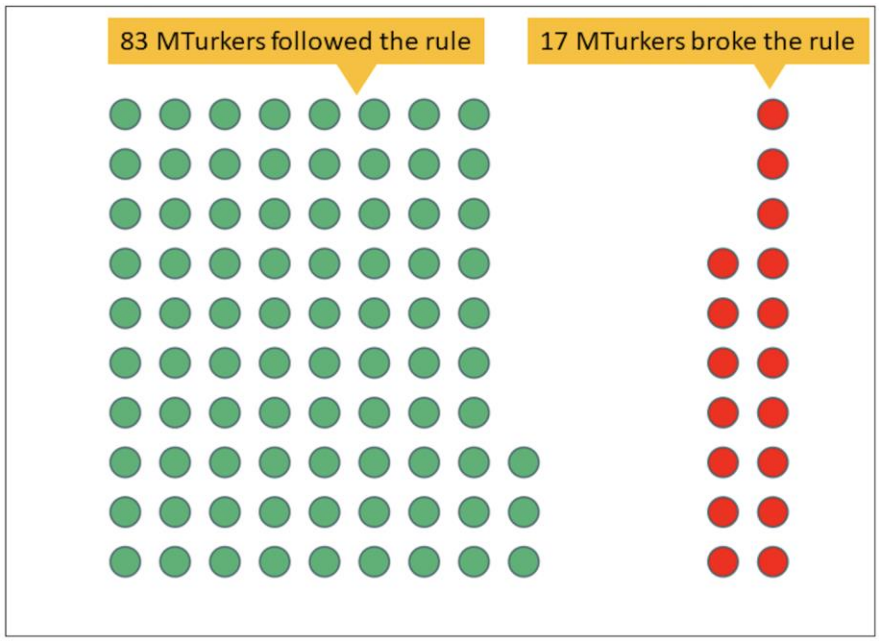
[73] Two scenarios were shown, one after the other in all treatment conditions. The number of rule-followers for all scenarios (2 per treatment; 8 in total) were: LOW: 92 followed, 83 followed; MID LOW: 67 followed, 58 followed; MID HIGH: 42 followed, 33 followed; HIGH: 17 followed, 8 followed. Since the treatments use the same set of instructions, we show only from LOW to avoid repetitions.

How satisfied do you think that the MTurkers who **broke the rule** and moved before the cross disappeared would be?

How satisfied do you think that the MTurkers who **followed the rule** and moved after the cross disappeared would be?

Now imagine that the results of the study would be as follows.

Of the 100 MTurkers who play the task, 83 follow the rule and move after the cross disappears, while 17 break the rule and move before the cross disappears. This outcome is shown in the figure below.



How satisfied do you think that the MTurkers who **broke the rule** and moved before the cross disappeared would be?

How satisfied do you think that the MTurkers who **followed the rule** and moved after the cross disappeared would be?

Which of the two imagined outcomes described above do you think is more realistic?

We actually ran the study that we just described to you with participants who, like you, were recruited via Amazon Mechanical Turk from the USA.

Please estimate the fraction (in percent) of MTurkers in this previous study who **moved before the cross had disappeared**. If your estimate is accurate with an error of at most +/- 3 percentage points, we will pay you $0.50 at the end of this experiment.

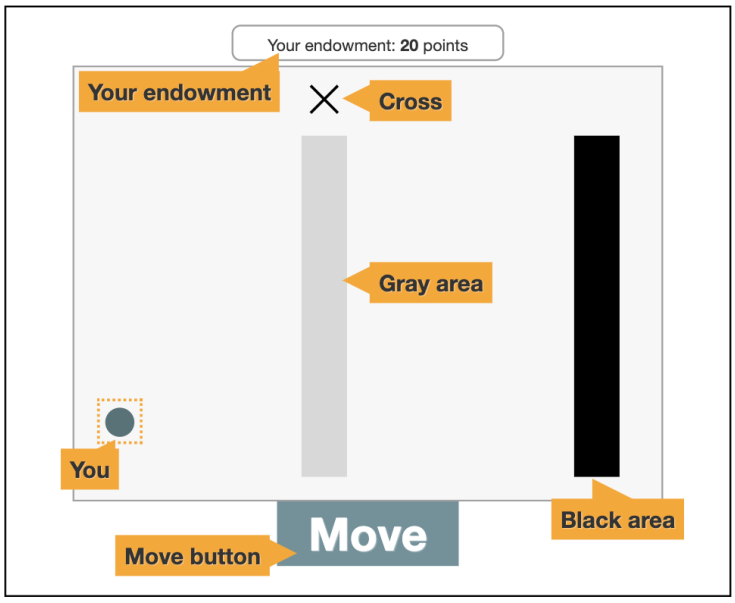
We also ran a version of the study where we asked a different group of MTurkers from the USA to tell us how appropriate they think it is to move before the cross disappears. **82%** told us that moving before the cross disappears is **inappropriate.**

