Experiments on discrimination and social norms

Tom Lane

University of Nottingham, Department of Economics

PhD thesis, submitted March 2017

Supervised by Abigail Barr and Daniele Nosenzo

Abstract

This dissertation presents three projects within the fields of behavioural and experimental economics. The first consists of a meta-analysis of lab experiments measuring economic discrimination. Most importantly, I find that the strength of discrimination in economics experiments varies depending on the dimension of identity across which discrimination is measured, and depending on the type of game used to measure it. The second project investigates the relationship between discriminatory behaviour and social norms. A lab experiments finds that discrimination is stronger when it is perceived to be more socially appropriate. In the third project, a field experiment investigates the effect of different nudges on voter registration rates. In particular, emphasising the possibility of being fined for failing to register is successful in raising registration rates, but offering the possibility of financial gain for registering is not. An online experiment in the same project suggests the conflicting normative effects of the two nudges may help explain these differences.

Table of contents

Chapter One: Introduction

This dissertation presents three research projects. All three projects fall within the fields of behavioural and experimental economics, although each uses a different methodology, the first being a meta-analysis, the second a lab experiment and the third a combination of field and online experiments. The research questions addressed also vary, but the two main themes of the dissertation are discrimination and social norms. The project presented in Chapter 2 focuses on discrimination. The project presented in Chapter 4 focuses on social norms. And the project in Chapter 3 focuses on norms of discrimination and how they vary across contexts.

Chapter 2 consists of a meta-analysis of lab experiments measuring economic discrimination. The chapter, entitled 'Discrimination in the laboratory: a meta-analysis of economics experiments', was published as a paper in the European Economic Review in 2016. Discrimination has long been regarded as an important economic matter (at least since Becker, 1957), and experimental research in economics measuring discrimination has proliferated in the last 15 years or so, since the emergence of such ground-breaking studies as Fershtman and Gneezy (2001). However, the results of this experimental literature have often been contradictory, and a clear consensus has been lacking regarding the types of experimental circumstances under which particularly strong or weak discrimination should be expected to appear. Given the now-large available sample of studies, a meta-analysis is an appropriate tool for addressing this issue, particularly as it is more objective than qualitative literature review.

The database of experimental results that I harvested was sufficiently large and diverse to address various questions. In particular, I investigated whether the strength of discrimination systematically differs according to the type of identity it is based upon, the type of game-setting in which it is elicited, and the characteristics of the subjects participating in the experiment. I also investigated whether the experimental literature provided more support for taste-based or statistical discrimination, and made male-female comparisons of behaviour in experiments studying gender discrimination. To maximise the likelihood of yielding unbiased answers to these questions, two crucial design aspects of the meta-analysis were: 1) ensuring a sufficiently thorough and rigorous literature search was conducted; and 2) deciding upon and consistently applying a set of criteria to determine which of these identified studies were included in the final database. Section 2 of the paper addresses these issues in depth.

A key finding of the meta-analysis was that the strength of discrimination in economics experiments varies depending on the dimension of identity across which discrimination is measured. Strikingly, discrimination has tended to be stronger in experiments with minimal groups (that is, groups whose identities are artificially induced during the experiment, using procedures similar to those first developed by Tajfel et al, 1971) than in experiments with groups based on ethnicity, nationality or religion. The identity type upon which the weakest discrimination has tended to be found is gender – with, in fact, a slight tendency towards subjects favouring members of the opposite gender. The strongest levels of discrimination have been found between groups with identities based on social or geographical characteristics.

Regarding game-setting, I found discrimination tends to be particularly strong in the ‘third-party allocator game’ (where decision-makers are required to allocate payoffs between passive in-group and out-group members). I found evidence throughout the literature for the existence of both taste-based and statistical discrimination, but that taste-based discrimination has tended to play the dominant role. I also found that levels of discrimination are not significantly affected by whether subjects are students or non-students and that, in gender experiments, the strength of male-to-female discrimination has not tended to significantly differ from that of female-to-male discrimination.

The relative strength of discrimination between groups with artificially-induced identities may seem counterintuitive: why would ‘minimal’ groups not yield minimal levels of discrimination? I designate some discussion to this question in Chapter 2, before attempting to address it empirically in Chapter 3. Chapter 3, entitled ‘On the social inappropriateness of discrimination’, reports the results of a lab experiment investigating the relationship between discriminatory behaviour and social norms. Social norms are increasingly being presented as important determinants of economic behaviour, particularly within the experimental literature instigated by Krupka and

Weber (2013), which attempts to directly measure social norms. We hypothesised that they might help explain why discrimination is more pronounced in some contexts than others.

In this experiment, we replicate under controlled conditions the finding from the meta-analysis that discrimination is stronger across minimal groups than across groups based on nationality. We did this by measuring discrimination in two treatments which differed only in the type of group identity primed amongst subjects. In one treatment subjects were split into groups based upon whether they were British or Chinese; in the other, they were split into groups on the basis of which colour of ball they randomly drew from a bag. Consistent with the result of the meta-analysis, discrimination was found to be significantly stronger in the treatment with minimal groups.

The experiment also measured the social norms pertaining to discrimination in each treatment. Following the methodology of Krupka and Weber (2013), we asked subjects to evaluate the social appropriateness of each action decision-makers in the experiment could take. This task was incentivised such that subjects whose evaluations matched those of others would receive money; thus, evaluators were encouraged to coordinate on the social norms. As we hypothesised, discrimination was perceived to be more socially inappropriate in the treatment with national groups. Our results suggest that social norms may affect discrimination and that cross-contextual variations in the social appropriateness of discriminatory behaviour may help explain cross-contextual variations in the strength of observed discrimination. In particular, the relatively strong discrimination observed in minimal group experiments may be the result of relatively weak social norms against discrimination in such settings. In the chapter, we discuss how the findings of our experiment are consistent with a theoretical framework based closely upon that of Akerlof and Kranton (2000, 2005).

Chapter 4, entitled ‘Nudging the electorate: a field experiment on raising voter registration for the UK General Election’, reports a field experiment ran in conjunction with Oxford City Council on the effect of various behavioural interventions on voter registration rates amongst the UK electorate. The paper is impact-focused, contributing to the literature on nudging (Thaler and Sunstein, 2008) in public policy. There is also a conceptual focus, as we explore social norms as a possible mechanism for the success or failure of particular nudges.

In the study, Oxford City Council sent postcards to unregistered student voters, encouraging them to register to vote ahead of a General Election. The postcards were designed such that only the wording of the messages on them varied between treatments, allowing us to isolate the effects on registration of particular persuasion strategies. Specifically, we tested the effects of emphasising to recipients the possibility that they could be fined for not registering; of offering potential monetary rewards, in the form of entry into a lottery to win cash prizes, for recipients who registered early; and of attempting to nudge recipients towards registering through the purely psychological means of a foot-in-the-door effect (Freedman and Fraser, 1966).

We found that emphasising the possibility of being fined raised registration rates substantially. However, offering entry into the lottery had no overall effect on registration and, once the deadline for the lottery had passed, subjects exposed to this intervention were in fact significantly less likely to register than those in a control treatment. Our attempts to invoke the foot-in-the-door effect were also unsuccessful. Therefore, our study offers clear advice to policymakers on how they should and should not attempt to raise registration rates.

As we discuss in the paper, there are multiple possible explanations for why emphasising the fine works well while offering monetary incentives for registering does not. We empirically explore one of these: the effect of each intervention on social norms. In a follow-up study, we investigated the effects of being exposed to each intervention on the perceived social appropriateness of registration behaviour, again using the incentivised norm-elicitation method of Krupka and Weber (2013). Exposure to the postcard emphasising the possibility of a fine strengthened the perceived social norm that one should register to vote, while exposure to the postcard offering entry into the lottery weakened it. We suggest that this strengthening of the norm may have contributed to the success of the fine treatment, while the weakening of the norm may help explain why, as in some previous economic research (e.g. Frey and Oberholzer-Gee, 1997), the offer of monetary rewards may crowd out people's intrinsic motivation to engage in a socially beneficial act.

References

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

Akerlof, G. A., & Kranton, R. E. (2005). Identity and the Economics of Organizations. *Journal of Economic Perspectives*, 9-32.

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

Fershtman, C., & Gneezy, U. (2001). Discrimination in a segmented society: An experimental approach. *Quarterly Journal of Economics*, 351-377.

Freedman, J. L., & Fraser, S. C. (1966). Compliance without pressure: the foot-in-the-door technique. *Journal of personality and social psychology*, *4*(2), 195.

Frey, B. S., & Oberholzer-Gee, F. (1997). The cost of price incentives: An empirical analysis of motivation crowding-out. *The American economic review*, *87*(4), 746-755.

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.

Tajfel, H., et al. (1971). "Social Categorization and Intergroup Behavior." European Journal of Social Psychology 1(2): 149-177.

Thaler, R. H., & Sunstein, C. R. (2008). *Nudge: Improving Decisions About Health, Wealth, and Happiness.* Yale University Press.

Chapter Two: Discrimination in the laboratory: a meta-analysis of economics experiments

Abstract

Economists are increasingly using experiments to study and measure discrimination between groups. In a meta-analysis containing 441 results from 77 studies, we find groups significantly discriminate against each other in roughly a third of cases. Discrimination varies depending upon the type of group identity being studied: it is stronger when identity is artificially induced in the laboratory than when the subject pool is divided by ethnicity or nationality, and higher still when participants are split into socially or geographically distinct groups. In gender discrimination experiments, there is significant favouritism towards the opposite gender. There is evidence for both taste-based and statistical discrimination; tastes drive the general pattern of discrimination against out-groups, but statistical beliefs are found to affect discrimination in specific instances. Relative to all other decision-making contexts, discrimination is much stronger when participants are asked to allocate payoffs between passive in-group and out-group members. Students and non-students appear to discriminate equally. We discuss possible interpretations and implications of our findings.

JEL classifications: C92 – Laboratory Experiments (Group Behaviour); D03 – Behavioural Microeconomics: Underlying Principles

Keywords: Discrimination; Meta-analysis

1. Introduction

Meta-analysis – a commonplace technique in medical science, psychology and, to a growing extent, economics – holds advantages over literature review in terms of objectivity and analytical rigour. In recent years, the experimental economics literature appears to have reached a critical mass at which researchers are finding meta-analyses useful.[1] The benefit of these works is that, by aggregating data across a large number of experiments and exploiting natural between-study design variation, they pinpoint behavioural regularities and the variables that modify them more precisely than could be done through qualitative review.

We run a meta-analysis on the body of studies investigating discrimination in lab and lab-in-the-field experiments, a sub-literature which has certainly reached the necessary critical mass for such a venture. Economists' interest in discrimination has been strong ever since Becker (1957), and with the growth of experimental economics in the last two decades, experiments have emerged as a popular complement to survey-based econometric studies.

These experiments create a controlled environment and therefore allow much cleaner measurements of discrimination than the analysis of naturally-occurring data, avoiding such problems as omitted variable bias and reverse causality. Furthermore, by testing for a very fundamental and general form of discrimination – simply, whether subjects treat others differently depending on which group those others belong to – experimental economists can produce findings of interest not only to their own discipline but also across the social sciences. Also, through the use of incentives, experiments hold a key advantage over questionnaire-based measures of discrimination, in that they elicit revealed rather than reported discrimination.

[1] Several meta-analyses of economics experiments have been released in recent years, including: Engel (2007) – oligopoly experiments; Prante et al. (2007) – Coasean bargaining; Jones (2008) – group cooperation in prisoners' dilemmas; Hopfensitz (2009) – the effects of reference dependence and the gambler's fallacy on investment; Percoco and Nijkamp (2009) – time discounting; Weiszäcker (2010) – social learning; Engel (2011) – dictator games; Johnson and Mislin (2011) – trust games.

Psychologists had already been studying discrimination in the lab for decades, and experimental economists have drawn on their knowledge, particularly regarding the minimal group paradigm. This technique was first introduced by Tajfel et al (1971) and has spawned a huge body of experiments wherein group identity is artificially induced in the laboratory. This is often done by, in a preliminary phase of an experiment, asking subjects to state their preference for one artist over another, or to randomly draw a colour. The experimenter then splits the subject pool into groups according to their art preference, or the colour they have drawn, and makes it known to participants that the division is based on these differences. Subsequent stages of such experiments involve interaction tasks between the groups and find discrimination surprisingly (at least to the early researchers) often.

To study discrimination, experimental economists set up games such as the dictator game, the trust game or the prisoner's dilemma, and invite a subject pool segregated along the lines of a particular identity-based characteristic (or else generate this segregation with artificial groups). They make subjects aware of the group affiliation of those they interact with, and then measure how their behaviour varies according to whether individuals they are interacting with share their identity (are in-group) or do not (are out-group).

The number of economics experiments of this type has grown rapidly since the turn of the century and now encompasses substantial diversity across several dimensions. Even after omitting many papers which investigate discrimination but do not meet our inclusion criteria devised to ensure a consistent approach (see Section 2), we are left with a dataset consisting of 441 experimental results (significant and null) from 77 studies – more data than most of the other experimental economics meta-analyses have had. In order to aid the progression of this literature, it is worth taking stock of what has been found to date, particularly as casual inspection reveals non-uniformity in the results; the strength of discrimination found against out-groups varies considerably, and some experiments even find discrimination in the opposite direction, i.e. against the in-group.

The aim of this meta-analysis is both to yield broad insights on discrimination and to inform the designers of future experiments testing for it. We first investigate the average strength of discrimination across the literature. We then inquire how it tends to vary according to specific experimental characteristics.

In particular, we are interested in whether the strength of discrimination depends on the type of identity being investigated. Comparing the level of discrimination between artificial (i.e. minimal) groups and various types of natural groups (such as those based on ethnicity, nationality, religion, gender and social/geographical affiliation) is particularly interesting. One might expect 'minimal' groups to yield minimal levels of discrimination. However, it is also conceivable that artificial identity inducement confers an experimenter demand effect in favour of discrimination, or that the experimental priming of sensitive natural identities reduces subjects' desire to discriminate owing to a preference not to engage in socially unacceptable behaviour. Evidence for these possibilities, in the form of relatively strong discrimination in artificial group experiments, could have implications for the external validity of certain experiments.

A further interesting question is whether the strength of discrimination varies according to the type of decision subjects are asked to make. This has implications in terms of the real-world circumstances in which discrimination can be most expected to appear and for the generalisability of findings.

We further ask whether experiments with students reveal greater or lesser discrimination than those with non-students. This is also important for the external validity of findings, and is a question worth pursuing as some studies (e.g. Bellemare and Kroger, 2007; Anderson et al, 2013) have found students are not entirely representative of wider populations in economics experiments.

This meta-analysis also aims to shed light on the motivations behind discrimination. Some experiments have been designed specifically to distinguish between taste-based discrimination and statistical discrimination – the two models that continue to dominate the theoretical literature in economics. The taste-based model, proposed by Becker (1957), entails individuals gaining direct utility from the act of discriminating against out-groups. Meanwhile, according to theories of statistical discrimination – beginning with Arrow (1972) – individuals aim to maximise their own payoffs given their beliefs and expectations about others' characteristics and behaviour, and discrimination occurs when those beliefs and expectations vary depending on the group to which the others belong. Understanding the relative importance of these two motivations will improve the focus of future research and the design of policies aimed at combating discrimination.

Finally, we include a subsection on experiments investigating gender discrimination. Gender is unique amongst the identity types in having the same two groups in each experiment. It is therefore simple to make a clean comparison between male-to-female discrimination and female-to-male discrimination.

In summary, the meta-analysis presented below aims to address the following questions: (1) What is the general pattern of discrimination across the literature? (2) How does the level of discrimination vary according to the type of identity groups are based upon? (3) How does the level of discrimination depend upon the decision-making context? (4) Do students discriminate any more or less than non-students? (5) Does the experimental literature provide more support for taste-based or statistical theories of discrimination? (6) In gender experiments, how does male-to-female discrimination compare with female-to-male discrimination?

Our main results, presented in Section 3, are as follows. (1) We find a moderate tendency towards discrimination against the out-group, with a majority of null results across the literature. (2) The strength of discrimination against the out-group does vary according to the type of group identity subjects are divided by. It is greater when identity is artificially instilled in a subject pool than when it is divided by nationality or ethnicity – minimal groups, it seems, are not so minimal after all. Discrimination is even stronger, though, when participants are divided into socially or geographically distinct groups. (3) The extent of discrimination against the out-group also depends on the role participants are given in an experiment: when subjects are asked to allocate payoffs between inactive players belonging to the in-group and out-group, it is stronger than in any other decision-making context. (4) Students do not appear to be differently inclined towards discrimination than non-students. (5) We find evidence in support of both taste-based and statistical discrimination. Tastes appear to drive the general tendency for discrimination against the out-group, but individual studies have found beliefs to affect discrimination. (6) In gender discrimination experiments the tendency for discrimination against the out-group is reversed, as subjects demonstrate slight but significant favouritism towards the opposite gender. Discriminatory behaviour in these experiments does not differ significantly between males and females. We discuss possible interpretations of these results in depth in Section 4.

We are aware of only one other meta-study attempting to analyse the experimental discrimination literature – Balliet et al (2014)[2], who take 214 estimates of discrimination from 78 studies. There is little overlap between our samples; Balliet et al take studies from across the social sciences but their search and inclusion criteria result in most of the experimental economics literature on discrimination not being included (26 of our studies – around a third – feature in Balliet et al's sample). They exclude decision-making contexts which we consider, such as being the second mover in a sequential game or a third-party allocator. They also exclude interactions between gender groups.

The present study and that of Balliet et al can be viewed as complements. Through focusing only on economic experiments, we enhance comparability and eliminate some studies using methodological elements that may not be acceptable to some social scientists. Our focus on the economic theories of taste-based and statistical discrimination differentiates our study from Balliet et al, who investigate

[2] Although nothing approaching a full meta-analysis of the in-group-out-group literature had previously been conducted, several social psychology meta-studies have investigated specific phenomena within it. Saucier et al (2005) analysed research measuring the degrees to which subjects would help white and black people; while not finding statistically significant aggregate discrimination against black people, they showed it increased in emergency situations and cases where helping was more difficult or risky. Bettencourt et al (2001) found high-status groups exhibited more in-group bias than low-status groups. Fischer and Derham (2010) concluded discrimination in minimal group experiments was stronger in countries whose societies are considered more individualistic. Aberson et al (2000) found greater in-group bias amongst individuals with higher self-esteem. Robbins and Krueger (2005) found social projection, 'the tendency to expect similarities between oneself and others', to be stronger towards in-groups than out-groups, and that this effect was amplified with artificial groups relative to natural ones. Although interesting, many of the studies included in these meta-analyses are considerably different from those we consider – often they do not relate specifically to economic behaviour, and even if they do they may not be incentivised.

psychological theories of discrimination. Throughout our analysis we compare our results to theirs. Their paper finds a similar overall tendency for discrimination to what we do. They find the extent of discrimination not to differ significantly between settings of natural and artificial identity, but do not split natural identity into subcategories as we do. The clearest difference in results between the two studies is that Balliet et al find discrimination is stronger by decision-makers who move simultaneously than by first movers in sequential exchanges, while we do not find it significantly differs between these settings.

2. Methodology and criteria for inclusion

We chose to restrict our study to the experimental economics literature. Almost all of the economics experiments have been conducted in the last 15 years and can reasonably be expected to have followed comparable procedures, which is important in a meta-analysis. We define an economics paper as follows: it must either have been published in an economics journal or have as at least one of its authors a person trained in economics or a business-related discipline, or who has at least once held a position in an economics or business-related department. Furthermore, we exclude economics papers which, it is clear to the reader, exhibit a breach of standard experimental economics practice – most notably, deception. For inclusion, an experiment must involve interaction between individuals whose decisions determine real material payoffs for participating players. In other words, it must be incentivised.

A serious pitfall meta-analyses can face is publication bias, also named the ‘file drawer problem’. Because null results are less likely to be published than significant ones, a meta-analysis risks including a disproportionately low number of studies finding small or no effects (Rosenthal, 1979; Rothstein, 2006). This can lead to an overestimation of average effect sizes. It can also, if null results are particularly unlikely to be published when combined with certain other features of a study, result in the meta-analysis overestimating the relationship between strong effects and these features; in our case, for instance, if null results in trust games were never published but null results in other games sometimes were, we would be in danger of estimating a spuriously strong relationship between trust games and significant results. To minimise such bias, a good meta-analysis should conduct the most thorough literature search possible in order to find all applicable studies, whether published or not. Our approach was threefold. In late 2013, we conducted RePEc searches for the keywords, ‘Discrimination experiment’, ‘Identity experiment’, ‘Ingroup experiment’ and ‘Outgroup experiment’, and carefully sifted through the output for candidate studies. We then followed the references and citations of all papers identified as relevant. Finally, we checked our list of included studies against that of Balliet et al (2014); this step added one study (Spiegelman, 2012).[3] One feature of the literature we meta-analyse is that studies tend to include various different treatments, and therefore report multiple results. This may act as a further curb on publication bias – insignificant findings make their way into papers alongside more interesting significant results (indeed, it turns out the majority of results in our dataset are null).[4]

Previous meta-analyses in experimental economics such as Engel (2011) and Johnson and Mislin (2011), which focus on a single game type, are able to use the average behaviour of subjects (amount sent in the dictator or trust game) as a continuous dependent variable, with one observation and an associated standard error for each treatment. In our case, we are pooling across different game types and therefore need a way of transforming the data to make meaningful comparisons between these settings. Our variable of interest is the difference between decision-makers’ behaviour towards their in-group and their out-group, whilst all other aspects of the experimental design are held constant – in essence, the discrimination effect size. There is typically one observation per every two treatments (one in-group and one out-group treatment) for each type of player active in the given game. The exception is when a decision-maker interacts with both the in-group and the out-group in the same treatment (either by making one decision which simultaneously

[3] The Balliet et al project was not in the public domain when we embarked upon ours, and we were unaware of it. We designed our search and inclusion criteria independently of theirs. However, learning of their meta-analysis provided the perfect opportunity to test the thoroughness of our search for studies. That Balliet et al include only one study which fits our inclusion criteria but which we had not independently found suggests it is unlikely we have missed many applicable papers.

[4] The number of observations generated by a single paper varies from 1 to 24, with Chen et al (2014) providing the most.

affects both, or by playing in the same role twice), in which case a within-treatment measure of discrimination is available.[5] The ideal approach would be to record an effect size for each comparison, and we attempt to do this. Consistent with Balliet et al (2014), the measure we use is Hedges' unbiased d: the mean difference in behaviour towards the in-group and the out-group, divided by the pooled standard deviation, with a minor correction for sample size (Hedges and Olkin, 1985).

However, a substantial number of studies do not report sufficient data for us to calculate effect sizes. This is particularly the case with null results, as when a difference is not significant authors are less likely to report the test statistic from which an effect size could be derived. We sent data requests to the authors of all papers for which we could not construct the measure using information provided in the paper. After receiving data from 22 of the 36 sets of contacted authors, we ended up with effect sizes on 364 of our 441 data-points. We therefore also employ a binary dependent variable, recording simply whether, for each comparison, behaviour significantly favours the in-group over the out-group at the 5% level.[6] The effect size is the inferior dependent variable in that it restricts the sample and may lead to greater under-representation of null results; but the superior one in terms of information content.

For simplicity, we define 'discrimination' as discrimination against the out-group, and 'out-group favouritism' as discrimination against the in-group, and will use these terms hereafter. Unlike some, we make no distinction between nepotism and discrimination; any result of favouritism towards one group relative to a second can equivalently be interpreted as discrimination against the second group. We therefore conceptualise 'discrimination' (against the out-group) as something which can be

[5] For a game to meaningfully measure discrimination, and therefore for us to include it, it must be possible to unambiguously rank the decision-maker's available actions in terms of how favourable they are to the decision-maker's partner. Certain coordination games cannot be included, since whether one action is more favourable depends upon the move a partner simultaneously makes. In Appendix A, Table A.2 we list all the game types included in our sample, and explain how they measure discrimination.

[6] We also do this for out-group favouritism, recording whether or not behaviour significantly favours the out-group over the in-group at the 5% level, and run separate regressions on this. These are reported in Appendix C, Table C.1.

measured on a continuum with positive and negative values. When discussing average effect sizes, we will describe a relatively low value as indicating ‘lower’ or ‘weaker’ discrimination, even if it is driven by highly negative effect sizes (i.e. even if it is driven by instances of strong discrimination against the in-group).

For an observation to meet our inclusion criteria, there must be an in-group and out-group, clearly defined on the basis of categorisation by a discrete identity-relevant variable, such as ethnicity, gender or – as with artificial groups – the preference for a particular artist or the colour randomly drawn. There must be controlled interaction within and between the groups, and decision-makers must be aware that they are interacting with individuals belonging to their in-group or out-group. We only consider an in-group to be appropriately defined as such if every one of its members shares the same categorisation as the decision-maker on the basis of the relevant variable. For an out-group to be appropriately so-defined, every member must take a different categorisation from the decision-maker. It is not required that all members of an out-group take the same categorisation as each other. For instance, Guillen and Ji (2011) use as their two groups Australian and non-Australian. In this case, for an Australian decision-maker the Australians are the in-group and the non-Australians the out-group, but for a non-Australian the other non-Australians should not count as their in-group. We then only record the observed behaviour of the appropriately defined group, the Australians in this example. Occasionally, we are forced to make a subjective decision on what can reasonably be considered a group. For example, from Chen et al (2011), which splits its US-based sample into white and Asian students, we record the behaviour of the white ‘group’ but not that of the Asians, as we believe that in American society white people can appropriately be defined as comprising a shared ethnicity, whilst those of Asian descent comprise a mixture of ethnicities.[7] Papers such as Falk and Zender (2007) which do not have clear groups but measure each subject’s position on a scale of social distance, based on a continuous variable, are not included.

[7] There were four cases where we made such subjective decisions, all listed in Appendix A. Our main results still hold regardless of the decisions we come to in these cases.

If an experimental design splits the sample up into more than two separate groups, on the basis of a single identity-relevant variable, we record separately how each group treats each other group relative to its own. If such a paper reports that Group A does not significantly discriminate against Group B or Group C but does significantly discriminate against Groups B and C combined, we record two results of no discrimination rather than one result of discrimination; and in the main text of this paper we report our results using this approach. We do this because, although Groups B and C combined could represent a single out-group as defined above, the experiment was set up to treat them as separate out-groups. Similarly, we do not include the reported results of statistical tests run on data pooling two or more treatment pairs. These are grey areas but we have re-run our main regression results for the binary dependent variables in the case of treating every result reported in our sample as an observation: this adds 16 extra data-points and does not qualitatively change our findings.

Sufficient data must be reported for it to be clear whether there is significant discrimination in each pair of treatments (or, when applicable, single treatment); if we cannot work out whether there is discrimination in one or more treatment pair, the whole paper is omitted from the study. This is because papers are less likely to report the results of statistical tests finding no discrimination, and if we failed to include a given study's non-results our analysis would overestimate the likelihood of this particular design finding discrimination. For similar reasons, if an experiment employs a cross-cutting design, dividing its subject pool by multiple identity types, it must report whether there is discrimination on the basis of each category. For example, an experiment which segregates the subjects by both gender and ethnicity must report, for each applicable treatment pair, whether each ethnic group discriminates against each other ethnic group or not, and also whether each gender discriminates against the other or not. Otherwise, we omit the study.

Experimenters using artificial groups generally conduct tests on pooled data; rather than reporting whether Group A discriminates against Group B and vice versa, they report whether individuals across the sample pool discriminate against out-group members. This makes sense because there is no obvious reason to doubt the relationship between two artificial groups is completely symmetrical. As such, we use pooled discrimination observations for artificial group experiments. Using similar reasoning, we also admit pooled discrimination observations for experiments dividing subjects by their real-world social groups. The pooling of certain types of data might lead to an increased chance of finding discrimination in certain experiments, which is one reason why we use the size of the sample from which the result is derived as a control variable in our regression analysis.

We limit our analysis to lab and lab-in-the-field experiments; we do not include pure field experiments, in which subjects do not know they are participants in a study. We therefore do not include the large body of field experiments in which applications are sent to employers, landlords or others to test for discrimination in markets (correspondence studies).

2.1 Analytical methods

Listed in the next subsection are descriptions of the independent variables we include in our regressions. Our basic model contains role and identity type dummies, and some controls. Because our samples are not large and most variables are dummies, we regard linear probability models (LPMs) with errors corrected for heteroskedasticity as the best specifications when employing the binary dependent variables. However, we also run as robustness checks logit models, which we report in Appendix C, Table C.2. In some cases the logits drop observations, which is a major disadvantage. Their results, however, are qualitatively similar to the LPMs. When using binary dependent variables, we treat each study within the meta-analysis as providing a cluster of observations.

When analysing the continuous dependent variable, we first use standard random effects meta-analysis procedures to determine average effect sizes for our full sample and for the subsamples based on identity type. These are simply aggregate estimates for the level of discrimination in the relevant subsample; they do not control for independent variables. The procedure takes into account that each observation has an associated standard error. It weights each observation by the inverse of this standard error, thus attaching more importance to results from larger samples and with smaller standard deviations. It then follows an unweighting process, the extent of which depends upon the heterogeneity in effect sizes. The more heterogeneity there is across effect sizes, the more equal will be the weights attached to observations with small or large standard errors (Harbord and Higgins, 2008).[8]

We then apply random effects meta-regressions, which allow the inclusion of independent variables in the analysis. These models follow the same processes of weighting and unweighting observations as described in the previous paragraph, but are otherwise standard linear regressions. Whereas with the binary dependent variable we must approach discrimination and out-group favouritism separately, the meta-regression analyses both simultaneously, since the effect sizes can be positive or negative. This can be one reason why the results of the meta-regressions may differ from those of the linear probability regressions. Another can be the reduction in sample – therefore, when the results of the meta-regressions do not match those of the LPM regressions on discrimination , we present the LPMs re-run on the reduced effect-size sample, in order to determine whether the disparity is due to the change in sample or the change in analytical approach.

2.2 Independent variables

Role type dummies: We include role type dummy variables to pursue the question of how different decision-making contexts affect the extent of discrimination. The games used in this literature feature either multilateral or unilateral decision-making. When decision-making is multilateral, the outcome of the game is determined by more than one player's actions. From such situations, we identify three different role types: *First Mover* (140 observations), where one's move does not finish the game; *Second Mover* (119 observations), where one determines the final payoffs in response

[8] The random effects approach is more suitable for our purposes than the fixed effects alternative, which excludes the unweighting step; the fixed effects process assumes there to be one true effect size across all studies, while random effects allow it to vary – the latter seems more plausible in our case, as we do not assume discrimination to be a universal constant.

to the co-player(s)' actions; and *Simultaneous Mover* (66 observations), where one makes the last move of the game at the same time as one's co-player(s).

When decision-making is unilateral, the final outcome of the game is determined by one player. From these situations, we identify a further three role types. First, there is the *Dictator* (67 observations), who allocates payoffs between another player and his- or herself. Next we have third-party allocators (*Allocator*, 30). These are players who must divide a pie between two or more passive players (who, in these experiments, are members of different groups), but whose own payoff does not depend on this decision. Finally, there are players tasked with selecting a partner (from a choice of in-group and out-group participants) with whom to play a subsequent game. We label this role *Partner Chooser* (19 observations).[9]

Identity type dummies: A second set of dummy variables records which type of group identity a given experimental sample has been divided according to. We consider identity to have been artificially induced if researchers split subjects into groups that, prior to the experiment, did not exist – in the sense of group members sharing characteristics that are not also shared by members of other groups in the study – and the subjects are aware they have been split into these groups.[10] 49 studies in the meta-analysis investigate natural identity, 32 artificially generate it, while the remaining four contain both natural and artificial treatments. We have 272 observations for natural identity types and 169 for artificial. We subdivide the natural observations into six specific categories of natural identity.

First, we have 82 observations from 13 studies in which subjects are divided by *Nationality*. Next, nine studies investigate *Ethnicity*-based identity, adding 63

[9] We ran further regressions in which we categorised the role types differently. In these models, dummy variables were assigned to specific game settings, such as the trust game sender and the trust game returner. The results are reported in Appendix B.

[10] There is some inconsistency in the literature on the definition of 'minimal groups'; some authors (e.g. Chen and Chen, 2011) categorise certain artificial groups as 'near minimal'. For our purposes, we use 'minimal groups' synonymously with 'artificially created groups.' In Appendix B, we explore the effects of inducing artificial identity using different methods, and show that it seems not to matter precisely how 'minimal' the groups are.

observations. A further seven studies generate 32 observations on *Gender* identity. 21 more observations are provided by five studies in which the subjects are split by *Religion*. 13 studies use a rather different approach, dividing the subject pool into groups based on real-world social and/or geographical identity. This is done in a variety of ways: for instance, using villages (Dugar and Shahriar, 2009), colleges within universities (Banuri et al, 2012) or friendship groups (Brands and Sola, 2010). However, all such designs share the common feature that each decision-maker has a clearly distinct social and/or geographical in-group – group identity here is defined with reference to the relative frequency with which one interacts with in- and out-group members in ordinary life. The 57 observations generated by these experiments are coded under the variable *Soc/Geo Groupings*. The remaining 17 results, from four papers, deal with other types of natural identity, which cannot appropriately be fitted into the above categories. These observations relate to political identity (Abbink and Harris, 2012), disability (Gneezy et al, 2012), caste (Hoff et al, 2011) and whether farmers are private or members of cooperatives (Hopfensitz and Miguel-Florensa, 2013). We pool them under the composite variable *Natural Other*[11].[12]

Other variables: In our regressions we include as a dummy variable (*Students*) whether each observation derives from a sample consisting predominantly of students or non-students. Even if not explicitly stated, we assume experiments run at universities have at most a very small number of non-student participants. Likewise, while we accept experiments in the field may include a few student subjects, their proportion is likely to be low (unless otherwise stated). As another control, we include the size of the active decision-making sample from which a given result is derived (*Sample Size)*.

[11] The distinction between *Soc/Geo Groupings* and *Natural Other* is not arbitrary: in- and out-groups in the *Natural Other* category are not necessarily socially or geographically distinct. However, if the *Natural Other* observations are incorporated into the *Soc/Geo Groupings* category, the *Soc/Geo Groupings* coefficients do not change substantially and all other results discussed in the paper remain unaffected.

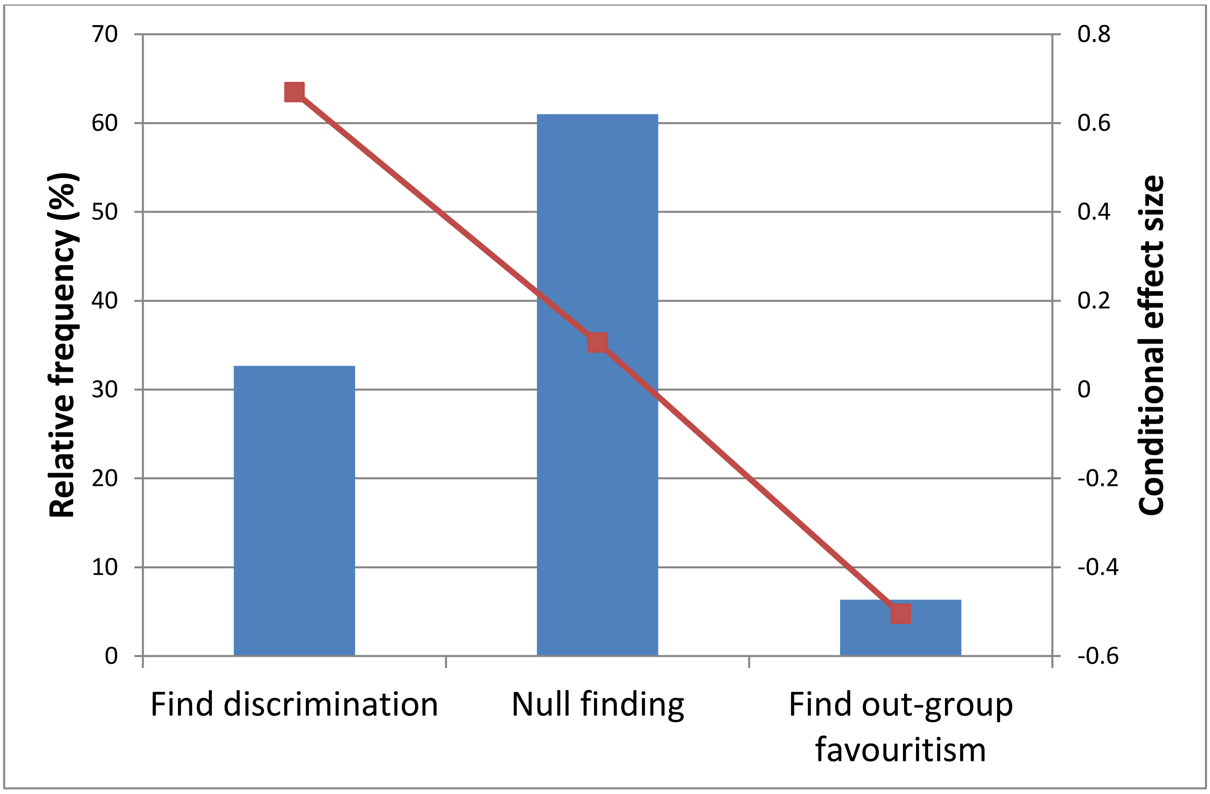
[12] Two papers provide separate results on more than one natural identity category.

3. Results

3.1 What is the general pattern of discrimination across the literature?

In total, as shown in Figure 1, there are 144 results indicating significant discrimination (32.65%), 28 indicating significant out-group favouritism (6.35%), and 269 indicating no significant discrimination or out-group favouritism (61.00%). 57 of our 77 studies record at least one result of discrimination, while only 15 record any results of out-group favouritism. 10 studies separately record results of discrimination and out-group favouritism. The general tendency, then, leans towards insignificant results, although only 15 studies consist entirely of nulls.

Figure 1: Breakdown of data-points by result type



Note: Blue bars show the percentage of observations in our dataset which find significant discrimination (at the 5% level), a null result, and significant out-group favouritism (at the 5% level). Red points show the average effect sizes for observations belonging to each category.

For the sub-sample where we are able to generate effect sizes (364 of 441 observations), the random effects meta-analysis finds an overall effect size of 0.256

(95% confidence range: 0.209 - 0.304). This can be interpreted as, on average, subjects' discriminating against out-groups by about a quarter of a standard deviation. This is not significantly different from Balliet et al (2014), who find an overall effect size of 0.32 (95% confidence range: 0.27 – 0.38). Figure 1 also displays point estimates for aggregate effect sizes, conditional on the type of result found for each observation. Observations finding significant discrimination have an average effect size of 0.67, those yielding null results have an average effect size of 0.11, and those finding significant out-group favouritism have an average effect size of -0.51; this confirms that the strength of the effect size tends to be closely related to the type of result found for a given observation.

Result 1: In general, there is limited discrimination against the out-group.

3.2 How does the level of discrimination vary according to the type of identity groups are based upon?

Table 1 displays a breakdown of our sample's observations by identity category, and the results of random effects meta-analyses run on these sub-samples. For most categories the tendency is towards null results. Only for *Soc/Geo Groupings* – which yields no results of out-group favouritism – are observations of discrimination more likely than insignificant results, and this is also the identity type with the highest average effect size. The category for which there is least discrimination and most out-group favouritism is gender; the average effect size for this sub-sample is negative.

Table 2a extends the analysis of Table 1 through the use of regressions. LPMa is a linear probability model with the dependent variable discrimination against the out-group (equal to 1 if discrimination is found, 0 otherwise). Metareg is a meta-regression with the dependent variable the discrimination effect size. In both models artificial identity and the dictator game are the benchmark categories[13]. In these

[13] By necessity, the choices of omitted categories are somewhat arbitrary – there are no variables to serve as obvious baselines for comparison. We selected *Artificial* because we regard comparisons between discrimination with artificial and natural forms of identity to be particularly interesting (as discussed in Section 1). For role type; we selected *Dictator* because it is a commonly used

regressions we test whether our identity-type variables still yield significantly different levels of discrimination after controlling for other factors. Table 2b presents the results of linear restriction tests run on the sets of dummy variables featuring in the models.

Table 1: Breakdown of data-points by result type and identity type

Category		Obs.	Find discrimi-nation (%)	Find null (%)	Find out-group favouritism (%)	Obs. with available effect sizes	Average Effect size (d) (with 95% C.I. below)
Artificial		169	42.0	55.6	2.4	150	0.365 (0.279 – 0.450)
Natural	National	82	18.3	68.3	13.4	52	0.164 (0.042 – 0.286)
	Ethnic	63	11.1	82.6	6.3	59	0.134 (0.013 – 0.255)
	Gender	32	9.4	65.6	25.0	28	-0.177 (-0.301 – -0.053)
	Religious	21	14.3	80.9	4.8	21	0.034 (-0.062 – 0.131)
	Soc/Geo Groupings	57	64.9	35.1	0.0	51	0.551 (0.432 – 0.669)
	Natural Other	17	47.1	52.9	0.0	7	-0.036 (-0.158 – 0.086)

Notes: Table 1 shows, for each identity type: the number of observations in our dataset; the percentage of these observations that find significant discrimination (at the 5% level), null results, and significant out-group favouritism (at the 5% level); the number of observations for which effect sizes are calculable; and the weighted average effect size across such observations, with associated 95% confidence intervals.

In both the linear probability models and meta-regression, the identity category linked to the strongest discrimination is social and geographical groupings. In Metareg it yields significantly higher discrimination, at the 1% level, than any of the other identity categories. In LPMa it does the same, except that the differences with *Artifical* and *Natural* Other are only significant at the 5% and 10% level respectively.

The identity category linked to the weakest discrimination is gender. Both the linear probability model and the meta-regression indicate weaker discrimination

game in experimental economics and arguably the simplest, making it a useful object for comparison.

between genders than between artificial groups, significant at the 1% level. The meta-regression also finds gender discrimination to be weaker than ethnic and national discrimination (at the 1% level), and religious discrimination (at the 10% level). However, LPMa does not find these differences to be significant.[14]

In LPMa the coefficients on the ethnic and national identity types are significantly negative at the 1% level, strongly indicating that discrimination is less likely to be observed when subject pools are split along these lines than on the basis of artificial identities. According to Metareg, however, ethnic and national identity experiments are only linked to significantly lower discrimination (i.e. less positive effect sizes) than artificial group experiments at the 10% level.

Given the inconsistency, Table 2a also reports LPMb, a linear probability model run on the reduced sample for which effect size calculation is possible. This helps to distinguish whether the losses of significance when moving from LPMa to Metareg are due to the reduction in sample or the change in measurement technique. For the comparison of national and artificial identity, the loss of significance appears to be due to the change in sample, as in LPMb the coefficient is also insignificant. The same cannot be said for *Ethnicity*, however, as the linear probability model on the reduced sample continues to report significantly less discrimination between ethnic than artificial groups at the 1% level. Doubt, therefore, is cast over the robustness of

[14] *Gender* is also found to be significant in a linear probability regression with out-group favouritism as the dependent variable, presented as LPMa1 in Appendix C, Table C.1. Our results show that gender experiments are more likely to yield observations of out-group favouritism than all other identity types except *Nationality*, with all differences significant at the 1% level. This model additionally finds experiments with socially or geographically distinct groups are less likely to provide results of out-group favouritism than those with artificial or national groups. Other identity types are not associated with significantly strong or weak out-group favouritism – however, we have few results of out-group favouritism across our sample. Where we do find significant identity type effects on out-group favouritism, they are in directions consistent with the results on discrimination – when an identity type is positively (negatively) associated with out-group favouritism, it will be negatively (positively) associated with discrimination.

Table 2a: Linear probability regressions on discrimination and meta-regressions on effect size

Dependent variable	Discrimination		d
	LPMa	LPMb	Metareg
Identity			
Ethnicity	-0.293***	-0.294***	-0.140*
	(0.067)	(0.070)	(0.079)
Religion	-0.235*	-0.256*	-0.164
	(0.131)	(0.144)	(0.125)
Nationality	-0.240***	-0.163	-0.145*
	(0.079)	(0.099)	(0.075)
Gender	-0.312***	-0.335***	-0.456***
	(0.068)	(0.072)	(0.099)
Soc/Geo Groupings	0.252**	0.229*	0.354***
	(0.099)	(0.122)	(0.089)
Natural Other	-0.056	-0.243	-0.236
	(0.165)	(0.197)	(0.192)
Role Types			
First Mover	-0.033	-0.025	0.136*
	(0.074)	(0.081)	(0.079)
Second Mover	-0.079	-0.066	0.050
	(0.065)	(0.075)	(0.085)
Simultaneous Mover	0.015	0.023	0.095
	(0.102)	(0.117)	(0.095)
Allocator	0.371***	0.408***	1.077***
	(0.094)	(0.140)	(0.155)
Partner Chooser	0.070	0.118	0.110
	(0.108)	(0.113)	(0.127)
Controls			
Students	0.005	-0.025	0.086
	(0.063)	(0.076)	(0.077)
Sample Size	6.6e^{-4}	4.9e^{-4}	6.6e^{-4}
	(4.8e^{-4})	(7.2e^{-4})	(4.8e^{-4})
Constant	0.406***	0.422***	0.144*
	(0.089)	(0.104)	(0.105)
R^2 (adjusted in Metareg)	0.201	0.196	0.240
N	441	364	364

Notes: *** p<0.01, ** p<0.05, * p<0.1; LPMa is linear probability model run on full sample, Metareg is meta-regression run on sample for which effect sizes are available, LPMb is linear probability model run on same sample as Metareg; omitted categories are Dictator (role type) and Artificial (identity); errors in LPM models are corrected for heteroskedasticity, with 77 clusters in LPMa and 67 in LPMb; standard errors in italics.