Your payment for taking part in today's HIT

On top of the money that you may earn if you have answered the question above correctly, we will pay you an additional sum of money for having taken part in this study.

To determine how much money, you will be paid we ask you to play the same computerized task as the participants in the study that we described before.

You will control a circle figure that you have to move across the screen into the black area.



When you click the Move button your circle will approach the grey area and stop to wait. To make the circle move again, you will have to again click the Move button.

The rule is to wait in the grey area until the cross disappears.

Your earnings are determined by the amount of time it takes you to move your circle into the black area. You begin with an initial endowment of 20 points, which decreases by 1 point for each second spent on the task. The endowment stops decreasing when your circle reaches the black area or when the task ends (after 20 seconds).

End of Task

This is the end of the task. Please click below to finish the task. You can view your earnings after filling out a brief questionnaire.

Questionnaire

What is your gender?

What is your age?

Please select your ethnicity (White/African American/Hispanic/Alaskan Native/Asian/Native Hawaiian/Others)

Which part of the country are you from?

(North/South/East/West/Central)

What is your total household income?

(Less than $10,000/$10,000-$60,000/ More than $60,000)

What is your level of education?

(High School/ Bachelor's Degree/ Master's degree)

What is your weekly income from Mturk experiments?

(Less than $1/ $1-$10/ $10-$20/More than $20)

In politics, people sometimes talk about the ‘left’ and the ‘right’. Where would you place your own views on a scale from 0 to 10? (0 means the most left and 10 means the most right)

Please provide your feedback from this HIT in the box below.

How easy did you find the instructions for this HIT?

How tedious did you find the HIT?

S-Method Instructions

Welcome!

Thank you for participating in our study. Including the time for reading these instructions, the study will take about 10 minutes to complete. During the study, please do not close this window or leave the HIT’s web pages in any other way. If you do close your browser or leave the HIT, you will not be able to re-enter, and we will not be able to pay you.

During this HIT you can earn points. These points will determine your bonus payment for this HIT. The number of points you earn depends on your decisions. At the end of the HIT, your points will be converted into real money **(20 points are worth $1).** In addition to this bonus, you will receive a guaranteed participation fee of $0.50 upon completion. You will receive a code to collect your payment via MTurk at the end of this HIT.

Note: If you are using a desktop/laptop to complete this HIT, we recommend that you maximise your browser screen before you start

Instructions 1/2

In this HIT, you control a circle figure that you have to move across the screen into the black area. Your circle started on the left-hand side of the screen. It has approached the grey area marked with the cross (X) and has stopped to wait.

It is your task to move your circle into the black area (the vertical bar on the right-hand side of the screen). The figure below shows a screenshot of the task. You have to decide when to make your circle move again. You can choose whether to wait until the cross(X) disappears or move before the cross disappears.

The rule is to wait in the grey area until the cross disappears.

Your earnings are determined by the amount of time it takes you to move your circle to the black area. Specifically, if you wait until the cross disappears, it will take your circle 11 seconds to reach the black area and your earnings will be 9 points. If you move before the cross disappears, it will take your circle 5 seconds to cross the finish line and your earnings will be 15 points.

Instructions 2/2

Over **100** MTurkers from the USA previously participated in a similar HIT as the one you are participating in today. These MTurkers also could choose whether to wait until the cross (X) disappears, or to move before it disappears.

In today's HIT, we will show you 4 possible outcomes of this previous HIT. We will ask you, for each outcome, whether you want to wait until the cross disappears or to move before it disappears.