Table 2b: Linear Restriction Tests on models presented in Table 2a

	P Value on two-tailed test		
Null Hypothesis	LPMa	LPMb	Metareg
Identity			
Ethnicity = Religion	0.662	0.776	0.848
Ethnicity = Nationality	0.533	0.201	0.952
Ethnicity = Gender	0.785	0.556	0.007***
Ethnicity = Soc/Geo Groupings	<0.001***	<0.001***	<0.001***
Ethnicity = Natural Other	0.144	0.786	0.612
Religion = Nationality	0.973	0.545	0.889
Religion = Gender	0.573	0.582	0.059*
Religion = Soc/Geo Groupings	<0.001***	0.001***	<0.001***
Religion = Natural Other	0.365	0.951	0.719
Nationality = Gender	0.341	0.033**	0.006***
Nationality = Soc/Geo Groupings	<0.001***	0.005***	<0.001***
Nationality = Natural Other	0.294	0.699	0.648
Gender = Soc/Geo Groupings	<0.001***	<0.001***	<0.001***
Gender = Natural Other	0.134	0.64	0.297
Soc/Geo Groupings = Natural Other	0.080*	0.018**	0.002***
Role Types			
First Mover = Second Mover	0.444	0.552	0.175
First Mover = Simultaneous Mover	0.617	0.639	0.589
First Mover = Allocator	<0.001***	0.004***	<0.001***
First Mover = Partner Chooser	0.309	0.17	0.831
Second Mover = Simultaneous Mover	0.329	0.399	0.579
Second Mover = Allocator	<0.001***	0.001***	<0.001***
Second Mover = Partner Chooser	0.129	0.082*	0.621
Simultaneous Mover = Allocator	0.004***	0.011**	<0.001***
Simultaneous Mover = Partner Chooser	0.636	0.434	0.902
Allocator = Partner Chooser	0.021**	0.064*	<0.001***

Note: *** p<0.01, ** p<0.05, * p<0.1; LPMa is linear probability model run on full sample, Metareg is meta-regression run on sample for which effect sizes are available, LPMb is linear probability model run on same sample as Metareg.

our finding on ethnicity – although the coefficient's sign is at least weakly significant[15].[16]

Result 2: The strength of discrimination depends upon the type of group identity under investigation. It is stronger when identity is artificially induced in the laboratory than when the subject pool is divided by ethnicity or nationality, and higher still when participants are split into socially or geographically distinct groups.

3.3 How does the level of discrimination depend upon the decision-making context?

Inspection of the coefficients on role type dummies in LPMa and Metareg (Table 2a) reveals discrimination is significantly stronger when the decision-maker is a third-party allocator than when he or she is a dictator (the omitted category). Linear restriction tests (Table 2b) also show the third-party allocator role is more likely to be associated with discrimination than all the other role types, with the difference always significant at the 1% level under both models. The size of the *Allocator* coefficients in the meta-regression (1.077) is worth noting – it indicates that discrimination in games of this type tends to be very large indeed, with on average more than one standard deviation between subjects' treatment of in- and out-groups.

[15] In Table 3, we will later present a meta-regression with the number of role type dummies reduced from five to one. The purpose of this model is to investigate taste-based and statistical discrimination. However, it is worth noting that in this model with fewer independent variables, the coefficient on *Ethnicity* is found to be significantly negative at the 5% level. This improves our confidence that there is an effect. The coefficient on *Nationality* is also significant (at the 1% level) in that model.

[16] We are particularly interested in the finding that discrimination is stronger in artificial group experiments than those employing certain types of natural identity. In an attempt to gain a greater understanding of what drives discrimination between artificial groups, we ran regressions focusing on just the artificial identity sample, coding for the method experimenters used to create artificial groups. We find it makes no difference whether groups are based on preferences (such as for a particular painting) or sheer randomisation. Furthermore, we do not find that team-building exercises designed to strengthen artificial group identity significantly increase the level of discrimination. These results are all presented in greater detail in Appendix B.

The other role types do not carry significantly different effects from one another. This is at odds with Kiyonari and Yamagishi (2004), and Balliet et al (2014), who find discrimination to be stronger by simultaneous movers than first movers (Balliet et al do not investigate second movers). In an attempt to discern why our result differs from that of Balliet et al, we re-ran our analysis keeping only the observations included in their study. We found there was still no significant difference between *First Mover* and *Simultaneous Mover* (the remaining sample on which to run this regression was small; however, we also compared the aggregate effect sizes for each category and found they are very similar). This suggests the significance of the finding in Balliet et al is driven by studies outside our dataset, i.e. outside the economics literature.[17]

Result 3: Third-party allocators discriminate more than decision-makers in all other roles.

3.4 Do students discriminate any more or less than non-students?

Most decision-makers in our analysis were students. Only 101 observations, from 22 studies, are produced by in-groups not comprised (at least in their near-entirely) of university students. 31.8% of the observations for students return discrimination, while 6.8% find out-group favouritism and 61.5% are null; for non-students 35.6% find discrimination, 5.0% yield out-group favouritism and 59.4% are null. The coefficient on *Students* is not significant in any of our regressions. That experiments with students do not generate significantly different levels of discrimination than those with non-students is an interesting non-result which suggests that, in this literature, working with student samples will not generate a biased perception of the extent and magnitude of discrimination by the wider population.[18]

[17] With out-group favouritism as the dependent variable (LPMa1 in Appendix C, Table C.1), we find no significant differences at all between any role type pair.

[18] In Appendix B, we also show that the country where an experiment is run is not a significant predictor of the extent of the discrimination found.

Result 4: Discrimination does not significantly differ between students and non-students.

3.5 Does the experimental literature provide more support for taste-based or statistical theories of discrimination?

For 262 (59.4%) of our observations, as a result of the experimental design any discrimination must be taste-based, as it cannot be statistical. Statistical discrimination cannot occur when a player is making the only or last move in a game, unless the game is to be repeated, or possibly if the move is made simultaneously with others. Discrimination by trust game returners, for example, can only be taste-based, because opponents then have no control over the final outcome and beliefs about their type are therefore irrelevant.[19] All observations under the *Dictator* and *Allocator* categories preclude the possibility of statistical discrimination, as do all except seven (due to the game being repeated) in the *Second Mover* category. All observations under the *First Mover* category permit the possibility of statistical discrimination, as do most in the *Partner Chooser* category and around a third in the *Simultaneous Mover* category.[20]

In Table 3, we run a linear probability regression on discrimination and a meta-regression on the discrimination effect size, with role types re-coded into two types: one, *Taste + Statistical*, where there is scope for both taste-based and statistical discrimination, the other (the omitted category) where there is scope only for taste-based discrimination. Note that in this literature any game-role contains scope for taste-based discrimination. The coefficient on *Taste + Statistical* is positive but insignificant in both the linear probability regression (p=0.87) and the meta-regression (p=0.18). This indicates there is no significant difference in the likelihood

[19] There is a grey area to be acknowledged here. One could have a model of statistical taste-based discrimination, in which people have a taste for discrimination against a group because of beliefs they hold about its members (for instance, about how rich they are). In this paper, we do not distinguish between this and any other type of taste-based discrimination (i.e. we do not consider root motivations for taste-based discrimination).

[20] In Appendix A, Table A.2, we list which specific games permit which forms of discrimination.

of observing discrimination, or in the predicted effect size, when scope for statistical discrimination is added.[21]

This would suggest taste-based discrimination is an important driver of behaviour in these experiments and statistical discrimination is not, but we probe further by analysing the results of individual experiments. Where there is scope for statistically-motivated discrimination, by design for 66.5% of these observations it is not possible to disentangle its effects from taste-based motivations. To be able to do so, an experiment must either use belief elicitation or include a control game in which behaviour can only be taste-based – the most common case of this is adding a dictator game to extricate taste-based from statistical discrimination by trust game senders[22]. In the 60 cases that it

[21] We also ran a linear probability model on out-group favouritism, with the equivalent specification to LPMa1 in Table 3. This is reported as LPMa2 in Appendix C, Table C.1. As with discrimination, there is no significant difference in the likelihood of observing out-group favouritism when scope for statistical discrimination is added.

[22] There are no precisely standard methods for disentangling taste-based and statistical discrimination. When using a control game in which only taste-based discrimination is possible, statistical discrimination is identified if this game finds significantly weaker discrimination than the setting with scope for both types of discrimination. When using belief elicitation, statistical discrimination is confirmed if beliefs about the in-group and out-group significantly differ, and there is significant discrimination in the direction that would maximise the decision-makers' payoffs based on these beliefs; taste-based discrimination is confirmed if there is still significant discrimination after controlling for the beliefs. Some studies use regression analysis, others non-parametric tests.

Table 3: Linear probability regressions on discrimination and out-group favouritism, and meta-regression on effect size, with or without scope for statistical discrimination.[23]

Dependent variable	Discrimination	D
	LPMa	Metareg
Type of discrimination possible		
Taste + Statistical	0.009	0.071
	(0.056)	(0.053)
Identity		
Ethnicity	-0.285***	-0.189**
	(0.063)	(0.080)
Religion	-0.279**	-0.232*
	(0.134)	(0.131)
Nationality	-0.237***	-0.225***
	(0.072)	(0.079)
Gender	-0.315***	-0.545***
	(0.064)	(0.100)
Soc/Geo Groupings	0.238**	0.265***
	(0.097)	(0.093)
Natural Other	0.074	-0.306
	(0.266)	(0.202)
Controls		
Students	0.043	0.116
	(0.072)	(0.080)
Sample Size	4.5e^{-4}	-2.7e^{-4}
	(4.7e^{-3})	(4.4e^{-3})
Constant	0.350***	0.238**
	(0.089)	(0.093)
R^2 (adjusted in Metareg)	0.157	0.138 (adjusted)
N	441	364

Notes: *** p<0.01, ** p<0.05, * p<0.1; LPMa is linear probability model run on full sample, Metareg is meta-regression run on sample for which effect sizes are available; omitted categories are taste-based only (type of discrimination possible) and Artificial (identity); errors in LPMa are corrected for heteroskedasticity, with 77 clusters; standard errors in parentheses.

[23] Table 3 does not present an LPMb model because in this case we are not interested in investigating any disparities between LPMa and Metareg – the Taste + Statistical coefficient is insignificant in both models.

is possible to distinguish between discriminatory motives, the authors find significant statistical discrimination to occur in 13 cases (10 times against the out-group and three times in favour of it). Within the same sample, for given beliefs or behaviour in a game with a belief-based component, they find significant taste-based deviations from own-payoff-maximisation in 26 cases (16 times against the out-group and 10 times in favour of it). In 26 cases neither statistical nor taste-based discrimination is found at the 5% level. We list all significant findings of taste-based and statistical discrimination from experiments designed to distinguish between the two in Appendix A, Table A.3.

Although the sample is small, tastes are found to affect behaviour more often than statistical beliefs. It seems, however, that beliefs do play some role in determining discriminatory behaviour in economics experiments. We conjecture that the insignificant regression results in Table 3 may be due to the fact that beliefs can either increase or reduce discrimination. This would be because individuals have favourable beliefs about the cooperativeness of out-groups, or because unfavourable beliefs about the out-group's cooperativeness can in some cases actually lead to statistical out-group favouritism. That is, depending on the game setting, self-serving optimal behaviour can either become more or less generous in response to the perception that one's partner is relatively uncooperative. In ultimatum games, for instance, if proposers expect out-group responders to treat them less favourably than in-group responders do, the self-serving optimum is to send them relatively kind offers. This is in contrast to how first mover behaviour would work in trust games, say, where a self-serving sender will send relatively low investments to an out-group responder if it expects to be treated unfavourably by them.[24][25]

[24] We are unable to explore this empirically. We can separate games into those where favourable beliefs about a partner's cooperativeness should either increase or decrease the selfish decision-maker's cooperation towards them, but we would need data on beliefs about in-groups and out-groups to predict the direction of discrimination this should result in.

[25] In Appendix B, section B.4, we analyse how the relative strength of discrimination in experiments featuring different identity types interacts with the type of discrimination possible. We show discrimination is only significantly stronger across artificial groups than across ethnic, religious or national groups when there is no scope for statistical discrimination, while discrimination is only

Result 5: There is evidence for both taste-based and statistical discrimination. Tastes appear to drive the general tendency for discrimination against the out-group, but individual studies have found beliefs to affect discrimination.

3.6 In gender experiments, how does male-to-female discrimination compare with female-to-male discrimination?

An immediately obvious finding is that gender acts very differently from other identity types. It is the only identity category which is more likely to be associated with a bias against the in-group than against the out-group, with eight results of the former and three of the latter out of a total 32 observations. On the reduced sample, the random effects meta-analysis finds an overall discrimination effect size of -0.177 (95% confidence range: -0.301 – -0.053) for gender experiments, representing significant out-group favouritism. There is obvious intuition why gender is different from the other identity categories: it is the only case in which the effects of sexual attraction – towards the out-group more than the in-group, for most subjects – and 'chivalry' (Eckel and Grossman, 2001) can be expected.

Every experiment on gender in the meta-analysis has a symmetrical male-female design, meaning that for every estimate of discrimination by men against women there is an identical treatment measuring discrimination by women against men. This allows a very clean comparison of these two behaviours across the sample. The only three significant results in our dataset of one gender discriminating against the other are female decision-makers discriminating against males, while six of the eight significant results of one gender favouring the other are male decision-makers favouring females. However, the calculated overall effect size for female decision-makers is actually slightly more negative than for males: -0.181 (95% confidence range: -0.35 - -0.013) for females and -0.173 (95% confidence range: -0.369 – 0.024) for males, although the difference is far from significant. Note that while the effect

significantly stronger across social/geographical groups than across artificial groups when there is scope for statistical discrimination.

size indicates females significantly favour males at the 5% level, the equivalent effect for male decision-makers is only significant at the 10% level.

Result 6: There is significant out-group favouritism in gender experiments. Females significantly favour males; males favour females but the effect is only weakly significant.

4. Discussion and Conclusions

A leading result of this paper is that discrimination in economics experiments varies by the type of identity groups are based upon. It is very strong when groups are socially or geographically distinct, and is relatively weak when they are based on ethnicity or nationality. Notably, it tends to be relatively strong in experiments using artificially-induced group identities – so it can confidently be stated that minimal groups do not produce the minimal level of discrimination. At first glance, this seems surprising.

It might be that artificial group manipulations are stronger priming instruments than natural identity experiments tend to use – after all, these dedicate an entire preliminary phase of the experiment to inducing feelings of identity, which will remain at the front of subjects' minds when they are then offered the chance to discriminate. This explanation is arguably supported by the evidence of Robbins and Krueger (2005), whose meta-analysis of psychology experiments shows subjects exhibit stronger in-group projection – that is, they perceive in-group members to be particularly similar to them, relative to out-group members – when identities are artificial than when they are natural. On the other hand, we do not find that team-building exercises, which are designed specifically to strengthen artificially-induced identity and would seem to amplify priming, have a significant effect on the level of discrimination (this is consistent with the findings of Chen and Li, 2009).

Conversely, it could be argued that, for the populations studied in the literature, membership of particular ethnic and national groups does not actually instil strong identity, so that even such trivial identities as can be artificially induced have a greater effect. There is evidence that the process of globalisation has weakened national and ethnic parochialism (Buchan et al, 2009), and in recent decades youth identity in the West and increasingly elsewhere has come to define itself to a large extent upon individuals' belonging to subcultures based on fashion and music tastes – preferences drawn from choice sets which are not, indeed, so different from the apparently arbitrary minimal group painting dichotomy. However, it would seem highly complacent to draw the conclusion from our results that racism and xenophobia are not big problems in many societies.

Another explanation may be that subjects in ethnic and national identity experiments are shying away from displaying 'politically incorrect'[26] behaviour, given that racism and xenophobia are taboo in most societies today. While the link between social acceptability and discrimination has not been well explored, the prejudice literature has yielded relevant findings: that expressions of prejudice correlate with perceptions towards the social acceptability of such prejudice (e.g. Crandall et al, 2003), and furthermore that this correlation is at least partly the result of norm-compliance (e.g. Blanchard et al, 1994).

It seems unlikely that discriminating on the basis of a stated preference for Klee's paintings over Kandinsky's carries any taboo similar to ethnic or national discrimination. Indeed, some subjects may regard an artificial group situation as a game in which they belong to one of the teams, wherein the social norm actively encourages favouritism of one's own group – the sheer strangeness of the setting may even lead subjects to perceive a demand for discrimination on the part of the experimenter (see e.g. Zizzo, 2010). Concerns about social acceptability could explain also why the *Soc/Geo Groupings* category produces significantly higher discrimination than other types of natural identity. Of course, it would not be surprising if relational and geographic proximity led to a stronger sense of belonging

[26] Political correctness is defined as 'The avoidance of forms of expression or action that are perceived to exclude, marginalize, or insult groups of people who are socially disadvantaged or discriminated against' (Oxford Dictionaries).

than shared ethnicity, religion or nationality, but bear in mind too that there is arguably no taboo against favouring friends over strangers[27].

If it were shown that discrimination in economics experiments is indeed limited by concerns about social acceptability, it might cast doubt over the external applicability of such studies' findings. It is possible that if participants guess an experiment is about a type of discrimination which is taboo, it will systematically generate a lower effect than if the subjects were unaware of its purpose. On the other hand, the very same concerns about social acceptability might also limit certain types of discrimination outside the lab.

It is noteworthy that gender is the identity category producing the weakest discrimination: in fact, here the meta-analysis finds a significant amount of out-group favouritism. However, gender discrimination clearly persists in the outside world. It may be that economics experiments do not find it because they poorly reflect the conditions under which it survives beyond the lab – in particular, in the job market.

It would be interesting to see more experiments designed to directly compare the effects of different types of group identity. This meta-analysis includes just four. Dugar and Shahriar (2009), Li et al (2011) and Goette et al (2012) all compare discrimination between social/geographical groups and artificial groups, while Abbink and Harris (2012) use artificial groups and political groups (which fall under the *Natural Other* category). The results of all four studies are consistent with ours – discrimination is always lower with artificial identity. However, direct comparisons between artificial group and ethnic or national discrimination are lacking, and it would be very illuminating to see whether such studies support – and if so, whether they can explain – the findings of this meta-analysis.

What implications does our research have for future experiments on discrimination? First, using artificially induced identities as a control against which to pit the results of natural identity treatments may not be recommendable, as the

[27] This does depend upon the context, however. There are strong taboos against nepotism in certain labour-market transactions. Possibly, the experiments in this literature do not recreate such circumstances.

artificial group manipulation appears not so much to capture the minimal level of discrimination that must result from priming any type of identity in a laboratory as to in fact often go beyond that.

Regarding role type, we find discrimination by third-party allocators is much stronger than by participants in any other game setting. If social acceptability does indeed limit discrimination, this is a counterintuitive result, as the allocator role essentially invites subjects to overtly and consciously favour one group over another and therefore seems to be the one that most obviously telegraphs the purpose of this type of experiment. One possibility is that the role carries an experimenter demand effect – whereby subjects feel they are encouraged to discriminate – or even an action bias effect, if the equal split feels like a default non-move. Another relevant factor may be that the third-party allocator is unique amongst our role types in the decision-maker's payoff being entirely disconnected from the extent to which they discriminate. In any case, experimenters should bear in mind that because they are more likely to identify significant discrimination when they employ the allocator role, they should be less confident that the same groups will discriminate against each other in different contexts.

We find the strength of discrimination does not significantly differ between student and non-student subject pools. This suggests – unlike in the context of social preferences (e.g. Bellemare and Kroger, 2007; Anderson et al, 2013) – student subjects are not a generally unrepresentative sample for questions relating to discrimination. However, we do not exclude the possibility that they are unrepresentative in specific instances, or within particular societies.

There is scope for more experimental research investigating taste-based and statistical discrimination. We show both are relevant, and the two types manifest themselves to different extents in different contexts. However, relatively few experiments have been designed to distinguish between taste-based and statistical discrimination, and more could be known about the mechanisms underlying them.

As a final observation, there is a great deal of variation in the findings of the experimental economics discrimination literature. Our analysis can explain some of it, but our LPM regressions typically have R^2 statistics below 0.2, and the meta-regressions' Adjusted R^{2}s are rarely above 0.35. As might be expected, discrimination does seem to vary idiosyncratically and is not easy to predict. The results of natural identity experiments do not seem very generalizable – they probably reflect more the characteristics of the specific groups under investigation, and the relationships between them, than aspects of the experimental design. Whilst a drawback for some research questions, this also means there is a great deal of scope for future experimental studies aimed at measuring the levels of discrimination within subject pools of specific interest. Furthermore, given the potential concerns we raise about experimenter demand effects and the external validity of lab experiments on discrimination, the important role of field experiments should be emphasised. Subjects in such studies are unaware they are being observed by experimenters and their behaviour can therefore not be influenced by the fact. Field experiments can test the generalisability of lab findings on discrimination.

Acknowledgements

Many thanks to Abigail Barr and Daniele Nosenzo for their support and input. I am very grateful to all authors who sent data for this project. I received helpful comments at presentations in Nottingham, Duisburg and Warwick from, amongst others, Swee-Hoon Chuah, Simon Gächter, Gianluca Grimalda, Heike Hennig-Schmidt and Gari Walkowitz. Thanks to the Network for Integrated Behavioural Science for financial support.

References

Papers included in the meta-analysis:

Abbink, K. and D. Harris (2012). In-group favouritism and out-group discrimination in naturally occurring groups, Technical report, Mimeo, Monash University.

Ahmed, A. M. (2010) "What is in a surname? The role of ethnicity in economic decision making." Applied Economics 42.21: 2715-2723.

Ahmed, A. M. (2007). "Group identity, social distance and intergroup bias." Journal of Economic Psychology 28(3): 324-337.

Banuri, S., et al. (2012). "Deconstructing nepotism." Available at SSRN 2248187.

Bauernschuster, S., et al. (2009). Social identity, competition, and finance: a laboratory experiment, Jena economic research papers.

Ben-Ner, A., et al. (2004). "Share and share alike? Gender-pairing, personality, and cognitive ability as determinants of giving." Journal of Economic Psychology 25(5): 581-589.

Bernhard, H., et al. (2006). "Group affiliation and altruistic norm enforcement." American Economic Review 96(2): 217-221.

Binzel, C. and D. Fehr (2013). "Social distance and trust: Experimental evidence from a slum in Cairo." Journal of Development Economics 103: 99-106.

Boarini, R., et al. (2009). "Interpersonal comparisons of utility in bargaining: evidence from a transcontinental ultimatum game." Theory and decision 67(4): 341-373.

Bouckaert, J. and G. Dhaene (2004). "Inter-ethnic trust and reciprocity: results of an experiment with small businessmen." European Journal of Political Economy 20(4): 869-886.

Brandts, J. and C. Sola (2010). "Personal relations and their effect on behavior in an organizational setting: An experimental study." Journal of Economic Behavior & Organization 73(2): 246-253.

Buchan, N. R., et al. (2006). "Let's get personal: An international examination of the influence of communication, culture and social distance on other regarding preferences." Journal of Economic Behavior & Organization 60(3): 373-398.

Büchner, S. and D. A. Dittrich (2002). I will survive!--Gender discrimination in a household saving decisions experiment, Max Planck Institute of Economics, Strategic Interaction Group.

Burns, J. (2004). "Race and trust in post-Apartheid South Africa." University of Cape Town, Centre for Social Science Research working paper 78.

Butler, J. V. (2014). Trust, Truth, Status and Identity: An Experimental Inquiry. The BE Journal of Theoretical Economics, 14(1).

Carpenter, J. and J. C. Cardenas (2011). "An Intercultural Examination of Cooperation in the Commons." Journal of Conflict Resolution 55(4): 632-651.

Chakravarty, S. and M. A. Fonseca (2013). Discrimination via Exclusion: An Experiment on Group Identity and Club Goods. University of Exeter Economics Department Discussion Papers Series. 13/02.

Charness, G., et al. (2007). "Individual behavior and group membership." The American Economic Review: 1340-1352.

Chen, R. and Y. Chen (2011). "The potential of social identity for equilibrium selection." The American Economic Review 101(6): 2562-2589.

Chen, Y. and S. X. Li (2009). "Group Identity and Social Preferences." American Economic Review 99(1): 431-457.

Chen, Y., et al (2014). Which hat to wear? Impact of natural identities on coordination and cooperation. Games and Economic Behavior, 84, 58-86.

Chuah, S. H., et al. (2007). "Do cultures clash? Evidence from cross-national ultimatum game, experiments." Journal of Economic Behavior & Organization 64(1): 35-48.

Chuah, S.-H., et al. (2013). "Fractionalization and trust in India: A field-experiment." Economics Letters 119(2): 191-194.

Costard, J. and F. Bolle (2011). Solidarity, responsibility and group identity, Discussion paper//European University Viadrina, Department of Business Administration and Economics.

Currarini, S. and F. Menge (2012). "Identity, homophily and in-group bias." services.bepress.com.

Daskalova, V. (2012). "Discrimination, Social Identity, and Coordination: An Experiment." webspace.qmul.ac.uk.

Delavande, A. and B. Zafar (2011). "Stereotypes and Madrassas." Federal Bank of New York Staff Reports 501.

Der Merwe, V., et al. (2008). "WHAT'S IN A NAME? RACIAL IDENTITY AND ALTRUISM IN POST-APARTHEID SOUTH AFRICA." South African Journal of Economics 76(2): 266-275.

Dugar, S. and Q. Shahriar (2010). "Group identity and the moral hazard problem: Evidence from the field." San Diego State University, Department of Economics Working Papers.

Eckel, C. C. and P. J. Grossman (2001). "Chivalry and solidarity in ultimatum games." Economic Inquiry 39(2): 171-188.

Etang, A., et al. (2011). "Does trust extend beyond the village? Experimental trust and social distance in Cameroon." Experimental economics 14(1): 15-35.

Etang, A., et al. (2011). "Trust and rosca membership in rural cameroon." Journal of International Development 23(4): 461-475.

Fehr, E., et al. (2013). "The development of egalitarianism, altruism, spite and parochialism in childhood and adolescence." European Economic Review 64: 369-383.

Ferraro, P. J. and R. G. Cummings (2007). "Cultural diversity, discrimination, and economic outcomes: an experimental analysis." Economic Inquiry 45(2): 217-232.

Fershtman, C., et al. (2005). "Discrimination and nepotism: The efficiency of the anonymity rule." Journal of Legal Studies 34(2): 371-394.

Fiedler, M., et al. (2011). "Social distance in a virtual world experiment." Games and Economic Behavior 72(2): 400-426.

Filippin, A. and F. Guala (2013). "Costless discrimination and unequal achievements in an experimental tournament." Experimental economics 16(3): 285-305.

Finocchiaro Castro, M. (2008). "Where are you from? Cultural differences in public good experiments." The Journal of Socio-Economics 37(6): 2319-2329.

Fong, C. M. and E. F. Luttmer (2011). "Do fairness and race matter in generosity? Evidence from a nationally representative charity experiment." Journal of Public Economics 95(5): 372-394.

Friesen, J., et al. (2012). "Ethnic identity and discrimination among children." Journal of Economic Psychology 33(6): 1156-1169.

Georg, S., et al. (2008). Distributive fairness in an intercultural ultimatum game, Jena economic research papers.

Gneezy, U., et al. (2012). Toward an understanding of why people discriminate: evidence from a series of natural field experiments, National Bureau of Economic Research.

Goette, L., et al. (2006). "The impact of group membership on cooperation and norm enforcement: Evidence using random assignment to real social groups." American Economic Review 96(2): 212-216.

Goette, L., et al. (2012). "The Impact of Social Ties on Group Interactions: Evidence from Minimal Groups and Randomly Assigned Real Groups." American Economic Journal-Microeconomics 4(1): 101-115.

Goette, L., et al. (2012). "Competition Between Organizational Groups: Its Impact on Altruistic and Antisocial Motivations." Management Science 58(5): 948-960.

Grossman, P. J. and M. Komai (2008). Leadership and Gender: An Experiment. repository.stcloudstate.edu. (accessed 14/11/13)

Guala, F., et al. (2013). "Group membership, team preferences, and expectations." Journal of Economic Behavior & Organization 86: 183-190.

Guillen, P. and D. Ji (2011). "Trust, discrimination and acculturation: Experimental evidence on Asian international and Australian domestic university students." The Journal of Socio-Economics 40(5): 594-608.

Gupta, G., et al. (2013). Religion, Minority Status and Trust: Evidence from a Field Experiment, Monash University, Department of Economics.

Guth, W., et al. (2005). "The effect of group identity in an investment game." Papers on Strategic Interaction.

Guth, W., et al. (2009). "Determinants of in-group bias: Is group affiliation mediated by guilt-aversion?" Journal of Economic Psychology 30(5): 814-827.

Haile, D., et al. (2008). "Cross-racial envy and underinvestment in South African partnerships." Cambridge Journal of Economics 32(5): 703-724.

Hargreaves Heap, S. P. and D. J. Zizzo (2009). "The value of groups." The American Economic Review: 295-323.

Harris, D., et al. (2009). Two's Company, Three's a Group: The impact of group identity and group size on in-group favouritism, CeDEx discussion paper series.

Hennig-Schmidt, H., et al. (2007). "Actions and Beliefs in a Trilateral Trust Game Involving Germans, Israelis and Palestinians." Unpublished manuscript.

Hoff, K., et al. (2011). "Caste and Punishment: the Legacy of Caste Culture in Norm Enforcement*." The Economic Journal 121(556): F449-F475.

Hopfensitz, A. and P. Miquel-Florensa (2013). Public good contributions among coffee farmers in costa rica: Cooperativists and private market participants, Toulouse School of Economics (TSE).

Houser, D. and D. Schunk (2009). "Social environments with competitive pressure: Gender effects in the decisions of German schoolchildren." Journal of Economic Psychology 30(4): 634-641.

Ioannou, C. A., et al. (2013). "Group Payoffs As Public Signals." christosaioannou.com. (accessed 10/12/2013)

Johansson-Stenman, O., et al. (2009). "Trust and Religion: Experimental Evidence from Rural Bangladesh." Economica 76(303): 462-485.

Kim, B.-Y., et al. (2013). Do institutions affect social preferences? Evidence from divided Korea, IZA Discussion Paper.

Lankau, M., et al. (2012). Cooperation preferences in the provision of public goods: An experimental study on the effects of social identity, Discussion Papers, Center for European Governance and Economic Development Research.

Li, S. X., et al. (2011). "Group identity in markets." International Journal of Industrial Organization 29(1): 104-115.

Masella, P., et al. (2014). Incentives and group identity. Games and Economic Behavior, 86, 12-25.

McLeish, K. N. and R. J. Oxoby, (2007). Identity, cooperation, and punishment, IZA Discussion Papers, No. 2572

Morita, H. and M. Servátka (2013). Group identity and relation-specific investment: An experimental investigation. European Economic Review, 2013, Vol. 58 (February), pp. 95-109

Netzer, R. J. and M. Sutter (2009). Intercultural trust. An experiment in Austria and Japan, Working Papers in Economics and Statistics.

Ortmann, A. and L. K. Tichy (1999). "Gender differences in the laboratory: evidence from prisoner's dilemma games." Journal of Economic Behavior & Organization 39(3): 327-339.

Pecenka, C. J. and G. Kundhlande (2013). "Theft in South Africa: An Experiment to Examine the Influence of Racial Identity and Inequality." The Journal of Development Studies 49(5): 737-753.

Ploner, M. and I. Soraperra (2004). Groups and Social Norms in the Economic Context: A Preliminary Experimental Investigation, Cognitive and Experimental Economics Laboratory, Department of Economics, University of Trento, Italia.

Ruffle, B. J. and R. Sosis (2006). "Cooperation and the in-group-out-group bias: A field test on Israeli kibbutz members and city residents." Journal of Economic Behavior & Organization 60(2): 147-163.

Shahriar, Q. (2011). "Identity In A Second-Price Sealed Bid Auction: An Experimental Investigation." The Manchester School 79(1): 159-170.

Slonim, R. and P. Guillen (2010). "Gender selection discrimination: Evidence from a Trust game." Journal of Economic Behavior & Organization 76(2): 385-405.

Solnick, S. J. (2001). "Gender differences in the ultimatum game." Economic Inquiry 39(2): 189-200.

Spiegelman, E. "«C'EST CE QUE JE VOUS DIS»: ESSAIS SUR L'ANALYSE ÉCONOMIQUE DE LA COMMUNICATION INTERPERSONNELLE." (2012).

Tremewan, J. (2010). "Group Identity and Coalition Formation: Experiments in one-shot and repeated games." webmeets.com.

Walkowitz, G., et al. (2004). Experimenting over a Long Distance: A method to facilitate intercultural experiments, Bonn econ discussion papers.

Willinger, M., et al. (2003). "A comparison of trust and reciprocity between France and Germany: Experimental investigation based on the investment game." Journal of Economic Psychology 24(4): 447-466.

Wu, F. (2009). "Cultural Affinity in International Joint Ventures-An Experimental Study." eale.nl conference paper.

Zizzo, D. J. (2011). "You are not in my boat: common fate and discrimination against outgroup members." International Review of Economics 58(1): 91-103.

Other sources cited:

Aberson, C. L., et al. (2000). "Ingroup bias and self-esteem: A meta-analysis." Personality and Social Psychology Review 4(2): 157-173.

Alesina, A., et al. (2003). "Fractionalization." Journal of Economic growth 8(2): 155-194.

Anderson, J., et al. (2013). "Self-selection and variations in the laboratory measurement of other-regarding preferences across subject pools: evidence from one college student and two adult samples." Experimental economics 16(2): 170-189.

Arrow, K. (1972). ‘Some mathematical models of race discrimination in the labor market’, in (A.H. Pascal, ed.), Racial Discrimination in Economic Life, pp. 187–204, Lexington, MA: D.C. Heath.

Balliet, D., et al. (2014). “Ingroup Favoritism in Cooperation: A Meta-Analysis.” Psychological Bulletin

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

Bellemare, C. and S. Kröger (2007). "On representative social capital." European Economic Review 51(1): 183-202.

Bettencourt, B., et al. (2001). "Status differences and in-group bias: a meta-analytic examination of the effects of status stability, status legitimacy, and group permeability." Psychological bulletin 127(4): 520.

Blanchard, F. A., et al. (1994). "Condemning and condoning racism: A social context approach to interracial settings." Journal of Applied Psychology 79(6): 993.

Buchan, N. R., et al. (2009). "Globalization and human cooperation." Proceedings of the National Academy of Sciences 106(11): 4138-4142.

Crandall, C. S., et al. (2002). "Social norms and the expression and suppression of prejudice: the struggle for internalization." Journal of personality and social psychology 82(3): 359.

Engel, C. (2007). "How much collusion? A meta-analysis of oligopoly experiments." Journal of Competition Law and Economics 3(4): 491-549.

Engel, C. (2011). "Dictator games: a meta study." Experimental Economics 14(4): 583-610.

Falk, A. and C. Zehnder (2007). Discrimination and in-group favoritism in a citywide trust experiment, IZA Discussion Papers.

Fischer, R. D., C. (2010). "Is the minimal group paradigm culture dependent? A cross-cultural multi-level analysis." http://www.psychology.org.au/ext/iaccp2010/saturday-10-july/6/0830/fischer-r.pdf. (accessed 5/11/13)

Harbord, R. M., and J. P. Higgins. "Meta-regression in Stata." Meta 8.4 (2008): 493-519.

Hedges, L. V., & Olkin, I. (1985). Statistical methods for meta-analysis. Orlando, FL: Academic Press.

Hopfensitz, A. (2009). "Previous outcomes and reference dependence: A meta study of repeated investment tasks with and without restricted feedback." mpra.ub.uni-muenchen.de. (accessed 10/11/13)

Johnson, N. D. and A. A. Mislin (2011). "Trust games: A meta-analysis." Journal of Economic Psychology 32(5): 865-889.

Jones, G. (2008). "Are smarter groups more cooperative? Evidence from prisoner's dilemma experiments, 1959–2003." Journal of Economic Behavior & Organization 68(3): 489-497.

Kiyonari, T., & Yamagishi, T. (2004). In-group cooperation and the social exchange heuristic. In R. Suleiman, D. V. Budescu, I. Fischer, & D. M. Messick (Eds.) Contemporary psychological research on social dilemmas (pp. 269-286). New York, NY: Cambridge University Press.

Oxford Dictionaries, definition of 'political correctness'. http://www.oxforddictionaries.com/definition/english/political-correctness (accessed 1/3/15)

Percoco, M. and P. Nijkamp (2009). "Estimating individual rates of discount: a meta-analysis." Applied Economics Letters 16(12): 1235-1239.

Prante, T., et al. (2007). "Evaluating coasean bargaining experiments with meta-analysis." Economics Bulletin 3(68): 1-7.

Robbins, J. M. and J. I. Krueger (2005). "Social projection to ingroups and outgroups: A review and meta-analysis." Personality and Social Psychology Review 9(1): 32-47.

Rosenthal, R. (1979). "The file drawer problem and tolerance for null results." Psychological bulletin 86(3): 638.

Rothstein, H. R., et al. (2006). Publication bias in meta-analysis: Prevention, assessment and adjustments, John Wiley & Sons.

Saucier, D. A., et al. (2005). "Differences in helping whites and blacks: A meta-analysis." Personality and Social Psychology Review 9(1): 2-16.

Tajfel, H., et al. (1971). "Social Categorization and Intergroup Behavior." European Journal of Social Psychology 1(2): 149-177.

Weizsäcker, G. (2010). "Do we follow others when we should? A simple test of rational expectations." The American Economic Review: 2340-2360.

Zizzo, D. J. (2010) "Experimenter demand effects in economic experiments." Experimental Economics 13.1: 75-98.

Appendix A

Table A.1: Subjective decisions on appropriately defined groups

Study	Notes
Burns, 2004	We consider ‘coloured’ to be an appropriate ethnic group in South Africa (as defined in comparison to ‘white’ and ‘black’).
Chen et al, 2011	We do not consider ‘Asian’ to be an appropriate ethnic group in the USA.
Ferraro and Cummings, 2007	We consider ‘Hispanic’ to be an appropriate ethnic group in the USA. (Justification, relative to ‘Asian’: Hispanic people in the USA share a more unified culture than those of Asian descent; they are descended from more linguistically homogeneous peoples than Asians)
Friesen et al, 2012	We do not consider ‘East Asian’ and ‘South Asian’ to be appropriate ethnic groups in Canada.

Table A.2: Game types in meta-analysis and how they measure discrimination

Role Type	How discrimination is measured	Type of discrimination possible
Trust Game Returner	Difference in proportion of amount received from the sender that is returned*, between in-group and out-group matching. *NB: Ploner and Soraperra (2004) use Indirect Trust Game, where the amount returned is not given to the sender but a group member of theirs	Taste-based only
Agent in Principal-Agent Game (Masella et al, 2012)	Difference in amount sent to principal, between in-group and out-group matching.	Taste-based only
Dictator; Proposer in Unilateral Power Game (Zizzo, 2011)	Difference in amount sent to recipient, between in-group and out-group matching. (NB: Buchner and Dittrich (2002) use a saving game where one player leaves the game early and decides how much to leave their partner. This decision is the equivalent of that faced by a dictator)	Taste-based only
Allocator	Difference in amount allocated to in-group and out-group member.	Taste-based only
Responder in Ultimatum Game, Hold-up Game (Morita and Servatka, 2013)	Difference in likelihood of rejecting an offer, controlling for its size, between in-group and out-group matching.	Taste-based only
Responder in Proposer-Responder Game (McLeish and Oxoby, 2007)	Difference in amount by which proposer's payoff is reduced, controlling for amount offered by proposer, between in-group and out-group matching.	Taste-based only
Responder in Proposer-Responder Game (Chen and Li, 2009; Currarini and Mengel, 2012)	Difference in rate of choosing more other-regarding response, between in-group and out-group matching.	Taste-based only
Third-party punisher	Difference in punishment level, controlling for behaviour of punishee, between in-group and out-group matching.	Taste-based only
One-shot Prisoner's Dilemma	Difference in rate of cooperation, between in-group and out-group matching.	Taste-based only
Trader in market games	For bidders: difference in price offered, between in-group and out-group matching. For sellers: difference in price accepted, between in-group and out-group matching.	Taste-based only for sellers in one-shot interactions; taste-based and statistical for bidders in one-shot interactions, and for all players in repeated games.
Public Goods Game	Difference in contribution level, between in-group and out-group matching.	Taste-based only in one-shot games (Hopfensitz, 2013);

		taste-based and statistical in repeated games.
Partner-Choosing Role	Difference in rate of choosing in-group partner and out-group partner.	Taste-based only if chosen partner does not become active decision-maker in subsequent games; taste-based and statistical if they do.
Repeated Common Pool Withdrawal Game (Carpenter and Cardenas, 2011)	Difference in withdrawal level, between in-group and out-group matching.	Taste-based and statistical.
Minimal Effort Game (Chen and Chen, 2011)	Difference in effort level, between in-group and out-group matching.	Taste-based and statistical.
Trust Game Sender; Principal in Principal-Agent Game (Masella et al, 2012); First Mover in Hold-up Game (Morita and Servatka, 2013)	With continuous action space: difference in amount sent, between in-group and out-group matching. With binary action space: difference in rate of choosing to trust, between in-group and out-group matching.	Taste-based and statistical.
Investor in Investment Game (Wu, 2009)	Difference in amount invested in manager's project, between in-group and out-group matching.	Taste-based and statistical.
Ultimatum Game Proposer; Second Mover in Hold-up Game (Morita and Servatka, 2013); First Mover in Proposer-Responder Game (McLeish and Oxoby, 2007)	Difference in amount offered, between in-group and out-group matching.	Taste-based and statistical.
Proposer in Proposer-Responder Game (Chen and Li, 2009; Currarini and Mengel, 2012)	Difference in rate of choosing more other-regarding first move, between in-group and out-group matching.	Taste-based and statistical.
Nash Demand Game (Ruffle and Sosis, 2006; Zizzo, 2011)	Difference in amount claimed, between in-group and out-group matching	Taste-based and statistical.
Stag Hunt	Difference in rate of choosing hawkish strategy, between in-group and out-group matching.	Taste-based and statistical.

Table A.3: List of significant results from studies designed to distinguish between taste-based and statistical discrimination

Result	Paper	Role	Groups
Taste-based and statistical discrimination	Banuri et al (2012)	Partner-choosing role	Colleges within university (Free Nepotism treatment)
	Bernhard et al (2006)	Dictator	Tribes
	Binzel and Fehr (2013)	Trust Game Sender	Social groups
	Currarini and Mengel (2012)	Partner-choosing role	Artificial (comparison of COORD, ENDO and LOWB treatments)
	Etang et al (2011a)	Trust Game Sender	Villages
Taste-based discrimination only	Burns (2004)	Trust Game Sender	Ethnic (coloured in-group, black out-group)
		Trust Game Sender	Ethnic (coloured in-group, white out-group)
	Chuah et al (2013)	Trust Game Sender	Religious (Hindu in-group, Muslim out-group)
		Trust Game Sender	Religious (Muslim in-group, Hindu out-group)
	Ferraro and Cummings (2007)	Ultimatum Game Proposer	Ethnic (Hispanic in-group, Navajo out-group)
	Guillen and Ji (2011)	Trust Game Sender	National (Australian in-group, non-Australian out-group)
	Kim et al (2013)	Trust Game Sender	National (North Korean in-group, South Korean out-group – sample 1)
		Trust Game Sender	National (North Korean in-group, South Korean out-group – sample 2)
	McLeish and Oxoby (2007)	First Mover in Proposer-Responder Game	Artificial (OP treatment)
		First Mover in Proposer-Responder Game	Artificial (NO treatment)
	Ruffle and Sosis (2006)	Nash Demand Game	Social/geographical (Kibbutz in-group, city out-group)
Statistical discrimination only	Banuri et al (2012)	Partner-choosing role	Colleges within university (Costly Nepotism treatment)
	Boarini et al (2009)	Ultimatum Game Proposer	National (French in-group, Indian out-group)
	Chen and Chen (2011)	Minimal Effort Game	Artificial (Enhanced treatment)
	Haile et al (2008)	Trust Game Sender	Ethnic (white in-group, black out-group)
	Masella et al (2014)	Principal in Principal-Agent Game	Artificial
Taste-based out-group favouritism only	Burns (2004)	Trust Game Sender	Ethnic (white in-group, black out-group)
		Trust Game Sender	Ethnic (black in-group, white out-group)
		Trust Game Sender	Ethnic (black in-group, coloured out-group)
	Hennig-Schmidt et al (2007)	Trust Game Sender	National (Israeli in-group, Palestinian out-group)
		Trust Game Sender	National (Palestinian in-group, Israeli out-group)
	Kim et al (2013)	Trust Game Sender	National (South Korean in-group, North Korean out-group – sample 1)
		Trust Game Sender	National (South Korean in-group, North Korean out-group – sample 2)
	Slonim and Guillen (2010)	Trust Game Sender	Gender (male in-group, Gender/Ability Selection treatment)
		Trust Game Sender	Gender (male in-group, No Selection treatment)
		Partner-choosing role	Gender (male in-group, Trust Game treatment)
Statistical out-group favouritism only	Boarini et al (2009)	Ultimatum Game Proposer	National (Indian in-group, French out-group)
	Hennig-Schmidt et al (2007)	Trust Game Sender	National (German in-group, Palestinian out-group)
	Slonim and Guillen (2010)	Partner-choosing role	Gender (female in-group, Trust Game treatment)

Appendix B: Further Results

B.1 Further analysis of role types

In this section we investigate the effects on discrimination of using specific game types. We recode the role type variables, assigning dummies to specific game settings in the following way. Trust games and similar principal-agent games provide two roles: senders (*TG Sender*, 98 observations) and returners (*TG Returner*, 81). The next most common role type is the *Dictator* (68). Prisoner's dilemmas, public goods games, and common pool withdrawal games are all social dilemmas, and are coded under a single category (*Social Dilemma*, 58). Next we have third-party allocators (*Allocator*, 33). Ultimatum games and similar bargaining settings are grouped together and split into two role types: first movers (*Proposer*, 31) and second movers (*Responder*, 27). Treating *Dictator* as the omitted category in our regressions, we form a set of binary independent variables from the other six role types, plus the additional variable *Game Other* (45 observations) into which are placed the remaining game settings that we did not think could be adequately categorised.[28]

Table C.3a in Appendix C displays the output of regressions incorporating these variables. These regressions are the equivalent of those presented in Table 2, the only change being the recoding of the role type variables. As above, LPMa1 and Metareg1 show discrimination to be significantly stronger when the decision-maker is a third-party allocator than when he or she is a dictator (the omitted category). Linear restriction tests (Table C.3b) also show the third-party allocator role is more likely to be associated with discrimination than all the other role types, with the difference always significant at the 1% level under both models. Again, the other role types do not consistently carry significantly different effects from one another. This is at odds with the analysis of Balliet et al (2014), who find discrimination is stronger by trust

[28] Specifically, the Game Other category consists of players in the following settings: unstructured bargaining games; the battle of the sexes; coordination games; indirect trust games; market-trading games; minimal effort games; Nash Demand games; partner-choosing situations; saving games; stag hunts; and third-party punishment games. Several of these could have been coded under a standalone category – coordination games and variants – but there would only be eight observations in such a category.

game senders than by dictators, and stronger still in social dilemmas. With out-group favouritism as the dependent variable (LPMa2), the only significant differences between role types are that proposers are less likely to engage in out-group favouritism than dictators, trust game senders, trust game returners and subjects in the *Game Other* category.