The 4 possible outcomes differ **in the fraction of previous MTurkers who broke the rule and moved before the cross (X) disappeared:**

A) Between 0% and 25% of previous MTurkers moved before the cross disappeared

B) Between 26% and 50% of previous MTurkers moved before the cross disappeared

C) Between 51% and 75% of previous MTurkers moved before the cross disappeared

D) Between 76% and 100% of previous MTurkers moved before the cross disappeared

For each possible outcome, you have to choose whether to **wait** in the grey area until the cross disappears or **move** before the cross disappears. Only one of these 4 choices will be used to compute your bonus earnings. At the end of the HIT, we will reveal the actual outcome of the previous HIT. This is the actual percentage of previous MTurkers who broke the rule and moved before the cross disappeared. We will use your choice corresponding to this outcome to compute your bonus earnings.

For example, suppose that the actual outcome of the previous HIT was that 50% of previous MTurkers broke the rule and moved before the cross disappeared. In that case your choice for outcome B will be used to compute your bonus earnings. Please click "Continue" below if you understand these instructions.

We also ran a version of the study where we asked a different group of MTurkers from the USA to tell us how appropriate they think it is to move before the cross disappears.

82% told us that moving before the cross disappears is **inappropriate.**

Remember: The rule is to wait in the grey area until the cross(X) disappears.

Make your choices

Choose whether to wait in the grey area until the cross (X) disappears or move before it disappears for each of the 4 possible outcomes listed below:

A. Between 0% and 25% of previous MTurkers moved before the cross disappeared.

Wait/Move

B. Between 26% and 50% of previous MTurkers moved before the cross disappeared.

Wait/Move

C. Between 51% and 75% of previous MTurkers moved before the cross disappeared.

Wait/Move

D. Between 76% and 100% of previous MTurkers moved before the cross disappeared.

Wait/Move

End of Task

This is the end of the task. Please click below to finish the task.

You can view your earnings after filling out a brief questionnaire[74].

[74] We use the same set of questions for the questionnaire as the D-method.

Cross-cultural Experiments

In all the experiments, subjects were first asked two qualifying questions, namely questions regarding their age and nationality. We recruited participants from both countries only if their age was within the range of 18-40 years. This is illustrated below only for Experiment 1 to avoid unnecessary repetitions. The questionnaire at the end uses the same questions for all studies.

Before you can participate in the study, please answer the following questions

What is your age?

What is your nationality? (Select from the list of 51 European Countries)

Part 1: The behavioural task: Abstract task

Welcome! Thank you for participating in our study!

The study will take about 10 minutes to complete. During the study, please do not close this window or leave the screen in any other way. If you do close your browser or leave the screen, you will not be able to re-enter, and we will not be able to pay you.

You will be paid XXX[75] points for taking part in this study. Moreover, you can earn **additional points depending on your decisions**. Each point you earn is worth approximately 0.5 Turkish Lira/2.75 Swedish Krona

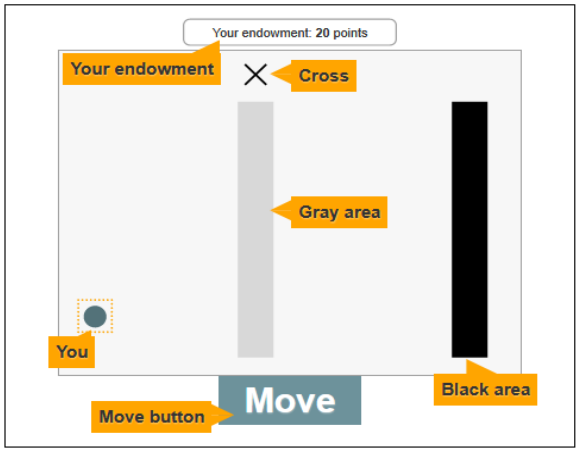
Note: If you are using a desktop/laptop to complete this study, we recommend that you maximise your browser screen before you start. Please click below to start the study.

[75] All payments were finalised by Qualtrics where respondents' incentivization is conducted via many different channels and can vary depending on the target group. They provided us with some general information, suggestive of the fact that payments could include cash, airline miles, gift cards, redeemable points, charitable donations, sweepstakes entrance and vouchers. We keep the incentives (in local currency value) XXX all for experiments.

Instructions

In this study you control a circle figure that you have to move across the screen into the black area (the vertical bar on the right-hand side of the screen).