Result A1: We do not find strong effects associated with such specific role types as the trust game sender or returner, players in social dilemmas, or bargaining game proposers or responders.

B.2 Does the strength of discrimination in artificial group experiments depend on the method used to induce identity?

The way in which identity is artificially instilled in subjects varies from experiment to experiment. However, we can identity two broad categories of artificial group creation. One follows the original Tajfel et al (1971) process of allowing subjects to self-select into groups. Typically this involves asking participants to choose a preference between the art of Klee and Kandinsky, although some studies elicit preferences on other choice sets, such as favourite colours. We code these observations under *Preferences*. The other main category gives subjects no control over which group they belong to. In such cases they are simply randomly assigned and labelled as belonging to, for instance, the 'red' or 'blue' group. We code these manipulations as *Labelling*. Occasionally, a different type of identity inducement is done – for example, groups can be based on subjects' tendency to overestimate or underestimate the number of dots on a screen (Guala et al, 2013; Ioannou et al, 2013), or by the time at which they undertake a particular task (Ahmed, 2007). These cases we code as *Other Method*.

Another way artificial group manipulations vary is by whether they contain additional stages in which group members interact, between being placed into groups and before the task upon which discrimination is measured. These stages often involve games in which group members must work together to earn monetary rewards, although on some occasions they merely interact non-strategically as a result of being permitted to converse electronically. Such stages are introduced as a mechanism to strengthen artificial group identities. We code their presence in studies under *Team Building*.

In order to test how these different procedures affect the extent of discrimination, we run LPM and meta-regressions on our sub-sample of observations for which identity is artificial. These are presented in Appendix C, Table C.4. We find there is no significant difference between whether groups are self-selected or randomly selected. Also, while the coefficients are in the direction of strengthening discrimination, we find the effect of team-building exercises not to be significant. From these results, we infer that the precise form of identity inducement is not crucial to the outcome of artificial group experiments. This is consistent with the findings of Chen and Li (2009), whose experiment addresses these questions.

Result A2: The strength of discrimination in artificial group experiments does not depend significantly on the method used to induce identity.

B.3 Can country-level variables explain discrimination?

Our meta-analysis encompasses geographical diversity, with data from 31 countries. Including cases where the out-group was located in a different country, 169 results from 34 studies come from Europe, 116 observations from 22 studies are from North America, 85 results from 17 studies are from Asia, 37 observations from seven studies come from Africa, nine results from three studies come from Latin America, and ten observations from three studies are from Australasia. Ten results from two papers have decision-makers located in more than one country, while one paper does not mention where its experiment took place. The country providing the most observations is the USA, with 106 from 19 studies.

This diversity allows us a further set of variables to test for relationships between discrimination and characteristics of the country in which an experiment is run. In Appendix C, Table C.5, therefore, we report regressions including location dummies for the *USA* and *Europe*, and country-level measures of *Individualism* (from the Hofstede Centre), ethno-linguistic-religious *Fractionalisation* (constructed from

Alesina et al, 2003, by averaging each country's scores for ethnic, linguistic and religious fractionalization[29]) and prosperity (*Log GDPpc*, the log of per capita national income at purchasing power parity, as estimated by the World Bank). Using these independent variables requires trimming the sample to exclude experiments conducted across countries, as well as those in locations for which data on *Individualism* is not available.

We do not find any country-level variables to be significant, with rare exceptions. In LPMa2, we find the probability of observing out-group favouritism is lower in the USA than in the rest of the world, significant at the 5% level. However, once controlling for country-level individualism, as in LPMa3, the effect disappears. *Individualism* itself only has a weakly significant effect of reducing the likelihood of out-group favouritism, after omitting the USA dummy in LPMa4.

While the insignificance of country-level variables in our analysis appears to show that results on discrimination can be generalised across cultures, we do not argue this is necessarily the case. The locations at which experiments on discrimination have been conducted are not a random global sample; in many cases they are handpicked by researchers who have prior reason to believe they have an interesting discrimination-related question to ask of a particular subject pool.

Result A3: Country-level variables are not found to significantly explain discrimination.

B.4 How does the experimental context affect the prevalence of each type of discrimination?

To investigate the strength of different types of discrimination in experiments with different types of identity, we run LPM and meta-regressions on the sub-sample of observations for which there is scope only for taste-based discrimination, and the sub-sample for which there is scope for both taste-based and statistical discrimination.

[29] We also ran regressions containing separate variables for ethnic, linguistic and religious fractionalization, none of which were found to have significance.

The results are presented in Appendix C, Table C.6a; LPMa1 and Metareg1 relate to the taste-based only sub-sample, while LPMa2 and Metareg2 relate to the both-types sub-sample. The table reports whether the coefficients for each identity category significantly differ between models 1 and 2; this is deduced by running pooled models with interaction terms. The results of linear restriction tests are also presented in Appendix C, Table C.6b.

When it can only be driven by taste, according to both the LPM and the meta-regression discrimination is significantly greater across artificial groups than across ethnicities, religions, nationalities or gender. All of these differences are significant at the 1% level, apart from the difference between *Artificial* and *Religion* in Metareg(1), which is significant at the 5% level. However, when discrimination can be driven both by tastes and statistical beliefs, neither model finds it to significantly differ between artificial group experiments and those on nationality, religion or ethnicity.[30]

With only taste-based discrimination possible, discrimination is not significantly different across artificial groups to across socially or geographically distinct groups. However, when there is also scope for statistical discrimination, discrimination is significantly higher (1% level) among socially or geographically distinct groups.

The only identity category whose coefficient significantly differs between the sample where only taste-based discrimination is possible and the sample where both types of discrimination are possible, in models ran on both dependent variables, is *Soc/Geo Groupings*. The coefficients on *Ethnicity*, *Religion* and *Nationality* do not significantly differ between samples. The test on the omitted category, *Artificial*, shows its coefficient also does not significantly differ between samples. We therefore interpret the narrowing of the discrimination gap between *Artificial* and *Ethnicity, Religion* and *Nationality* when scope is added for statistical discrimination as being driven by beliefs either reducing discrimination in artificial identity experiments, or enhancing it in experiments with ethnicity, religion and nationality, or both. We interpret the widening of the discrimination gap between *Artificial* and *Soc/Geo*

[30] Except with *Ethnicity* at the 10% level in the LPM.

Groupings when scope is added for statistical discrimination as being driven primarily by beliefs enhancing discrimination between social and geographical groups.

Result A4: Discrimination is only significantly stronger between artificial groups compared to between ethnic, religious and national groups when there is scope only for taste-based discrimination. Discrimination is only significantly stronger between social/geographical groups compared to between artificial groups when there is scope for both types of discrimination.

Appendix C: Additional Regression Output

Table C.1: Linear probability regressions on out-group favouritism

Dependent variable	Out-group favouritism	
	LPMa1	LPMa2
Type of discrimination possible		
Taste + Statistical		-0.003
		(0.020)
Role Types		
First Mover	-0.031	
	(0.029)	
Second Mover	-0.012	
	(0.027)	
Simultaneous Mover	-0.025	
	(0.038)	
Allocator	-0.049	
	(0.037)	
Partner Chooser	0.031	
	(0.063)	
Identity		
Ethnicity	0.041	0.035
	(0.040)	(0.036)
Religion	-0.004	-0.008
	(0.052)	(0.050)
Nationality	0.118*	0.111
	(0.069)	(0.068)
Gender	0.222***	0.231***
	(0.046)	(0.047)
Soc/Geo Groupings	-0.051**	-0.045*
	(0.024)	(0.023)
Natural Other	-0.025	-0.041
	(0.034)	(0.026)
Controls		
Students	-0.023	-0.026
	(0.041)	(0.038)
Sample Size	$4.9e^{-4}$	$2.0e^{-4}$
	$(7.2e^{-4})$	$(2.9e^{-4})$
Constant	0.051	0.038
	(0.046)	(0.047)
R^2	0.088	0.084
N	441	441

Notes: *** p<0.01, ** p<0.05, * p<0.1; LPMa1 and LPMa2 are linear probability models run on full sample; omitted categories are Dictator (role type) and Artificial (identity); errors are corrected for heteroskedasticity, with 77 clusters; standard errors in italics.

Table C.2: Logistic regressions on discrimination and out-group favouritism

Dependent variable	Discrimination	Out-group favouritism
	LOGITa1	LOGITa2
Identity		
Ethnicity	-0.260***	0.062
	(0.046)	(0.069)
Religion	-0.195**	-0.003
	(0.098)	(0.061)
Nationality	0.216***	0.172*
	(0.065)	(0.095)
Gender	-0.252***	0.326***
	(0.045)	(0.097)
Soc/Geo Groupings	0.244**	(dropped)
	(0.111)	
Natural Other	-0.056	(dropped)
	(0.138)	
Role Types		
First Mover	-0.031	-0.022
	(0.081)	(0.017)
Second Mover	-0.080	-0.007
	(0.066)	(0.017)
Simultaneous Mover	0.025	-0.011
	(0.113)	(0.030)
Allocator	0.413***	-0.035**
	(0.103)	(0.014)
Partner Chooser	0.075	0.035
	(0.125)	(0.045)
Controls		
Students	0.011	-0.037
	(0.072)	(0.053)
Sample Size	1.7e^{-4}	1.9e^{-4}
	(4.9e^{-4})	(1.8e^{-4})
Pseudo R^2	0.167	0.132
N	441	367

Notes: *** p<0.01, ** p<0.05, * p<0.1; LPMa1 and LPMa2 are linear probability models run on full sample; omitted categories are Dictator (role type) and Artificial (identity); errors are corrected for heteroskedasticity, with 77 clusters in LOGITa1 and 66 in LOGITa2; standard errors in parentheses; for dummy variables, dy/dx is for discrete change from 0 to 1.

Table C.3a: Linear probability regression on discrimination and meta-regression on effect size (further analysis of role type)

Dependent variable	Discrimination		d	Out-group favouritism
	LPMa1	LPMb1	Metareg1	LPMa2
Identity				
Ethnicity	-0.287***	-0.282***	-0.113	0.032
	(0.078)	(0.084)	(0.081)	(0.041)
Religion	-0.219	-0.220	-0.092	-0.027
	(0.141)	(0.156)	(0.128)	(0.057)
Nationality	-0.234***	-0.133	-0.112	0.107
	(0.085)	(0.106)	(0.078)	(0.067)
Gender	-0.314***	-0.339***	-0.500***	0.233***
	(0.065)	(0.069)	(0.098)	(0.042)
Soc/Geo Groupings	0.242**	0.242*	0.365***	-0.057**
	(0.102)	(0.124)	(0.091)	(0.023)
Natural Other	-0.069	-0.286	-0.203	-0.034
	(0.176)	(0.179)	(0.193)	(0.035)
Role Types				
TG Sender	-0.045	-0.045	0.002	-2.6e^{-4}
	(0.126)	(0.086)	(0.081)	(0.027)
TG Returner	-0.126*	-0.147*	-0.112	0.016
	(0.074)	(0.083)	(0.087)	(0.024)
Social Dilemma	0.014	0.010	-0.022	-0.011
	(0.106)	(0.115)	(0.097)	(0.039)
Allocator	0.348***	0.400***	0.991***	-0.042
	(0.104)	(0.141)	(0.154)	(0.034)
Proposer	-0.081	-0.131	-0.012	-0.088**
	(0.101)	(0.090)	(0.106)	(0.041)
Responder	-0.029	0.129	0.120	-0.017
	(0.099)	(0.109)	(0.135)	(0.055)
Game Other	0.045	0.087	0.034	0.004
	(0.098)	(0.119)	(0.104)	(0.037)
Controls				
Students	0.004	-0.028	0.103	-0.018
	(0.068)	(0.077)	(0.077)	(0.040)
Sample Size	1.2e^{-4}	5.7e^{-5}	-7.1e^{-4*}	2.7e^{-4}
	(4.3e^{-4})	(4.7e^{-3})	(4.3e^{-3})	(3.2e^{-3})
Constant	0.416***	0.442***	0.236**	0.037
	(0.087)	(0.100)	(0.106)	(0.048)
R^2 (adjusted in Metareg1)	0.206	0.214	0.237	0.094
N	441	364	364	441

Notes: *** p<0.01, ** p<0.05, * p<0.1; LPMa1 and LPMa2 are linear probability models run on full sample, Metareg1 is meta-regression run on sample for which effect sizes are available, LPMb1 is linear probability model run on same sample as Metareg1; omitted categories are Dictator (role type) and Artificial (identity); errors in LPM models are corrected for heteroskedasticity, with 77 clusters in LPMa1 and LPMa2, and 67 in LPMb1; standard errors in italics.

Table C.3b: Linear Restriction Tests on models presented in Table C.3a

	P Value on two-tailed test			
Null Hypothesis	LPMa1	LPMb1	Metareg1	LPMa2
Identity				
Ethnicity = Religion	0.613	0.652	0.872	0.43
Ethnicity = Nationality	0.553	0.157	0.991	0.256
Ethnicity = Gender	0.705	0.452	0.001***	<0.001***
Ethnicity = Soc/Geo Groupings	<0.001***	<0.001***	<0.001***	0.066*
Ethnicity = Natural Other	0.203	0.983	0.642	0.164
Religion = Nationality	0.909	0.571	0.884	0.126
Religion = Gender	0.496	0.431	0.009***	<0.001***
Religion = Soc/Geo Groupings	0.001***	0.002***	<0.001***	0.544
Religion = Natural Other	0.47	0.735	0.586	0.902
Nationality = Gender	0.286	0.017**	0.001***	0.085*
Nationality = Soc/Geo Groupings	<0.001***	0.007***	<0.001***	0.017**
Nationality = Natural Other	0.374	0.424	0.653	0.017**
Gender = Soc/Geo Groupings	<0.001***	<0.001***	<0.001***	<0.001***
Gender = Natural Other	0.169	0.769	0.16	<0.001***
Soc/Geo Groupings = Natural Other	0.092*	0.004***	0.003***	0.452
Role Types				
TG Sender = TG Returner	0.25	0.171	0.119	0.53
TG Sender = Social Dilemma	0.605	0.648	0.782	0.808
TG Sender = Allocator	<0.001***	0.008***	<0.001***	0.2
TG Sender = Proposer	0.709	0.343	0.887	0.030**
TG Sender = Responder	0.891	0.25	0.37	0.758
TG Sender = Game Other	0.381	0.282	0.748	0.907
TG Returner = Social Dilemma	0.206	0.176	0.31	0.526
TG Returner = Allocator	<0.001***	0.001***	<0.001***	0.138
TG Returner = Proposer	0.647	0.863	0.341	0.016**
TG Returner = Responder	0.414	0.061*	0.085*	0.551
TG Returner = Game Other	0.086*	0.050*	0.155	0.758
Social Dilemma = Allocator	0.014***	0.019**	<0.001***	0.398
Social Dilemma = Proposer	0.415	0.213	0.932	0.058*
Social Dilemma = Responder	0.761	0.458	0.309	0.912
Social Dilemma = Game Other	0.788	0.564	0.616	0.726
Allocator = Proposer	<0.001***	<0.001***	<0.001***	0.158
Allocator = Responder	0.005***	0.107	<0.001***	0.626
Allocator = Game Other	0.018**	0.050**	<0.001***	0.234
Proposer = Responder	0.722	0.093*	0.367	0.183
Proposer = Game Other	0.186	0.021**	0.699	0.034**
Responder = Game Other	0.542	0.806	0.554	0.699

Note: *** p<0.01, ** p<0.05, * p<0.1; LPMa is linear probability model run on full sample, Metareg is meta-regression run on sample for which effect sizes are available, LPMb is linear probability model run on same sample as Metareg.

Table C.4 Linear probability regressions on discrimination and meta-regressions on effect size for artificial identity experiments only

Dependent variable	Discrimination	d
	LPM	Metareg
Role Types		
First Mover	-0.190	-0.074
	(0.165)	(0.127)
Second Mover	-0.102	-0.098
	(0.142)	(0.135)
Simultaneous Mover	-0.211	-0.005
	(0.154)	(0.139)
Allocator	0.236*	0.898***
	(0.130)	(0.170)
Partner Chooser	0.061	0.083
	(0.213)	(0.182)
Controls		
Students	0.032	0.128
	(0.123)	(0.207)
Sample Size	0.002***	0.001
	(0.001)	(0.001)
Identity Inducement Method		
Labelling	-0.117	0.030
	(0.081)	(0.085)
Other Method	-0.140	0.102
	(0.125)	(0.141)
Team Building	0.085	0.031
	(0.095)	(0.691)
Constant	0.409*	0.102
	(0.207)	(0.247)
R^2 (adjusted for Metareg)	0.154	0.262
N	169	146

Notes: *** p<0.01, ** p<0.05, * p<0.1; LPM is linear probability model run on artificial identity sample, Metareg is meta-regression run on artificial identity sample for which effect sizes are available; omitted categories are Dictator (role type) and Preferences (Identity inducement method); errors in LPM are corrected for heteroskedasticity, with 32 clusters; standard errors in parentheses.

Table C.5: Linear probability regressions on discrimination and out-group favouritism, and meta-regression on effect size, with country-level variables included

Dependent variable	Discrimination	d	Out-group favouritism		
	LPMa1	Metareg1	LPMa2	LPMa3	LPMa4
Identity					
Ethnicity	-0.220**	-0.171	0.053	0.044	0.042
	(0.086)	(0.104)	(0.045)	(0.045)	(0.042)
Religion	-0.146	-0.198	0.014	-0.063	-0.069
	(0.192)	(0.179)	(0.040)	(0.090)	(0.079)
Gender	-0.294***	-0.386***	0.254***	0.244***	0.239***
	(0.100)	(0.113)	(0.047)	(0.048)	(0.045)
Soc/Geo Groupings	0.260**	0.273***	-0.029*	-0.052*	-0.055**
	(0.103)	(0.095)	(0.016)	(0.029)	(0.025)
Natural Other	0.008	-0.311	-0.023	-0.093	-0.099*
	(0.177)	(0.213)	(0.017)	(0.068)	(0.057)
Role Types					
First Mover	-0.027	0.245***	-0.013	-0.012	-0.012
	(0.101)	(0.088)	(0.024)	(0.025)	(0.025)
Second Mover	-0.086	0.182*	0.014	0.014	0.014
	(0.080)	(0.094)	(0.025)	(0.026)	(0.026)
Simultaneous	-0.080	0.206*	-0.004	-0.004	-0.006
	(0.108)	(0.110)	(0.027)	(0.028)	(0.029)
Allocator	0.423***	1.178***	-0.026	-0.046	-0.049*
	(0.142)	(0.165)	(0.020)	(0.029)	(0.027)
Partner Chooser	0.058	0.208	0.049	0.067	-0.066
	(0.130)	(0.141)	(0.060)	(0.063)	(0.063)
Controls					
Fractionalisation	0.015	0.414			
	(0.265)	(0.257)			
LogGDPpc	0.020	-0.031			
	(0.083)	(0.089)			
Europe	0.100	0.277*			
	(0.131)	(0.150)			
USA	0.038	0.096	-0.049**	-0.012	
	(0.132)	(0.163)	(0.022)	(0.034)	
Individualism	-0.001	-0.001		-0.002	-0.002*
	(0.004)	(0.003)		(0.002)	(0.001)
Constant	0.191	0.174	0.035*	0.167	0.179*
	(0.731)	(0.813)	(0.021)	(0.119)	(0.094)
R^2 (adjusted in Metareg1)	0.217	0.256	0.112	0.122	0.122
N	345	304	359	345	345

Notes: *** p<0.01, ** p<0.05, * p<0.1; LPMa1, LPMa3 and LPMa4 are linear probability models run on full sample excluding experiments conducted across countries and in countries for which data on Individualism is not available, LPMa2 is linear probability model run on full sample excluding experiments conducted across countries, Metareg1 is meta-regression run on sample for which effect sizes are available excluding experiments conducted across countries and in countries for which data on Individualism is not available; omitted categories are Dictator (role type) and Artificial (identity); errors in LPM models are corrected for heteroskedasticity, with 60 clusters in LPMa1, LPMa3 and LPMa4, and 65 in LPMa2; standard errors in parentheses.

Table C.6a: Linear probability regressions on discrimination and meta-regressions on effect size, with scope only for taste-based discrimination (Model 1) and scope for both types of discrimination (Model 2)

Dependent variable	Discrimination			d		
	Taste-based only	Taste + Statistical	Test of coefficient difference	Taste-based only	Taste + Statistical	Test of coefficient difference
	LPMa1	LPMa2		Metareg1	Metareg2	
Identity						
Ethnicity	-0.351***	-0.195*		-0.291***	0.020	
	(0.078)	(0.100)		(0.088)	(0.169)	
Religion	-0.474***	-0.159		-0.443**	-0.048	
	(0.125)	(0.234)		(0.172)	(0.206)	
Nationality	-0.292***	-0.162		-0.312***	-0.133	
	(0.089)	(0.122)		(0.115)	(0.113)	
Gender	-0.290***	-0.352***		-0.474***	-0.642***	
	(0.090)	(0.080)		(0.122)	(0.167)	
Soc/Geo Groupings	0.081	0.408***	**	0.015	0.532***	***
	(0.159)	(0.126)		(0.125)	(0.142)	
Natural Other	0.146	-0.224		-0.301	-0.227	
	(0.263)	(0.262)		(0.329)	(0.281)	
Controls						
Students	0.041	-0.047		0.034	0.178	
	(0.101)	(0.095)		(0.101)	(0.142)	
Sample Size	$6.6e^{-4}$	$4.9e^{-4}$		$1.7e^{-5}$	$-3.1e^{-4}$	
	($4.8e^{-4}$)	($7.2e^{-4}$)		($5.4e^{-4}$)	($7.2e^{-4}$)	
Constant	0.380***	0.383***		0.355***	0.178	
	(0.117)	(0.120)		(0.114)	(0.153)	
R^2 (adjusted in Metaregs)	0.175	0.196		0.117	0.174	
N	262	179		204	160	

Notes: *** p<0.01, ** p<0.05, * p<0.1; LPMa1 is linear probability model run on the sample for which discrimination can only be taste-based, LPMa2 is linear probability model run on the sample for which dicrimination can be both taste-based and statistical, Metareg1 is meta-regression run on the sample for which discrimination can only be taste-based and effect sizes are available, Metareg2 is meta-regression run on the sample for which discrimination can be both taste-based and statistical and effect sizes are available; 'test of coefficient difference' reports whether coefficients differ significantly between models 1 and 2; the omitted category is Artificial (identity); errors in LPM models are corrected for heteroskedasticity, with 65 clusters in LPMa1 and 59 in LPMa2; standard errors in parentheses.

Table C.6b: Linear Restriction Tests on models presented in Table C.6a

	P Value on two-tailed test			
Null Hypothesis	LPMa1	LPMa2	Metareg1	Metareg2
Ethnicity = Religion	0.171	0.876	0.38	0.756
Ethnicity = Nationality	0.353	0.788	0.87	0.405
Ethnicity = Gender	0.348	0.045**	0.163	0.003***
Ethnicity = Soc/Geo Groupings	0.002***	<0.001***	0.016**	0.005***
Ethnicity = Natural Other	0.052*	0.911	0.977	0.373
Religion = Nationality	0.082*	0.989	0.509	0.699
Religion = Gender	0.097*	0.413	0.878	0.020**
Religion = Soc/Geo Groupings	<0.001***	0.013**	<0.005***	0.006***
Religion = Natural Other	0.018**	0.839	0.679	0.532
Nationality = Gender	0.985	0.057*	0.29	0.005***
Nationality = Soc/Geo Groupings	0.012**	<0.001***	0.04**	<0.001***
Nationality = Natural Other	0.099*	0.823	0.974	0.745
Gender = Soc/Geo Groupings	0.015**	<0.001***	0.003***	<0.001***
Gender = Natural Other	0.097*	0.624	0.613	0.192
Soc/Geo Groupings = Natural Other	0.81	0.015**	0.33	0.007

Note: *** p<0.01, ** p<0.05, * p<0.1

Chapter Three: On the social inappropriateness of discrimination[31]

Abstract

We experimentally investigate the relationship between discriminatory behaviour and the perceived social inappropriateness of discrimination. We test the framework of Akerlof and Kranton (2000, 2005), which suggests discrimination will be stronger when social norms favour it. Our results support this prediction. Using a Krupka-Weber social norm elicitation task, we find participants perceive it to be less socially inappropriate to discriminate on the basis of social identities artificially induced, using a trivial minimal group technique, than on the basis of nationality. Correspondingly, we find that participants discriminate more in the artificial identity setting. Our results suggest norms and the preference to comply with them affect discriminatory decisions and that the social inappropriateness of discrimination moderates discriminatory behaviour.

JEL classifications: C71 – Cooperative games; C92 – Laboratory Experiments (Group Behavior); D03 – Behavioral Microeconomics: Underlying Principles

Keywords: Discrimination; Social norms; Krupka-Weber method; Allocator game

[31] This chapter was co-authored with Abigail Barr and Daniele Nosenzo.

1. Introduction

Economic theories seeking to explain discrimination focus on two mechanisms. First, in the presence of incomplete information, profit- or income-maximizing agents use aggregate group characteristics to form statistical beliefs about individual characteristics and then act in accordance with those beliefs by, potentially, treating members of different groups differentially (Arrow, 1972). Second, individuals are assumed to derive direct utility from favouring certain groups relative to others, i.e. they are assumed to have a ‘taste for discrimination’ (Becker, 1957). Such tastes explain why discrimination is observed even in settings where asymmetric or incomplete information is not an issue (e.g. Chen and Li, 2009; Abbink and Harris, 2012). However, given their empirical importance, the psychological foundations of such tastes or preferences for discrimination have received remarkably little attention in the literature.

In this paper we use experimental methods to test whether tastes for discrimination are systematically shaped by *social norms*, i.e. by collectively recognised rules of behaviour that define which actions are viewed as socially appropriate within a specific social group.[32] The importance of norms for discriminatory behaviour has been suggested by Akerlof and Kranton (2000, 2005). In their framework, individuals mentally place themselves in social categories (or identity groups), thereby assigning themselves social identities.[33] They have perceptions of the specific prescriptions (norms) that mandate how individuals within these identity groups are expected to behave, and gain utility from conforming to the prescriptions that apply to their own group, as it ‘affirms [their] self-image, or identity’ (Akerlof and Kranton, 2000, p. 716). Within this framework, intergroup discrimination arises if the behaviours that are prescribed to the members of one group involve differential treatment or consideration of in-group and out-group members.[34]

[32] See Elster (1989) and Ostrom (2000) for definitions of social norms.

[33] See also Huang and Wu (1994) and Montgomery (1994) for related approaches.

[34] As an extreme example, consider the case of caste discrimination in South Asia and the belief that caste 'purity' (identity) can be 'polluted' by interactions with the individuals at the bottom of the caste system (known as 'Dalits'). This led to the so-called 'untouchability practices', a set of strongly discriminatory norms against Dalits, which, for example, impose segregation and

The Akerlof and Kranton framework implies a positive correlation between in-group members' beliefs about the appropriateness of discrimination and the incidence of discriminatory behaviour. Similar correlations have been found in relation to other types of economic behaviour. Following Krupka and Weber (2013), experiments have shown that in a variety of economic contexts people are more likely to take an action if they perceive it to be more socially appropriate (e.g. Burks and Krupka, 2012 – corporate ethics; Gachter et al, 2013 – gift-exchange; Krupka et al, 2016 – informal contract enforcement; Banerjee, 2016 – bribery). There is also non-experimental evidence suggesting norms drive economically-relevant behaviour (e.g. Buonanno et al, 2009). Therefore, economists are increasingly invoking social norms and norm-compliance to explain empirical behaviour. In driving behaviour, social norms may effectively substitute for laws (e.g. Huang and Wu, 1994), or may complement them (e.g. Sunstein, 1990; Kubler, 2001; Lazzarini et al, 2004; Posner, 2009; Benabou and Tirole, 2011).

However, a correlation between individuals' beliefs about the appropriateness of discrimination and the prevalence of discriminatory behaviour is a challenge to empirically document using naturally occurring data, not least of all because of the difficulties associated with accurately measuring such beliefs.[35]

Occasionally, attitudinal surveys include questions that can be interpreted as eliciting respondents' perceptions of the appropriateness of discrimination. For instance, the 2002 wave of the Scottish Social Attitudes Survey asked respondents whether they believed that 'sometimes there is good reason for people to be prejudiced against certain groups'. One can interpret positive responses to this question as an imperfect proxy for the perceived social appropriateness of racial discrimination. Using this interpretation, we calculated the percentage of residents in each local authority area of Scotland who agreed with the survey question. For each area, Figure 1 plots this

restrictions on occupation, prohibit inter-caste marriage, and limit or prohibit access to public places and services.

[35] See Krupka and Weber (2013) and Mackie et al. (2015) for a discussion of the difficulties of measuring social norms empirically.

variable against the number of racist incidents[36], per 100 non-white residents[37], reported to the police in the financial year 2003-4 (Scottish Executive Statistical Bulletin, 2007). A correlation of 0.27 between the two variables indicates a positive relationship between the social appropriateness of racial discrimination and the incidence of racially discriminatory behaviour, which is consistent with the Akerlof and Kranton framework.

The acceptability of prejudiced-based humour has sometimes been used as a proxy for the normative appropriateness of discrimination (see, e.g., Crandall et al., 2002). Figure 2 plots, over the period 2004 to 2014, the frequency of Google searches in the US for 'N***** jokes' (we apply the censorship for this paper; the original search term was uncensored[38]), as a proportion of all Google searches in the US (Google Trends, 2016). Searching for racist jokes about black people can be treated as evidence that the searcher perceives discrimination against black people to be socially appropriate. Figure 2 also plots, on an annual basis over the same period, the number of incidents in the US involving hate crimes motivated by an anti-black bias that were reported to the FBI, per every 100 people living in areas where the hate crimes are reported[39] (United States Department of Justice, 2005, 2006, 2007, 2008, 2009, 2010, 2011, 2012, 2013, 2014, 2015). Both the frequency of anti-black joke searches and the rate of anti-black hate crime incidents declined considerably over the period. This is suggestive of a positive relationship in the US between the change over time in the social appropriateness of discrimination against black people and the change over time in discriminatory behaviour against black people.

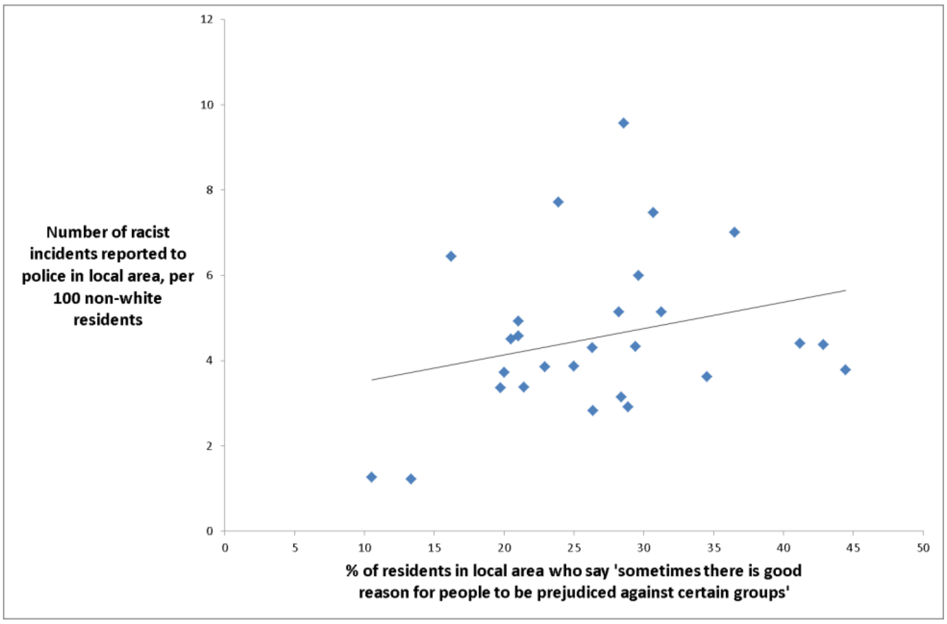
[36] The Scottish police define a racist incident as 'any incident which is perceived to be racist by the victim or any other person.' (Scottish Executive Statistical Bulletin, 2007)

[37] The contemporaneous proportion of non-white residents in each Scottish local area is taken from the 2001 UK Census (National Records of Scotland, 2011).

[38] We deliberated over our decision to censor the word, but eventually concluded that we felt uncomfortable using it uncensored even in a scientific context. We expect readers will be able to guess the extremely derogatory term describing black people that we refer to.

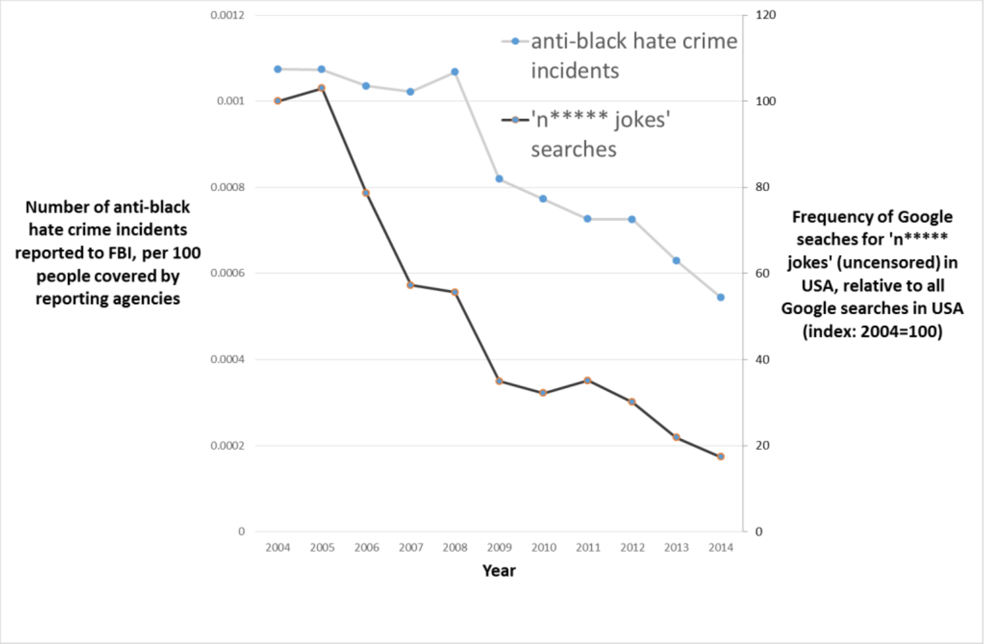
[39] We report this, rather than the absolute number of hate crimes, to adjust for the fact that the population covered by the FBI's hate crime statistics varies from year to year. The proportion of black people in the covered population is not available.

Figure 1: Variations in attitudes towards racial prejudice and race crimes across Scottish local authority areas



Note: Figure 1 plots, at the level of local authority area, the relationship between attitudes to prejudice, as reported in the Scottish Social Attitudes Survey 2002, and the frequency of racist incidents reported to the police in the financial year 2003-4. Each data-point represents one local authority area in Scotland.

Figure 2: Google searches for racist jokes about black people and anti-black hate crimes in the US, 2004-14



*Note: The light grey line plots the number of anti-black hate crime incidents reported to the FBI each year, adjusted for the population size covered by reporting agencies at the time. The dark grey line plots the relative frequency, amongst Google searches in the US, of the search term 'N***** jokes' (censorship applied retrospectively) – monthly data was recovered using the Google Trends tool, and is averaged over the course of each year.*

In spite of these examples, the paucity of useful naturally occurring data with which to investigate the empirical relevance of Akerlof and Kranton's framework to the issue of discrimination advances the case for using experimental methods to address the question. Our paper does this, with an empirical strategy relying on four main elements.

First, we use standard experimental techniques to prime participants to think about particular dimensions of their identities. The priming aims to trigger the process of social identification that is central to Akerlof and Kranton's approach by encouraging subjects to identify with half of the participants in their experimental session and not with the other half.

Second, in the decision-making phase of the experiment we ask subjects to distribute a given amount of money between two potential recipients, one an individual sharing their primed identity, the other an individual not sharing their primed identity. This simple allocation task allows us to measure discrimination as the extent to which individuals are willing to favour members of their own social group at the expense of the out-group.

Third, we vary the dimension of identity that is primed. Applying Akerlof and Kranton's framework, the distributive decision that an individual makes within our experiment will depend on the normative prescriptions that apply, given the individual's own social identity and the way the social identities of each of the two recipients relate to it. This implies that the content of the normative prescriptions pertaining to discrimination depend on what dimension of identity is salient within the decision-making context.[40] Focusing on this aspect of Akerlof and Kranton's framework, we design two identity treatments, aimed at inducing different perceptions of the appropriateness of discrimination, while holding other aspects of the decision-making context constant. Under one treatment, social identities are based on nationality; we form groups in the laboratory based on whether participants are British or Chinese. Under the other treatment, social identities are entirely artificial; groups are formed according to the colour of ball that each participant draws blindly from a bag. We expect the prescriptions that mandate how a decision-maker should treat in-groups and out-groups in our experiment to differ across the two treatments. Specifically, we expect discrimination against out-group and in favour of in-group members to be perceived as *less* appropriate when identity groups are formed on the basis of nationality, than when they are artificially formed on the basis of the colour

[40] For example, norms may render it appropriate to discriminate against others who support a different football team or listen to a different type of music from oneself, but not appropriate to discriminate against others who are different in terms of ethnicity or gender; and individuals may moderate their behaviour accordingly.

of balls randomly picked. Therefore, if discrimination is systematically shaped by norms, we expect discrimination to be stronger between the artificial groups.

Fourth, as well as measuring discrimination, we directly measure the perceived social appropriateness of discrimination in each treatment. We do this by employing the 'norm-elicitation' task introduced by Krupka and Weber (2013), in which participants are described the allocation game and are asked to evaluate the social appropriateness of each and every possible action available to the allocator. We use this norm-elicitation task to construct an incentivized measure of the extent to which participants' perceptions of the appropriateness of discrimination vary across our two treatments and to examine the extent to which these differences in perceived appropriateness translate into differences in discriminatory behaviour in the allocation task.

Our results show that, in both treatments, discriminatory actions are viewed as socially inappropriate. However, as expected, discrimination is perceived to be significantly less appropriate in the nationality treatment compared to the artificial identity treatment. The results of the decision task match these differences in perceived appropriateness: while few participants discriminate in either treatment, discrimination is significantly stronger between artificial groups than between nationality groups. These results are consistent with the Akerlof-Kranton framework: the perceived social appropriateness of discrimination varies according to the way identity groups are defined, and this forms the basis for individuals' revealed preferences for discrimination.

Our study's main contribution is in linking discrimination to social norms and social identity theory. In this sense, our study is closely related to the paper by Chang et al. (2015), who apply Akerlof and Kranton's framework to investigate the effect of priming US citizens' political identities on redistributive behaviour. They show that individuals' primed political identities (Democratic or Republican) determine their perceptions of the social appropriateness of redistribution, and that this explains differences in redistributive behaviour between Democrats and Republicans. Like Chang et al., our experiment also shows that both individuals' distributive decisions and their perceptions of the social appropriateness of such decisions are sensitive to the dimension of identity that is salient in a given context. However, while the normative prescriptions upon which Chang et al focus relate to the social identity of the decision-maker alone, we focus on the social identities of *both* the decision-makers *and* other individuals affected by their behaviour, *and* on how those social identities relate one to another. Thus, unlike Chang et al., in our experiment both the priming and the distributive decisions have an *intergroup* component which allows us to investigate the relationship between social identities, social norms, and discriminatory behaviour.

Our paper is also related to work on the associations between social identity and norm enforcement.[41] Bernhard et al. (2006) and Goette et al. (2006), for instance, use third-party punishment games to study whether the willingness to enforce norms of sharing and cooperation depends on the social identities of the norm violator and of the victim of the norm violation and on how those identities relate to that of the norm enforcer. Both papers find that social identity systematically affects the patterns of norm enforcement: enforcers are generally more willing to mete out punishment against violators when the victim of the norm violation is an in-group rather than an out-group member. Also related is Harris et al. (2014), who study whether in-group favouritism is proscribed by social norms by observing the extent to which individuals are willing to incur costs to punish it. They find that in-group favouritism goes largely unpunished when the punisher belongs to the same identity group as the norm violator or when she belongs to a neutral group. In-group favouritism is instead frequently punished when the punisher belongs to a different identity group. Harris et al. conclude that in-group favouritism is not always considered a violation of social norms, as this depends on the identities of the agents involved in the interaction.

[41] Also relevant is the research, mostly undertaken by psychologists, on the associations between social norms and the expressions of prejudiced views – a related but different phenomenon to acts of discrimination. Crandall et al (2002), for instance, found that expressions of prejudice towards groups are very strongly correlated with reported beliefs on the social appropriateness of such prejudice. Other studies have shown that the degree to which individuals are willing to express prejudice can easily be swayed by the views of others (Blanchard et al, 1994; Zitek and Hebl, 2007), or by an experimenter deceptively varying the social norm that is presented to them (Nesdale et al, 2005), suggesting that normative consideration may play an important role on the expression of prejudice.

While these studies strongly suggest an association between discrimination and social norms and identities, none of them has directly measured the norms that underlie the observed patterns of behaviour. Moreover, none of these studies has investigated whether variations in primed social identity trigger differences in norms that, in turn, predict variations in discrimination. Thus, our study fills an important gap in this literature, as we are the first to provide direct evidence not only that discrimination co-varies with group norms, but also that these norms vary across particular dimensions of an individual's identity.

The rest of the paper is set out as follows: Section 2 sketches a simple theoretical model of identity and norm-compliance that we use to motivate and inform our empirical strategy. Section 3 outlines our experimental design; Section 4 presents our results; Section 5 concludes and discusses our findings.

2. Theoretical framework

Our simple model of social norm-compliance closely follows Krupka and Weber (2013), and in particular Chang et al (2015), who based theirs on Akerlof and Kranton (2000, 2005). We first assume that individuals have multiple social identities, the salience of which depends on the decision-making context.

An individual i's utility U_i depends on the actions of him- or herself and others, $a = (a_i, a_{-i})$, and the salient social identities of him- or herself and others, $I = (I_i, I_{-i})$:

$$U_i(a, I) = V_i(a) + \gamma_i N(a_i | a_{-i}, I)$$

We assume that the decision-maker's utility can be broken into two components. The first component, $V_i(a)$, describes individual i's utility over material payoffs, which in turn depend upon his or her own actions and the actions of others. Note that this accommodates standard self-regarding preferences, where the individual only cares about his or her own material payoff, as well as various forms of outcome-based other-regarding preferences, where individual i's utility also depends on others' material payoffs (e.g. Fehr and Schmidt, 1999; Bolton and Ockenfels, 2000).

The second component of utility is derived from complying with normative prescriptions and is captured by the function $N(.)$. The normative prescriptions define, for each action a_i available to individual i, the social appropriateness of that action, given the actions of other players. Crucially, as in Akerlof and Kranton (2000, 2005), we assume that normative prescriptions also depend on the salient identities of i and other players, and on how these relate to one another. This takes into account that differently defined identity groups may normatively prescribe different behaviours, and therefore that the same action may be viewed as more or less socially appropriate depending on the salient dimension of the identity of the decision-maker in a given context, as well as the identities of the other players with whom the decision-maker interacts. Finally, γ_i is an individual-specific parameter defining the importance that individual i attaches to complying with social norms.

In our experiment, subjects face a simple allocation task (described in detail in the next section), which measures the extent to which they are willing to treat differently those who belong to the same identity group as themselves from those who belong to a different one. In all treatments of the experiment, we keep constant the set of material payoffs available to players and the mapping from actions into payoffs. Thus, the first component $V_i(a)$ of the utility function above is held constant across treatments.

Our treatments vary the dimension of identity I that is made salient to the decision-makers and, hence, the process by which the relevant identity groups are defined in the experiment. As we describe in detail in the next section, in one treatment participants are encouraged to form identity groups on the basis of a random event, while in the other treatment identity groups are based on a meaningful personal characteristic. An implication of this treatment manipulation is that the normative prescriptions, $N(a_i|a_{-i}, I)$, that regulate the second component of the utility function described above may differ across treatments. Specifically, the same action a_i available to the decision-maker may be evaluated differently depending on how identity groups are formed. We employ a norm-elicitation technique, based on the task introduced by Krupka and Weber (2013), to quantify, in an incentive-compatible way, the function $N(.)$ in each treatment. This allows us to assess the extent to which normative prescriptions do indeed differ across treatments; and therefore to examine the extent to which differences between treatments in the level of discrimination in the allocation task are explained by differences in the perception of its appropriateness.