The figure below shows a screenshot of the task.



When you click the Move button at the bottom of your screen, your circle will approach the grey area marked with the cross (X) and stop to wait. To make the circle move again, you have to click the Move button once more.

The rule is to wait in the grey area until the cross disappears.

The additional number of points you will earn today is determined by the amount of time it takes your circle to reach the black area. You begin with an initial endowment of 20 points, which decreases by 1 point for each second spent on the task. The endowment stops decreasing when your circle reaches the black area or when the task ends (after 20 seconds).

Questions

Before performing the task, please answer two questions to ensure you understand the task.

1) How many additional points will you earn if you move your circle **after** the cross has disappeared and finish the task in 13 seconds?

2) How many additional points will you earn if you move your circle **before** the cross has disappeared and finish the task in 4 seconds?

Remember:

When you are ready, click the Move button once to start moving your circle.

After the first click, the circle will stop on its own when it reaches the grey area.

Click the Move button again to move the circle into the black area.

The rule is to wait in the grey area until the cross disappears.

OK

Questionnaire

This completes the task. Please fill out this brief questionnaire before we show you how many points you have earned today.

What is your gender? (Male / Female / Other)

Which country were you born in?

Which city do you live in?

What is the total monthly income of your household?

What is the highest level of education that you have completed? (High School / Bachelor's degree / Master's degree / PhD)

Your earnings in the study

This is the end of the study. In the task, you left the grey area before (after) the cross disappeared. You earned XXX additional points (depends on how long the subject took to complete the task). Your participation fee is: XXX points. So, in total, you have earned XXX points from this study. Each point is worth 0.5 Turkish Lira/2.75 Swedish Krona. We will pay this amount of points in your account in due course. Thank you for taking part in the study!

Experiment 2: Belief Elicitation

Thank you for participating in our study!

The study will take about 10 minutes to complete. During the study, please do not close this window or leave the screen in any other way. If you do close your browser or leave the screen, you will not be able to re-enter, and we will not be able to pay you.

You will be paid XXX points for taking part in this study. Moreover, you can earn **additional points depending on your decisions**. Each point you earn is worth approximately 0.5 Turkish Lira/2.75 Swedish Krona.

Note: If you are using a desktop/laptop to complete this study, we recommend that you maximise your browser screen before you start. Please click below to start the study

Instructions (1 of 3)

We have recruited over 100 participants from your country to take part in a computerized task. In this study you will read a description of this task and answer a few questions about it. The number of additional points you will earn today **depends on how you answer these questions**, so please read the instructions carefully.

Instructions (2 of 3)

In two of the questions, we will ask you to evaluate the behaviour of a **person from your own country** who has participated in the task that we will describe to you. In each question, we will ask you to rate how **socially appropriate** the behaviour of this person is:

Very socially appropriate	Socially appropriate	Socially inappropriate	Very socially inappropriate

By "**socially appropriate**" we mean behaviour that most people agree is "**the right thing to do**". Another way to think about what we mean is that, if someone were to take a "**socially inappropriate**" action, then most people would **disapprove** of their behaviour.

Instructions (3 of 3)

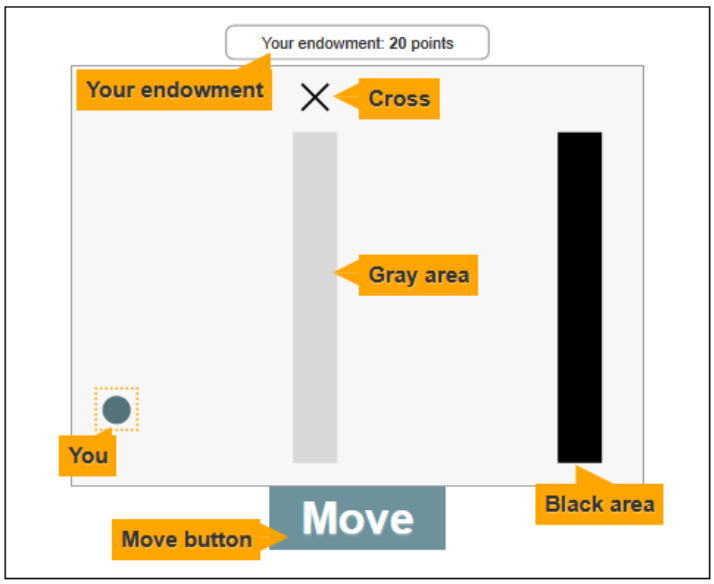
Your answers to the two questions will determine **how many additional points** you earn today. This works as follows. We will ask the same two questions to about **200 other people recruited from your own country**. After the study, we will randomly select one of the two questions and we will check how **most people** have rated the behaviour described in this question. If your rating matches the most common rating, we will pay you **20 additional points**.