3. Experimental design

Measuring discrimination – the allocator game

In the allocator game, one participant was endowed with £16 and asked to allocate it between two passive players, one belonging to his or her own identity group and the other belonging to a different identity group.[42] The decision-maker could not keep any of the money for him- or herself but knew he or she would receive a payment, between £6 and £10, which the computer would randomly pick at the end of the experiment.[43] Allocators could split the money any way they liked between the other two players, as long as each amount was a multiple of two. Thus, the allocator had to choose one of nine possible allocations of money between the two passive players, ranging from (£16; £0) to (£0; £16). In order to maximize sample sizes, we elicited decisions using a role randomisation method: all participants were asked to make a decision in the allocator role knowing that their actual role would be determined at random at the end of the experiment (participants had a one-third chance of being assigned the allocator role and a two-thirds chance of being assigned the passive player role). Role assignment was implemented at the end of experiment, once everyone had submitted an allocation decision. Decisions were made anonymously and the only information allocators had about their recipients was the identity group that each of them belonged to.

We chose the allocator game as our discrimination-eliciting device for the following reasons. First, given our focus on the micro-foundations of taste-based discrimination, we wanted a decision-making task within which statistical discrimination had no relevance; in the allocator game the decision-maker's material payoff does not

[42] See Supplementary Online Materials A for a copy of the instructions used in the experiments.

[43] The possible payments were £6, £8 and £10; each had 1/3 probability of occurring. Our aim was to pay allocators £8 on average. However, had we had made this payoff a certainty it might have inflated the salience of the (8,8) split in the allocator game, as this allocation would ensure payoff equality across all three players.

depend on what any other player does, so statistical beliefs about other players are irrelevant.[44] Second, to maximise our chances of discerning treatment differences, we wanted a task that reliably produces discriminatory behaviour; in a meta-analysis Lane (forthcoming) found the allocator game to be the experimental task that yielded the strongest discrimination. Finally, in the allocator game, discrimination is measured within-participants, so it is obvious to participants what the experiment is about and any observed discrimination is interpretable as conscious rather than subconscious. Thus, the game is an ideal subject for a norm-elicitation task; it is much simpler to assess the social appropriateness of conscious behaviour than of subconscious behaviour.

Measuring the social appropriateness of discrimination – the Krupka-Weber norm-elicitation task

We elicited the social appropriateness of discrimination in the allocator game using an adaptation of the task design pioneered by Krupka and Weber (2013). Participants were described the allocator game, were presented with a table listing the nine possible actions the allocator could take, and were asked to evaluate the social appropriateness of each by marking one option on a four-point scale: 'Very socially inappropriate', 'Somewhat socially inappropriate', 'Somewhat socially appropriate' or 'Very socially appropriate.' To ensure that the relevant perceptions of appropriateness are measured, the evaluators should be, to the greatest extent possible, in the mind-set of the person making the decision they are evaluating. In our experiment, participants in the norm-elicitation task were the same as those playing the allocator game, although we varied which task came first (participants were unaware of the content of the second task until they had completed the first). All participants were assigned to identity groups before their first task, so those taking the norm-elicitation task first had had their identities primed in exactly the same way as the allocator game participants whose behaviour they were evaluating. Each

[44] Note that given the non-strategic nature of the allocator game certain elements of the utility function set out in the previous section are redundant. This notwithstanding the proposed framework remains relevant. In section 4, for the purpose of analysis, we set out a parameterised version of the utility function that is directly and entirely relevant to the game.

individual in the norm-elicitation task only evaluated the appropriateness of actions made by allocators of the same identity group as that individual.

The evaluation of actions was incentivised. Participants were told that, at the end of the experiment, one of the nine actions they had evaluated would be randomly selected, and each participant’s evaluation of the action would be compared to that of another randomly selected participant. If a participant’s evaluation matched that of the person they were compared with, that participant would earn £8; otherwise they would earn nothing. The incentives transform the task into a coordination game, where participants are incentivised to match other participants’ evaluations of appropriateness. Krupka and Weber (2013) argue that this gives participants an incentive to reveal their perception of what is commonly regarded as appropriate or inappropriate behaviour in the decision situation, rather than their own personal evaluation of the actions they are asked to consider. This is important because social norms are collectively recognized rules of behaviour, rather than personal opinions about appropriate behaviours (e.g. Elster, 1989; Ostrom 2000).

Moreover, because we wanted to incentivise participants to coordinate on *identity-specific* social norms (i.e. the social norms that were recognised by those belonging to a specific identity group), participants were told that the person whose evaluation theirs would be compared to would be a member of their own identity group. Participants were told:

> ‘By socially appropriate, we mean behaviour that you think most participants [of your group] would agree is the "correct" thing to do. Another way to think about what we mean is that if [the allocator] were to select a socially inappropriate action, then another participant [of your group] might be angry at [the allocator].’

Treatments

Our treatments, labelled *Nationality* and *Artificial*, differed in the way identity groups were formed. In *Nationality* participants in the experiment were segregated into identity groups based on nationality (previous economics studies taking this approach include Hennig-Schmidt et al, 2007; Netzer and Sutter, 2009; Guillen and Ji, 2011). In *Artificial* participants were split into ‘minimal groups’, using a variant of the technique first introduced by Tajfel et al (1971), wherein social identities are artificially instilled in participants during the experiment.

For both treatments we recruited British and Chinese students at the UK campus of the University of Nottingham, a British institution which hosts a large number of students from China.[45] In the *Nationality* treatment, upon arrival, the British were seated on one side of the lab and the Chinese on the other. At every computer terminal on the British (Chinese) side was placed a sign reading 'YOU ARE ON THE BRITISH (CHINESE) SIDE OF THE ROOM. ALL PARTICIPANTS ON THIS SIDE OF THE ROOM ARE BRITISH (CHINESE)' (see Supplementary Online Materials B). In the instructions at the beginning of the experiment, it was again made explicitly clear that the lab and the participants had been divided based on nationality.

In the *Artificial* treatment, upon arrival, participants blindly drew a ball from a bag. In each session the bag initially contained equal numbers of green and yellow balls, and participants continued to draw from it until the bag was empty, thus ensuring an equal split of green and yellow balls drawn. Those with green balls were then seated on one side of the lab, and those with yellow on the other. Consistent with the *Nationality* treatment, signs were placed at each terminal, reading 'YOU ARE ON THE (GREEN/YELLOW) SIDE OF THE ROOM. ALL PARTICIPANTS ON THIS SIDE OF THE ROOM DREW A (GREEN/YELLOW) BALL', and it was again made explicit at the beginning of the instructions that the lab and the participants had been divided on the basis of ball colour.

[45] Participants were recruited using ORSEE (Greiner, 2015), an online database of experimental participants, upon which participants are asked to state their nationality when they sign up. We were able to cross-check nationalities using the University of Nottingham's central student register system, which lists students' official nationalities. Note that we based the groups in our experiment on official nationalities, rather than self-identified ones (e.g. we did not invite Malaysian students who listed their nationality as Chinese). Chinese participants were mainlanders, with none from Hong Kong, Macao or Taiwan.

As in the *Nationality* treatment, we invited an equal mix of British and Chinese students to the *Artificial* sessions. This ensures comparability between the two treatments.[46]

We chose our treatment manipulation because we conjectured that it would produce the differences that we needed to test the Akerlof and Kranton framework. Specifically, we conjectured that discrimination would be stronger in the *Artificial* compared to the *Nationality* condition. This conjecture was based primarily on existing evidence from previous research: experiments priming national identity (e.g. Hennig-Schmidt et al, 2007; Netzer and Sutter, 2009; Willinger et al, 2003) have often not found significant discrimination, while experiments involving minimal group identity (e.g., Ahmed, 2007; Chen and Li, 2009; Hargreaves-Heap and Zizzo, 2009) do so more frequently, and according to a recent meta-analysis by Lane (forthcoming), on average, discrimination is significantly weaker in the former compared to the latter type of experiment. There are also theoretical reasons why discrimination would be stronger in the *Artificial* condition. First, members of newly formed groups may be more inclined to draw boundaries between in- and out-groups than members of more established groups (Jetten et al., 1996). In our experiment, the groups in the *Artificial* treatment are new, while those in the *Nationality* treatment are not. Second and more closely related to the theoretical framework above, the extent to which individuals are willing to behave prejudicially may be related to how easily the expression of prejudice can be justified to oneself or others (Crandall et al., 2002).[47] In our experiment, discrimination against people who randomly drew a ball of a

[46] Given the relatively small Chinese community in Nottingham, Chinese participants in our experiment were more likely to know each other than were the British. This could be problematic if, particularly in the *Nationality* treatment, participants based their behaviour on the number of friends they had on either side of the lab. We controlled for this by asking each participant, in the post-experimental questionnaire, how many people on each side of the lab they had previously met. Chinese participants were indeed more likely to know each other, but there was no association between the number of friends on either side of the lab and participants' behaviour in either treatment (available on request).

[47] For instance, Crandall et al. (2002) show that there are large differences in the perceived appropriateness of prejudice against Blacks vis-à-vis members of the American Nazi Party. They argue that this is related to the differences in "… an outside perceiver's sense of the justification of the prejudices …. the justifications of the prejudice against Nazis are widely accepted; the justifications of the prejudice against Blacks are not." (p. 361).

different colour may be easier to justify than discrimination on the basis of nationality. Third and again closely related to the theoretical framework, relatively weak norms against discrimination in the *Artificial* treatment could arise because it triggers group identity akin to sports fandom, a dimension of identity across which discrimination, via competition, is expected. In contrast, there may be stronger norms proscribing discrimination against foreign nationals, given the historical sensitivities this could arouse.

Procedure

All participants participated in both the allocator game and the norm-elicitation task, as well as completing a post-experimental questionnaire. In each session, everyone received payment either for the allocator game or for the norm-elicitation task, as determined by a coin toss at the end of the experiment. Participants also received a £4 show-up fee. The order in which the tasks were performed was randomised between sessions, so that we could check for ordering effects. We do not find such effects (see Supplementary Online Materials C for the analysis), which is consistent with the findings of Erkut et al (2015) and D'Adda et al (2016). Therefore, in the analysis below we pool across ordering conditions. All sessions had 24 participants – twelve belonging to each group – and were conducted in March or April 2015, using z-Tree (Fischbacher, 2007). We conducted ten sessions, with 120 participants participating in each treatment.[48]

4. Results

Treatment differences – social norms

We look first at the social appropriateness of discrimination in each treatment, as measured by the norm-elicitation task. Figure 3 plots the mean appropriateness ratings assigned to each allocation in the *Nationality* and *Artificial* treatments. Following the approach of Krupka and Weber (2013), we assign evenly-spaced values of -1 for the rating 'very socially inappropriate', -0.33 for the rating

[48] We conducted one additional session in the *Artificial* treatment which we exclude from the analysis. This is due to procedural issues that resulted from a low turn-up rate. Excluding the session does not meaningfully affect any important results.

‘somewhat socially inappropriate’, 0.33 for the rating ‘somewhat socially appropriate’ and 1 for the rating ‘very socially appropriate.’ The table at the bottom of the figure displays the distribution of evaluations for each allocation in each treatment, and presents the results of randomisation tests on the treatment differences in mean ratings. Our results are corrected for the fact that we are performing multiple tests; applying the Benjamini-Hochberg False Discovery Rate method (Benjamini and Hochberg, 1995), we sort our p-values in ascending rank and multiply each by the number of separate tests being performed (in our case nine, one for each possible allocation) before dividing each by its rank – thus the greatest adjustment is made to smaller p-values.[49]

In each treatment the mean and modal evaluations follow the same general pattern. Participants tend to regard extreme discrimination against recipients belonging to either identity group to be very socially inappropriate, while the equal split is generally regarded as very socially appropriate. There is a lack of strong consensus on allocations mildly favouring members of one group or the other. This pattern is consistent with a social norm of equality. However, in both treatments the perceived social appropriateness decays faster as allocations move away from equality towards favouring the out-group member than when they move towards favouring the in-group member, indicating that social norms against discrimination are stronger when the victim is a member of one’s own identity group.[50]

By design, any treatment differences in the ratings assigned to a given allocation can only be driven by contextual differences in the perceived appropriateness of discrimination. We observe subtle but significant treatment differences. Whereas 95% of participants in the *Nationality* treatment perceive the equal split to be very appropriate, the equivalent figure is only 84.2% in the *Artificial* treatment; mean

[49] All p-values reported in this paper are two-sided and based on Fisher randomisation tests and corrected using the Benjamini-Hochberg False Discovery Rate method. See Moir (1998) for a discussion of the randomisation test, and Kaiser and Lacy (2009) for information on the Stata command used to apply it.

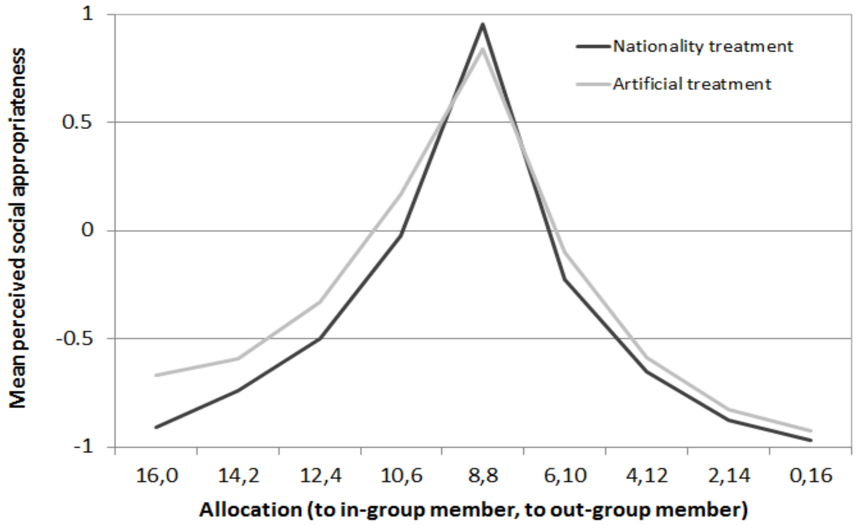
[50] OLS regressions confirm that, in both treatments, the rate of decay of appropriateness of allocations favouring the out-group is significantly higher than that of allocations favouring the in-group (both p-values < 0.002).

ratings for the equal split are significantly higher in the *Nationality* treatment. Furthermore, as the allocations move away from the equal split towards favouring the in-group, the appropriateness ratings decline at a faster rate in the *Nationality* treatment than in the *Artificial* treatment. For the extreme (16,0) split, 92.5% of participants in the *Nationality* treatment opt for 'very inappropriate', while only 80.8% do so in the *Artificial* treatment. And while only 5% of participants rate the (16,0) allocation as socially appropriate in the *Nationality* treatment, 18% do so in the *Artificial* treatment. In fact, Figure 3 shows that, for any in-group-favouring allocation, there are more participants in the *Artificial* than *Nationality* treatment who find discrimination to be socially appropriate.[51]

As a consequence, all in-group-favouring allocations are on average perceived to be more appropriate in the *Artificial* treatment, and the differences are statistically significant at the 5% level or lower in three out of four possible

[51] In Supplementary Online Materials D we show that these treatment differences in the perceived norms are driven by variations in the within-subject response patterns to the norm-elicitation task across treatments. In particular, in the *Artificial* treatment we find relatively more subjects who assign their highest appropriateness rating to the (16,0) allocation and then monotonically decrease their ratings of appropriateness as more money is given to the out-group member. Such a pattern indicates the perception of a social norm of in-group favouritism.

Figure 3: Perceived social appropriateness of actions in allocator game



	16,0	14,2	12,4	10,6	8,8	6,10	8,12	2,14	0,16
Nationality treatment									
Very appropriate	1.7	2.5	2.5	5.0	95.0	0.8	0.0	0.0	0.0
Somewhat appropriate	3.3	5.8	10.0	51.7	3.3	35.0	4.2	1.7	1.7
Somewhat inappropriate	2.5	20.0	47.5	28.3	1.7	43.3	44.2	15.0	1.7
Very inappropriate	92.5	71.7	40.0	15.0	0.0	20.8	51.7	83.3	96.7
Mean rating	**-0.91**	**-0.74**	**-0.50**	**-0.02**	**0.96**	**-0.23**	**-0.65**	**-0.88**	**-0.97**

Allocation (to in-group member, to out-group member)

	16,0	14,2	12,4	10,6	8,8	6,10	8,12	2,14	0,16
Artificial treatment									
Very appropriate	12.5	7.5	5.0	9.2	84.2	0.8	0.8	1.7	3.3
Somewhat appropriate	5.8	10.8	18.3	64.2	10.0	46.7	6.7	2.5	0.8
Somewhat inappropriate	0.8	17.5	49.2	19.2	3.3	39.2	46.7	15.8	0.0
Very inappropriate	80.8	64.2	27.5	7.5	2.5	13.3	45.8	80.0	95.8
Mean rating	**-0.67**	**-0.59**	**-0.33**	**0.17**	**0.84**	**-0.10**	**-0.58**	**-0.83**	**-0.92**
p-value (Benjamini-Hochberg)	**0.011**	**0.085**	**0.024**	**0.031**	**0.024**	**0.082**	**0.323**	**0.354**	**0.335**

Notes: Figure 3 presents the distribution of social appropriateness ratings of each allocation in the two treatments. Allocations (e.g. 16,0) are denoted by the amount given to the in-group member on the left (£16), and the amount given to the out-group member on the right (£0). Shaded cells represent the modal ratings for each allocation in each treatment. Mean ratings are taken by assigning values of 1, 0.33, -0.33 and -1 for the ratings 'very appropriate', 'somewhat appropriate', 'somewhat inappropriate' and 'very inappropriate' respectively, and averaging the values for all participants in a given treatment. Benjamini-Hochberg-corrected p-values are reported from randomisation tests.

cases (the exception being the allocation 14,2 for which the difference is significant at the 10% level). Moreover, the differences in perception of appropriateness of discrimination only pertain to in-group favouritism and not to *any* form of discrimination; Figure 3 shows that, while out-group-favouring allocations are on average perceived to be slightly more appropriate in the *Artificial* treatment, only for the (6,10) allocation is the difference significant, and then only at the 10% level.[52]

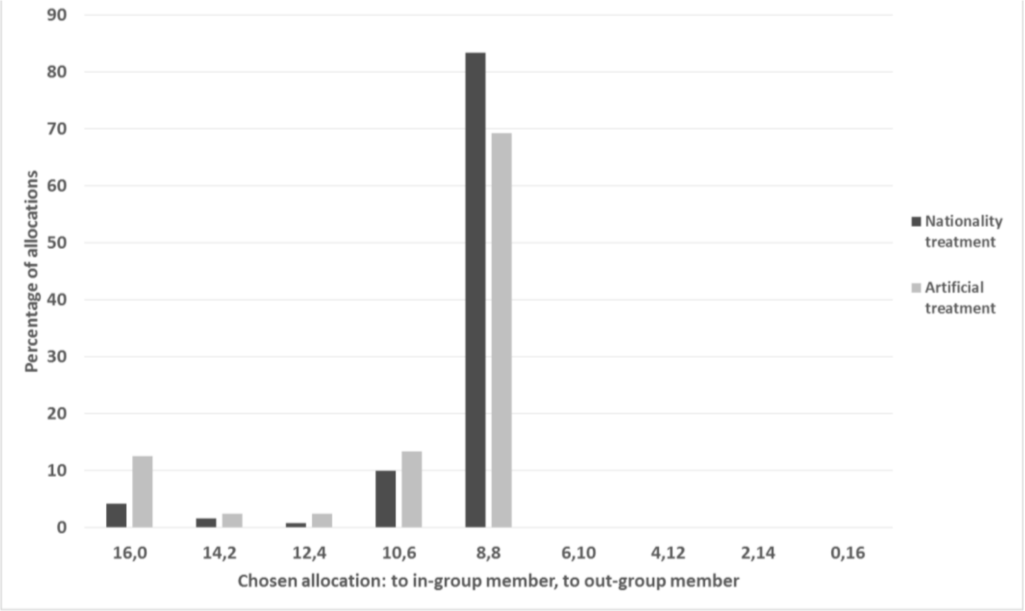
Treatment differences – discrimination

Figure 4 shows the distribution of decisions made in the allocator game in each treatment. In the *Nationality* treatment, 83.3% of participants choose to allocate the money evenly between the in-group member and the out-group member. Only 69.2% of the participants in the *Artificial* treatment make this choice. The remainder of participants in each treatment discriminate against out-group members; no individual in either treatment allocates more money to the out-group member than the in-group member. 12.5% of participants in the *Artificial* treatment allocate all the money to the in-group member, while only 4.2% do so in the *Nationality* treatment.

In the *Nationality* treatment, participants allocate an average of £8.67 to the in-group member and £7.33 to the out-group member, resulting in a mean difference of £1.33. In the *Artificial* treatment, participants allocate an average of £9.52 to the in-group member and £6.48 to the out-group member, resulting in a mean difference of £3.03. A randomisation test indicates that the mean difference in the *Artificial* treatment is significantly higher than that in the *Nationality* treatment (p=0.007). This is consistent with the conjecture that discrimination is stronger in the treatment where it is perceived to be more socially appropriate. It suggests that norm-compliance moderates discriminatory behaviour.

[52] In Supplementary Online Materials E, we analyse cross-national differences in responses to the norm-elicitation task. We show that in both the *Artificial* and *Nationality* treatments, Chinese participants perceived discrimination to be more socially appropriate than British participants.

Figure 4: Discrimination in the allocator game



Notes: Figure 4 shows the percentage of participants in each treatment who choose each allocation. Allocations are denoted by the amount given to the in-group member on the left, and the amount given to the out-group member on the right – e.g. (16,0) denotes allocating £16 to the in-group member and £0 to the out-group member.

In Table 1, an OLS regression confirms that the treatment effect on discrimination is robust to the inclusion of various controls – such as age, gender, nationality and the extent to which participants understand the tasks.[53]

[53] In addition to the regression in Table 1, we ran further models on the British and Chinese subsamples to investigate the effects on discrimination of several other variables which were nationality-specific. These variables were not significant. For the British, we found no significant effect on discrimination of: ethnicity, political persuasion, views on immigration, or hostility towards foreign students. For the Chinese, we found no significant effect of: views towards foreigners in China, feeling welcome in the UK, or hostility towards domestic students. Output is available on request.

Table 1: OLS regressions of treatment differences in discrimination

	Dependent variable = Difference in amount allocated to in-group member and out-group member			
	OLS model		OLS model	
Treatment				
Artificial	1.700***	(0.605)	1.974***	(0.626)
Controls				
Male			0.229	(0.664)
Age			-0.121	(0.229)
Year of study			-0.280	(0.379)
Chinese			2.212***	(0.802)
Misunderstanding			0.875	(0.687)
Rural background			0.637	(0.705)
Economics student			-0.060	(0.775)
Constant	1.333	(0.428)	3.113	(4.003)
R^2	0.032		0.082	
N	240		234	

*Note: *** p<0.01, ** p<0.05, * p<0.1; Standard errors in parentheses. Misunderstanding = number of control questions answered incorrectly at first attempt; Six observations dropped from model with controls owing to missing data for age and year of study.*

Econometric analysis of individual behaviour

So far we have analysed the link between behaviour and norms at the group level, by showing that there is more discrimination in the treatment where it is perceived as less socially inappropriate. We now exploit the within-subject nature of our experiment to extend the analysis to the individual level. Specifically, we investigate whether a model that incorporates a preference for norm compliance is better able to explain the behavioural regularities in our experiment than a model that does not incorporate such a factor.