For example, imagine that, for the selected question, most people from your country have rated the behaviour as "Socially inappropriate". If you have also rated the behaviour described in that question as "Socially inappropriate" you will earn 20 points. Otherwise, you will earn 0 points.

Description of the task you have to evaluate (1 of 2)

We have recruited over 100 participants from your country to take part in the following task. In the task, each participant controlled a circle figure that he/she had to move across the screen into the black area (the vertical bar on the right-hand side of the screen). The figure below shows a screenshot of the task.

When the participant clicked the Move button at the bottom of the screen, the circle approached the grey area marked with the cross (X) and stopped to wait. To make the circle move again, the participant had to click the Move button once more.



We told participants that **the rule is to wait in the grey area until the cross disappears.**

The earnings of the participant were determined by the amount of time it took the circle to reach the black area. The participants began with an initial endowment of 20 points, which decreased by 1 point for each second spent on the task. The endowment stopped decreasing when the circle reached the black area or when the task ended (after 20 seconds).

Description of the task you have to evaluate (2 of 2)

Before performing the task, participants completed a quiz to check their understanding of the task.

1) They were asked to calculate how many points they would earn if they moved the circle **after** the cross disappeared and finished the task in 13 seconds. [Correct answer: 7 points]

2) They were asked to calculate how many points they would earn if they moved the circle **before** the cross disappeared and finished the task in 4 seconds. [Correct answer: 16 points]

Note: your questions from today's study will appear on the next screen. **Your earnings today depend on how you answer these questions**. Once you click "CONTINUE", you will not be able to go back to the description of the task you have to evaluate, or to your instructions. If you want to read them again, go back now. Otherwise click "CONTINUE"

Questions

Think about **a person from your own country** who has participated in the task we have just described to you.

Question 1

Imagine that this person **broke the rule and moved before the cross disappeared**, finishing the task in 15 seconds, and earning 5 points. How socially appropriate do you think their behaviour is? (Click on one of the ratings to answer). Remember: if your rating of this behaviour matches how most people from your country also rate the behaviour, we will pay you 20 additional points. Otherwise, you will earn 0 points.

Think again about **a person from your own country** who has participated in the task.

Question 2

Imagine that this person **followed the rule and moved after the cross disappeared**, finishing the task in 9 seconds and earning 11 points. How socially appropriate do you think their behaviour is? (Click on one of the ratings to answer). Remember: if your rating of this behaviour matches how most people from your country also rate the behaviour, we will pay you 20 additional points. Otherwise, you will earn 0 points.

We have one final question for you. We now want you to think about the **over 100 participants whom we have recruited in your country** to actually perform the task we have described.

How many of them [out of 100] do you think **broke the rule** and moved the circle before the cross disappeared? If your answer is accurate with an error of at most +/- 3 percentage points, we will pay you **20 additional points** at the end of this study

Questionnaire

Please fill out this brief questionnaire before the end of the study[76].

Your earnings in the study

This is the end of the study. As we explained before, when we will have collected all responses from this study, we will select **one of the first two questions** at random and calculate the most common rating in that question. If your answer in that question matches the most common rating, we will pay you 20 points.

We will also compare your answer to **the third question** (about how many people followed the rule) to the actual fraction of participants who have participated in the task and followed the rule. If your answer is accurate (+/- 3 percentage points), we will pay you an additional 20 points. On top of the points, you earn from these questions, we will pay you a **participation fee** of XXX points. Each point is worth 0.5 Turkish Lira/2.75 Swedish Krona. We will pay this amount of points in your account in due course. Thank you for taking part in the study!

[76] We use the same set of questions for all the 4 set of experiments.

Experiment 3a: Preference Elicitation with empirical information

Thank you for participating in our study!

The study will take about 10 minutes to complete. During the study, please do not close this window or leave the screen in any other way. If you do close your browser or leave the screen, you will not be able to re-enter, and we will not be able to pay you.

You will be paid XXX points for taking part in this study. Moreover, you can earn **additional points depending on your decisions**. Each point you earn is worth approximately 0.5 Turkish Lira/2.75 Swedish Krona.

Note: If you are using a desktop/laptop to complete this study, we recommend that you maximise your browser screen before you start.

Instructions (1 of 3)

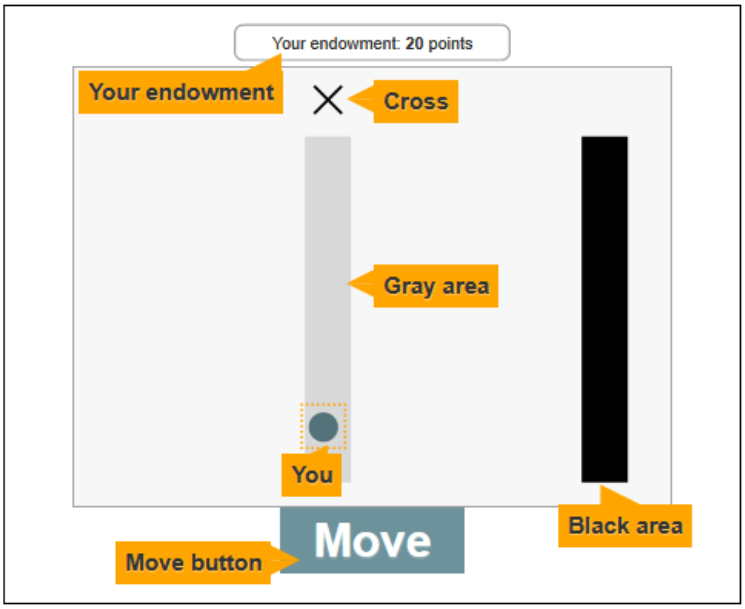
In this study, you control a circle figure that you have to move across the screen into the black area (the vertical bar on the right-hand side of the screen). The figure below shows a screenshot of the task.

You have to decide when to make your circle move. You can choose whether to wait until the cross disappears or move before the cross disappears.

The rule is to wait in the grey area until the cross disappears.

Your earnings in today's study are determined by the amount of time it takes your circle to reach the black area. Specifically, if you wait until the cross disappears, it will take your circle 15 seconds to reach the black area and your earnings will be 5 points. If you instead move

before the cross disappears, it will take your circle 4 seconds to reach the black area and your earnings will be 16 points.



Instructions (2 of 3)

You will decide whether to wait in the grey area until the cross disappears or move before it disappears **in 4 different scenarios**. These scenarios correspond to 4 different outcomes of a study similar to the one you are taking part in today, that we have already conducted with over 100 participants from your own country. The 4 outcomes differ in the **fraction of these previous participants who broke the rule** and moved **before** the cross disappeared**:**

(A) Between **0% and 25% broke the rule** and moved before the cross disappeared

(B) Between **26% and 50% broke the rule** and moved before the cross disappeared

(C) Between **51% and 75% broke the rule** and moved before the cross disappeared

(D) Between **76% and 100% broke the rule** and moved before the cross disappeared

For each scenario, you will decide whether to **wait** until the cross disappears or **move** before it disappears.

Instructions (3 of 3)

You will therefore make 4 choices in total, one for each possible outcome of the previous study that we have already conducted in your country.

(A) Between **0% and 25% broke the rule** and moved before the cross disappeared

(B) Between **26% and 50% broke the rule** and moved before the cross disappeared

(C) Between **51% and 75% broke the rule** and moved before the cross disappeared

(D) Between **76% and 100% broke the rule** and moved before the cross disappeared

However, only one of these 4 choices will be used to calculate **your additional points in today's study**. At the end of the study, we will reveal the **actual percentage of previous participants from your own country who moved before the cross disappeared**. We will use your choice corresponding to this outcome to compute your additional points. For example, suppose that 50% of previous participants moved before the cross disappeared. In that case your choice for scenario (B) will be used to compute your additional points.

Note: your task is about to begin. **Your earnings today depend on your choices in the task**. Once you click "CONTINUE", you will not be able to go back to the instructions for the task. If you want to read them again, go back now. Otherwise click "CONTINUE".

Just so you know: We also ran a version of this study where we asked a different group of participants from your country to tell us **how socially appropriate it is to move before the cross disappears. 72%** (**SWE)/ 46% (TUR)** told us that moving before the cross disappears is **inappropriate.**

Make Your Choices

Choose whether to wait in the grey area until the cross disappears or move before it disappears for each of the 4 possible scenarios listed below. Remember: **The rule is to wait until the cross disappears**

Scenario (A)

Between **0% and 25%** of participants from your country moved **before** the cross disappeared

WAIT/MOVE

Scenario (B)

Between **26% and 50%** of participants from your country moved **before** the cross disappeared

WAIT/MOVE

Scenario (C)

Between **51% and 75%** of participants from your country moved **before** the cross disappeared

WAIT/MOVE

Scenario (D)

Between **76% and 100%** of participants from your country moved **before** the cross disappeared

WAIT/MOVE

Questionnaire

This completes the study. Please fill out this brief questionnaire before we show you how many points you have earned today.

Your earnings in the study

This is the end of the study. In the previous task, 45% SWE (61% TUR) of participants broke the rule and moved before the cross disappeared. We will therefore use your choice in scenario (B/C) to compute your earnings. In this scenario, you left the grey area before (after) the cross disappeared. You earned XXX additional points. (Follow: 5 points/ violate: 16 points)

Your participation fee is: XXX points. So, in total, you have earned XXX points from this study. Each point is worth 0.5 Turkish Lira/2.75 Swedish Krona. We will pay this amount of points in your account in due course. Thank you for taking part in the study!

Experiment 3b: Preference Elicitation with normative information

Thank you for participating in our study!

The study will take about 10 minutes to complete. During the study, please do not close this window or leave the screen in any other way. If you do close your browser or leave the screen, you will not be able to re-enter, and we will not be able to pay you.

You will be paid XXX points for taking part in this study. Moreover, you can earn **additional points depending on your decisions**. Each point you earn is worth approximately 0.5 Turkish Lira/2.75 Swedish Krona.

Note: If you are using a desktop/laptop to complete this study, we recommend that you maximise your browser screen before you start.

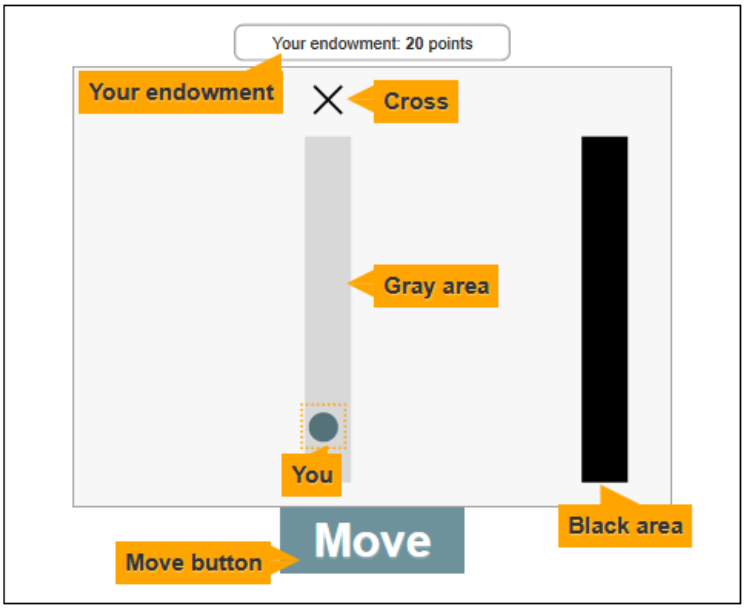
Instructions (1 of 3)

In this study, you control a circle figure that you have to move across the screen into the black area (the vertical bar on the right-hand side of the screen). The figure below shows a screenshot of the task. You have to decide when to make your circle move. You can choose whether to wait until the cross disappears or move before the cross disappears.

The rule is to wait in the grey area until the cross disappears.

Your earnings in today's study are determined by the amount of time it takes your circle to reach the black area. Specifically, if you wait until the cross disappears, it will take your circle 15 seconds to reach the black area and your earnings will be 5 points.

If you instead move before the cross disappears, it will take your circle 4 seconds to reach the black area and your earnings will be 16 points.



Instructions (2 of 3)

You will decide whether to wait in the grey area until the cross disappears or move before it disappears **in 4 different scenarios**. These scenarios correspond to 4 different outcomes of a study similar to the one you are taking part in today, that we have already conducted with over 100 participants from your own country.

In this previous study, we asked participants to tell us how "**socially appropriate**" it is to break the rule and move before the cross disappears. By "socially appropriate" we mean behaviour that most people agree is "**the right thing to do**". Another way to think about what we mean is that, if someone were to take a "**socially inappropriate**" action, then most people would **disapprove** of their behaviour.

The 4 outcomes we will show you today differ in the **fraction of previous participants who told us that breaking the rule** and moving before the cross disappears **is socially inappropriate:**

(A) Between **0% and 25%** said that **breaking the rule** and moving before the cross disappears is **inappropriate**

(B) Between **26% and 50%** said that **breaking the rule** and moving before the cross disappears is **inappropriate**

(C) Between **51% and 75%** said that **breaking the rule** and moving before the cross disappears is **inappropriate**

(D) Between **76% and 100%** said that **breaking the rule** and moving before the cross disappears is **inappropriate**

For each scenario, you will decide whether to **wait** until the cross disappears or **moving** before it disappears.

Instructions (3 of 3)

You will therefore make 4 choices in total, one for each possible outcome of the previous study that we have already conducted in your country.

(A) Between **0% and 25%** said that **breaking the rule** and moving before the cross disappears is **inappropriate**

(B) Between **26% and 50%** said that **breaking the rule** and moving before the cross disappears is **inappropriate**

(C) Between **51% and 75%** said that **breaking the rule** and moving before the cross disappears is **inappropriate**

(D) Between **76% and 100%** said that **breaking the rule** and moving before the cross disappears is **inappropriate**

However, only one of these 4 choices will be used to calculate **your additional points in today's study**. At the end of the study, we will reveal the **actual percentage of previous participants from your own country who told us that moving before the cross disappears is inappropriate**. We will use your choice corresponding to this outcome to compute your additional points. For example, suppose that 50% of previous participants told use that moving

before the cross disappears is inappropriate. In that case your choice for scenario (B) will be used to compute your additional points.

Note: your task is about to begin. **Your earnings today depend on your choices in the task**. Once you click "CONTINUE", you will not be able to go back to the instructions for the task. If you want to read them again, go back now. Otherwise click "CONTINUE"

Just so you know: We also ran a version of this study where we asked a different group of participants from your country to **simply choose whether to wait until the cross disappears or move before it disappears**. **45% SWE/ 61% TUR** chose to **move before the cross disappears.**

Make Your Choices

Choose whether to wait in the grey area until the cross disappears or move before it disappears for each of the 4 possible scenarios listed below.

Remember: **The rule is to wait until the cross disappears**

Scenario (A)

Between **0% and 25%** of participants from your country said that moving **before** the cross disappears is **inappropriate**.

WAIT/MOVE

Scenario (B)

Between **26% and 50%** of participants from your country said that moving **before** the cross disappears is **inappropriate.**

WAIT/MOVE

Scenario (C)

Between **51% and 75%** of participants from your country said that moving **before** the cross disappears is **inappropriate.**

WAIT/MOVE

Scenario (D)

Between **76% and 100%** of participants from your country said that moving **before** the cross disappears is **inappropriate.**

WAIT/MOVE

Questionnaire

This completes the study. Please fill out this brief questionnaire before we show you how many points you have earned today.

Your earnings in the study

This is the end of the study. In the previous task, **72%** (**SWE)/ 46% (TUR)** of participants told us that moving before the cross disappeared is inappropriate. We will therefore use your choice in scenario (C/B) to compute your earnings. In this scenario, you left the grey area before (after) the cross disappeared. You earned XXX additional points (5 or 15 points). Your participation fee is: XXX points. So, in total, you have earned XXX points from this study.

Each point is worth 0.5 Turkish Lira/2.75 Swedish Krona. We will pay this amount of points in your account in due course. Thank you for taking part in the study!