Following the theoretical framework introduced in section 2, we assume that the utility that allocators derive from choosing allocation x depends on two components, one defined on material payoffs and the other defined on normative prescriptions. We assume that the first component depends on the absolute difference between the material payoffs of the two passive players implied by allocation x. The second component depends on the social appropriateness of the allocation. For allocator i,

$$U_i(a_x) = v\left|\pi_j(a_x) - \pi_k(a_x)\right| + \gamma N_i(a_x)$$

where $\pi_j(a_x)$ and $\pi_k(a_x)$ are the material payoffs that the two passive players j and k receive from allocation x, and $N_i(a_x)$ is the social appropriateness that allocator i ascribes to allocation x, as measured in the norm-elicitation task.[54]

The parameter v captures the weight that the allocator places on the material payoff component of the utility function, while the parameter γ captures the weight that allocators place on norms. Note that the material payoff component of the utility function is blind to the identities of the passive players, and allocations that implement unequal payoffs carry the same weight to utility, regardless of whether the inequality favours the in-group or out-group. Thus, the parameter v simply captures (identity-blind) preferences associated with payoff inequality. In contrast, the normative component of the utility function allows allocations to weigh differently in the utility function depending on the identities of the passive players. Hence, the parameter γ captures the weight that allocators place on a wider array of normative considerations, including both norms of equality and identity-related prescriptions.[55]

Following Gaechter et al. (2013) and Krupka and Weber (2013), we use fixed-effects conditional logit regressions to estimate the weights v and γ on the two components of the utility function. Specifically, we assume that allocators choose allocations

[54] Recall that our experiment delivers, for each allocator i, a measurement of i's perceived social appropriateness of allocation x. This is because each participant in our experiment made decisions in both the allocation task and the norm-elicitation task.

[55] While the allocation that equalises payoffs between passive players is a strict maximum for both the first and second component of the utility function for many participants, this is not true for all. In particular, about 15% of participants in *Artificial* and 3% in *Nationality* do not identify the equal-split as the most socially appropriate allocation. Moreover, for another 7% of participants in *Artificial* and 5% in *Nationality* the equal-split is not a strict maximum for the normative component of the utility function.

following a logit choice rule, whereby the likelihood of choosing each of the nine possible allocations depends on the utility associated with that choice, $U(a_x)$, relative to the utility associated with the alternative allocations:

$$Pr(a = a_x) = \frac{exp\{U(a_x)\}}{\sum_{l=1,\dots,9} exp\{U(a_l)\}}, x = 1, \dots, 9$$

Our objective, here, is to show that a norm-augmented model is better able to capture treatment differences in choices than a model which is identity-blind. Thus, in Table 2 we report the output of two fixed-effects conditional logit models, each estimated using all of the allocation decisions and all of the social appropriateness evaluations generated under either the *Nationality* or the *Artificial* treatment. In the first model we impose the restriction $\gamma = 0$ to the utility function and, thus, estimate a choice model where the decision-maker is purely concerned with identity-blind payoff inequality. In the second model this restriction is removed and utility is allowed to depend on both payoff inequality and wider normative prescriptions.

Table 2: Conditional logit regressions of the likelihood of choosing an action

Dependent variable = 1 if action is chosen; 0 otherwise		
Model	(1)	(2)
υ (weight on payoff inequality)	-0.338***	-0.111***
	(0.021)	(0.028)
γ (weight on normative prescriptions)		1.081***
		(0.119)
Pseudo R^2	0.424	0.511
Bayesian Information Criterion	615.17	531.02
Number of Observations	2,160	2,160

*Note: *** p<0.01, ** p<0.05, * p<0.1; Standard errors are in parentheses.*

The significant negative estimates of v in both models indicate actions that yield larger payoff inequality are less likely to be chosen. The significant positive estimate of γ in model (2) indicates that more appropriate actions are more likely to be chosen. The significant estimate of γ in a model that also includes the v parameter indicates that the normative component of the utility function can explain variation in choice behaviour that cannot be entirely captured by pure (identity-blind) inequality considerations. This also explains why the Bayesian Information Criterion is significantly lower for model (2) than (1) ($p < 0.001$ on a likelihood-ratio test) indicating that the norm-augmented model fits the data significantly better than the model without norms.

The reason why the norm-augmented model performs better is made clear in Figure 5, in which the aggregate action choice rates predicted by each of the models are graphed next to the actual choice rates (as displayed in Figure 4). The left-hand panel of Figure 5 presents the choice rates predicted by model (1). The right-hand panel presents the choice rates predicted by model (2). For ease of comparison, actual choice rates are reproduced in both panels. In each panel, the predicted choice rates (striped bars) and actual choice rates (shaded bars) of the *Nationality* (*Artificial*) treatment are shown in dark (light) grey.

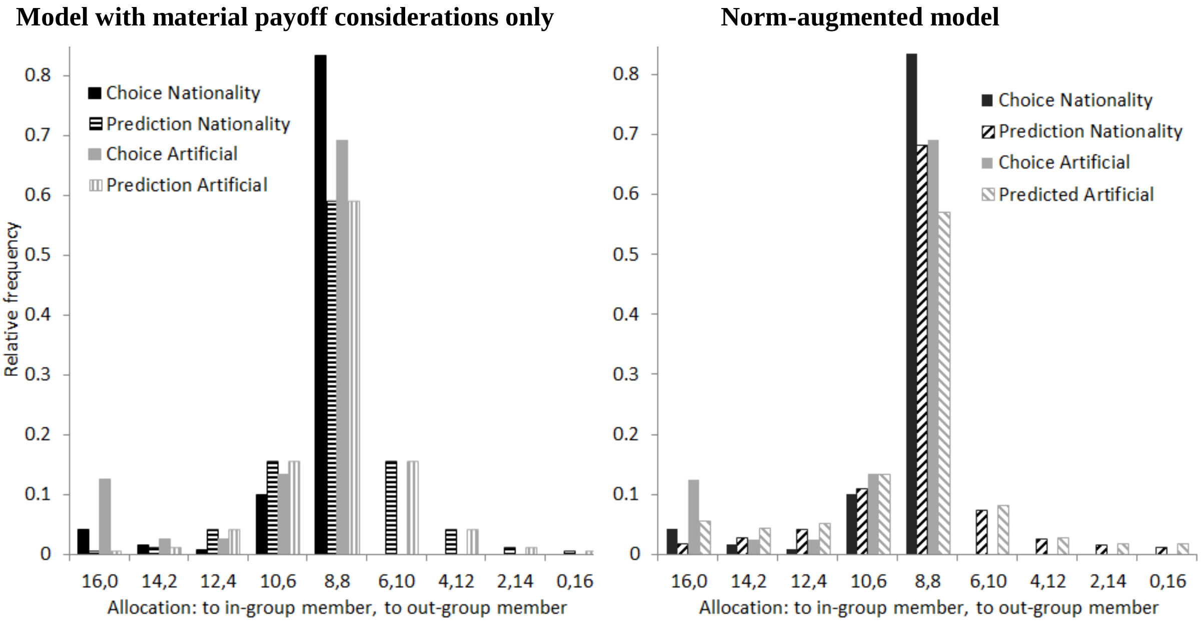
The model in which participants are identity-blind and care only about inequality fails to capture the most important features of the choice data. Most notably, the model is unable to predict any treatment differences in choice, since the implied inequality of an allocation is not different across the *Artificial* and *Nationality* treatments. Moreover, the model predicts that deviations from equality are symmetric across the choice space. That is, the probability of choosing an in-group-favouring allocation is predicted to be the same as that of choosing an allocation which favours the out-group by the equivalent amount. This is not the case in the actual choice data, as no-one chooses out-group-favouring allocations, while 24% of participants choose an in-group-favouring allocation.

In contrast, the norm-augmented model predicts a lower probability of choosing the equal split allocation and higher probabilities of choosing in-group-favouring allocations in the *Artificial* than *Nationality* treatment. This is in line with what we observed in the experiment. Moreover, although the model still assigns positive probabilities to out-group-favouring allocations, it predicts lower probabilities for them than for the comparable in-group-favouring allocations.

5. Conclusion

We show that discrimination is perceived to be socially inappropriate. However, the extent of this perceived inappropriateness depends on the identities upon which discrimination is based: when the identities are defined with reference to a brief, random event, discrimination in favour of the in-

Figure 5: Actual choice rates in the allocator game and choice rates predicted by conditional logits



Notes: Figure 5 shows the percentage of participants in each treatment who choose each allocation, compared to the percentages of participants choosing each allocation in each treatment as predicted by conditional logit models; left-hand panel: model only taking into account considerations for material payoffs, right-hand panel: model augmented by normative considerations; allocations are denoted by the amount given to the in-group member followed by amount given to the out-group member – e.g. (16,0) denotes allocating £16 to the in-group member and £0 to the out-group member.

group is viewed as more appropriate than when the identities are based on nationality. Furthermore, we show that discrimination in the allocator game is stronger in the setting where it is perceived to be more appropriate, and the econometric analysis confirms that the differences in perceived appropriateness predict actual behaviour.

These findings are strongly supportive of Akerlof and Kranton (2000, 2005) providing a useful framework within which to think about and model taste-based discrimination. We offer direct evidence that differences in the way identity groups are defined translate into differences in the perceived normative prescriptions, in choice contexts that are otherwise identical. This offers direct support for Akerlof and Kranton's conjecture that the process of social identification plays a key role in the formation of normative prescriptions.

Acknowledgements

This research was funded by the Economic and Social Research Council via the Network for Integrated Behavioural Sciences (Award No. ES/K002201/1). We are grateful for help provided in running the experiment by Erika Barbosa, Cindy Fu, Xueheng Li and Till Weber. We received helpful comments at presentations at the Universities of Nottingham and East Anglia from, amongst others, Despoina Alempaki, Ben Beranek, Robin Cubitt and Emily Wyman, and we also received helpful written advice from Ilyana Kuziemko.

References

Abbink, K. and D. Harris (2012). In-group favouritism and out-group discrimination in naturally occurring groups, *Technical report, Mimeo, Monash University*.

Ahmed, A. M. (2007). "Group identity, social distance and intergroup bias." *Journal of Economic Psychology* 28(3): 324-337.

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

Akerlof, G. A., & Kranton, R. E. (2005). Identity and the Economics of Organizations. *Journal of Economic Perspectives*, 9-32.

Arrow, K. (1972). Some mathematical models of race discrimination in the labor market, in (A.H. Pascal, ed.), *Racial Discrimination in Economic Life*, pp. 187–204, Lexington, MA: D.C. Heath.

Banerjee, R. (2016). On the interpretation of bribery in a laboratory corruption game: Moral frames and social norms. *Experimental Economics*, *19*(1), 240-267.

Becker, G. S. (1957). *The economics of discrimination*, University of Chicago press.

Benabou, R., & Tirole, J. (2011). *Laws and norms* (No. w17579). National Bureau of Economic Research.

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)*, 289-300.

Bernhard, H., Fehr, E., & Fischbacher, U. (2006a). Group affiliation and altruistic norm enforcement. *American Economic Review, 96*(2), 217–221.

Blanchard, F. A., Crandall, C. S., Brigham, J. C., & Vaughn, L. A. (1994). Condemning and condoning racism: A social context approach to interracial settings. *Journal of Applied Psychology*, *79*(6), 993.

Bolton, G. E., & Ockenfels, A. (2000). ERC: A theory of equity, reciprocity, and competition. *American Economic Review*, 166-193.

Buonanno, P., Montolio, D., & Vanin, P. (2009). Does social capital reduce crime?. *Journal of Law and Economics*, *52*(1), 145-170.

Burks, S. V., & Krupka, E. L. (2012). A multimethod approach to identifying norms and normative expectations within a corporate hierarchy: Evidence from the financial services industry. *Management Science*, *58*(1), 203-217.

Chang, D., Chen, R., & Krupka, E. (2015). Social norms and identity dependent preferences. *Unpublished manuscript.*

Chen, Y. and S. X. Li (2009). Group Identity and Social Preferences. *American Economic Review* 99(1): 431-457.

Crandall, C. S., Eshleman, A., & O'Brien, L. (2002). Social norms and the expression and suppression of prejudice: the struggle for internalization. *Journal of Personality and Social Psychology*, *82*(3), 359.

D'Adda, G., Drouvelis, M., & Nosenzo, D. (2016). Norm elicitation in within-subject designs: Testing for order effects. *Journal of Behavioral and Experimental Economics*, *62*, 1-7.

Elster, J. 1989. Social Norms and Economic Theory. *The Journal of Economic Perspectives* 3(4), 99–117.

Erkut, H., Nosenzo, D., & Sefton, M. (2015). Identifying social norms using coordination games: Spectators vs. stakeholders. *Economics Letters*, *130*, 28-31.

Fehr, E., & Schmidt, K. M. (1999). A theory of fairness, competition, and cooperation. *Quarterly Journal of Economics*, 817-868.

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

Gächter, S., Nosenzo, D., & Sefton, M. (2013). PEER EFFECTS IN PRO-SOCIAL BEHAVIOR: SOCIAL NORMS OR SOCIAL PREFERENCES? *Journal of the European Economic Association*, *11*(3), 548-573.

Goette, L., Huffmann, D., & Meier, S. (2006). The impact of group membership on cooperation and norm enforcement: Evidence using random assignment to real social groups. *American Economic Review,* 96(2), 212–216.

Google Trends. (2016). (https://www.google.co.uk/trends/explore#q=nigger%20jokes&geo=US&cmpt=q&tz=Etc%2FGMT-1) (accessed June 8, 2016)

Greiner, B. (2015) Subject pool recruitment procedures: organizing experiments with ORSEE. *Journal of the Economic Science Association*, 1-12.

Guillen, P., & Ji, D. (2011). Trust, discrimination and acculturation: Experimental evidence on Asian international and Australian domestic university students. *The Journal of Socio-Economics*, *40*(5), 594-608.

Hargreaves Heap, S. P. and D. J. Zizzo (2009). "The value of groups." The *American Economic Review*: 295-323.

Harris, D., Herrmann, B., Kontoleon, A., & Newton, J. (2014). Is it a norm to favour your own group?. *Experimental Economics*, 1-31.

Hennig-Schmidt, H., Selten, R., Walkowitz, G., Winter, E., & Dakkak, I. (2007). Actions and Beliefs in a Trilateral Trust Game Involving Germans, Israelis and Palestinians. *Unpublished manuscript*.

Huang, P. H., & Wu, H. M. (1994). More order without more law: A theory of social norms and organizational cultures. *Journal of Law, Economics and Organization*, *10,* 390.

Jetten, J., Spears, R., & Manstead, A. S. (1996). Intergroup norms and intergroup discrimination: distinctive self-categorization and social identity effects. *Journal of Personality and Social Psychology*, *71*(6), 1222.

Kaiser, J., & Lacy, M. G. (2009). A general-purpose method for two-group randomization tests. *Stata Journal*, *9*(1), 70.

Krupka, E. L., Leider, S., & Jiang, M. (2016). A meeting of the minds: informal agreements and social norms. *Management Science*.

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.

Kübler, D. (2001). On the regulation of social norms. *Journal of Law, Economics, and Organization*, *17*(2), 449-476.

Lane, T. (forthcoming). Discrimination in the laboratory: a meta-analysis. *European Economic Review (forthcoming).*

Lazzarini, S. G., Miller, G. J., & Zenger, T. R. (2004). Order with some law: Complementarity versus substitution of formal and informal arrangements.*Journal of Law, Economics, and Organization*, *20*(2), 261-298.

Mackie, G., Moneti, F., & Shakya, H. (2015). What are Social Norms? How are They Measured? *UNICEF discussion paper, www.unicef.org/protection/files/4_09_30_Whole_What_are_Social_Norms.pdf*

Moir, R. (1998). A Monte Carlo analysis of the Fisher randomization technique: reviving randomization for experimental economists. *Experimental Economics*,*1*(1), 87-100.

Montgomery, J. D. (1994). Revisting Tally's Corner Mainstream Norms, Cognitive Dissonance, and Underclass Behavior. *Rationality and Society*, *6*(4), 462-488.

National Records of Scotland. (2011). 2011 Census: Aggregate data (Scotland) [computer file]. *UK Data Service Census Support. Downloaded from: http://infuse.ukdataservice.ac.uk.*

Nesdale, D., Maass, A., Durkin, K., & Griffiths, J. (2005). Group norms, threat, and children's racial prejudice. *Child Development*, *76*(3), 652-663.

Netzer, R. J., & Sutter, M. (2009). Intercultural trust. An experiment in Austria and Japan (No. 2009-05). *Working Papers in Economics and Statistics*.

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

Posner, E. A. (2009). *Law and social norms*. Harvard university press.

Scottish Executive Statistical Bulletin. (2007). RACIST INCIDENTS RECORDED BY THE POLICE IN SCOTLAND, 2003/04 TO 2005/06. *Criminal Justice Series.*

Sunstein, C. R. (1990). Norms in Surprising Places: The Case of Statutory Interpretation. *Ethics*, *100*(4), 803-820.

Tajfel, H., Billig, M. G., Bundy, R. P., & Flament, C. (1971). Social categorization and intergroup behavior. *European Journal of Social Psychology*,*1*(2), 149-178.

United States Department of Justice, Federal Bureau of Investigation. (2005). *Hate Crime Statistics, 2004*. Retrieved (8th June, 2016), from (https://www2.fbi.gov/ucr/hc2004/hctable1.htm).

United States Department of Justice, Federal Bureau of Investigation. (2006). *Hate Crime Statistics, 2005*. Retrieved (8th June, 2016), from (https://www.fbi.gov/about-us/cjis/ucr/hate-crime/2005).

United States Department of Justice, Federal Bureau of Investigation. (2007). *Hate Crime Statistics, 2006*. Retrieved (8th June, 2016), from (https://www2.fbi.gov/ucr/hc2006/table1.html).

United States Department of Justice, Federal Bureau of Investigation. (2008). *Hate Crime Statistics, 2007*. Retrieved (8th June, 2016), from (https://www.fbi.gov/about-us/cjis/ucr/hate-crime/2007).

United States Department of Justice, Federal Bureau of Investigation. (2009). *Hate Crime Statistics, 2008*. Retrieved (8th June, 2016), from (https://www2.fbi.gov/ucr/hc2008/data/table_01.html).

United States Department of Justice, Federal Bureau of Investigation. (2010). *Hate Crime Statistics, 2009*. Retrieved (8th June, 2016), from (https://www2.fbi.gov/ucr/hc2009/data/table_01.html).

United States Department of Justice, Federal Bureau of Investigation. (2011). *Hate Crime Statistics, 2010*. Retrieved (8th June, 2016), from (https://www.fbi.gov/about-us/cjis/ucr/hate-crime/2010/tables/table-1-incidents-offenses-victims-and-known-offenders-by-bias-motivation-2010.xls).

United States Department of Justice, Federal Bureau of Investigation. (2012). *Hate Crime Statistics, 2011*. Retrieved (8th June, 2016), from (https://www.fbi.gov/about-us/cjis/ucr/hate-crime/2011/tables/table-1).

United States Department of Justice, Federal Bureau of Investigation. (2013). *Hate Crime Statistics, 2012*. Retrieved (8th June, 2016), from (https://www.fbi.gov/about-us/cjis/ucr/hate-crime/2012/tables-and-data-declarations/1tabledatadecpdf/table_1_incidents_offenses_victims_and_known_offenders_by_bias_motivation_2012.xls).

United States Department of Justice, Federal Bureau of Investigation. (2014). *Hate Crime Statistics, 2013*. Retrieved (8th June, 2016), from (https://www.fbi.gov/about-us/cjis/ucr/hate-crime/2013/tables/1tabledatadecpdf/table_1_incidents_offenses_victims_and_known_offenders_by_bias_motivation_2013.xls).

United States Department of Justice, Federal Bureau of Investigation. (2015). *Hate Crime Statistics, 2014*. Retrieved (8th June, 2016), from (https://www.fbi.gov/about-us/cjis/ucr/hate-crime/2014/tables/table-1).

Willinger, M., et al. (2003). A comparison of trust and reciprocity between France and Germany: Experimental investigation based on the investment game. *Journal of Economic Psychology* 24(4): 447-466.

Zitek, E. M., & Hebl, M. R. (2007). The role of social norm clarity in the influenced expression of prejudice over time. *Journal of Experimental Social Psychology*, *43*(6), 867-876.

Supplementary Online Materials

A: Experimental instructions

A.1 Instructions for subjects in the Nationality treatment, playing allocator game first

Instructions

Welcome to this experiment. This is an experiment about decision-making. During the experiment, we request that you remain quiet and do not attempt to communicate with other participants. Participants not following this request may be asked to leave without receiving payment. If you have any questions, please raise your hand and the experimenter will come to you. For your participation, you will be paid a show-up fee of £4. You may also receive some additional money based on your choices and the choices of others in the tasks described below.

There will be two tasks for all participants to perform. At the end of the experiment, the experimenter will toss a fair coin. If it lands on heads, all participants will receive payment for the first task only; if it lands on tails, all participants will receive payment for the second task only. As you will not know until the end of the experiment which task you will receive payment for, *please make your decisions in each task carefully*. You will not receive feedback on the outcome of any task until the end of the experiment, and your decisions in the first task will have no effect on the nature or outcome of the second task. You will not receive any instructions for or information about the second task until you have completed the first task. After the second task, there will also be a questionnaire. The anonymity of your responses to all parts of all tasks and questions is guaranteed.

Please now answer two questions on your screen, to ensure you understand the process of the experiment.

In this experiment, the room has been divided into two sections on the basis of nationality. On one side everyone is British; on the other side everyone is Chinese. The sign on your desk reminds you whether you are on the British or Chinese side of the room.

Task One

In this experiment, one third of you will be randomly assigned by the computer into a role entitled ‘Individual A’. The decisions made by Individual As during the task will determine the payments from the task received by the other two thirds of participants. Each of you has an equal chance of being an Individual A. Exactly who the Individual As are will not be revealed until the end of the experiment. In the meantime, we ask all participants to make a decision **as if** they are an Individual A.

Please make your decision carefully, as it may be used to determine participants’ payments.

Assume for the rest of this paragraph that you are an Individual A. Your task will be to decide how to divide £16 between two other participants in the experiment, one who has the same nationality as you, and another who has a different nationality from you. You may divide the money any way you like so long as the amount allocated to each person is a multiple of two. You may not allocate any of the money to yourself. However, you will also receive a payment. This payment might be £6, £8 or £10. This will be randomly decided at the end of the experiment by the computer, which is equally likely to select any of these amounts.

Please now answer two questions on your screen, to ensure you understand this part of the experiment.

Task Two

In the second part of this experiment, you will receive a description of a situation. This description corresponds to a situation in which one person, “Individual A,” must decide how to act. You will be given a description of various possible actions Individual A can choose to take.

After you receive the description of the situation, you will be asked to evaluate each of the various possible actions Individual A can choose to take. You must indicate, for each of the possible actions, whether taking that action would be "socially appropriate" or "socially inappropriate". By socially appropriate, we mean behaviour that you think most participants of your nationality would agree is the "correct" thing to do. Another way to think about what we mean is that if Individual A were to select a socially inappropriate action, then another participant of your nationality might be angry at Individual A.

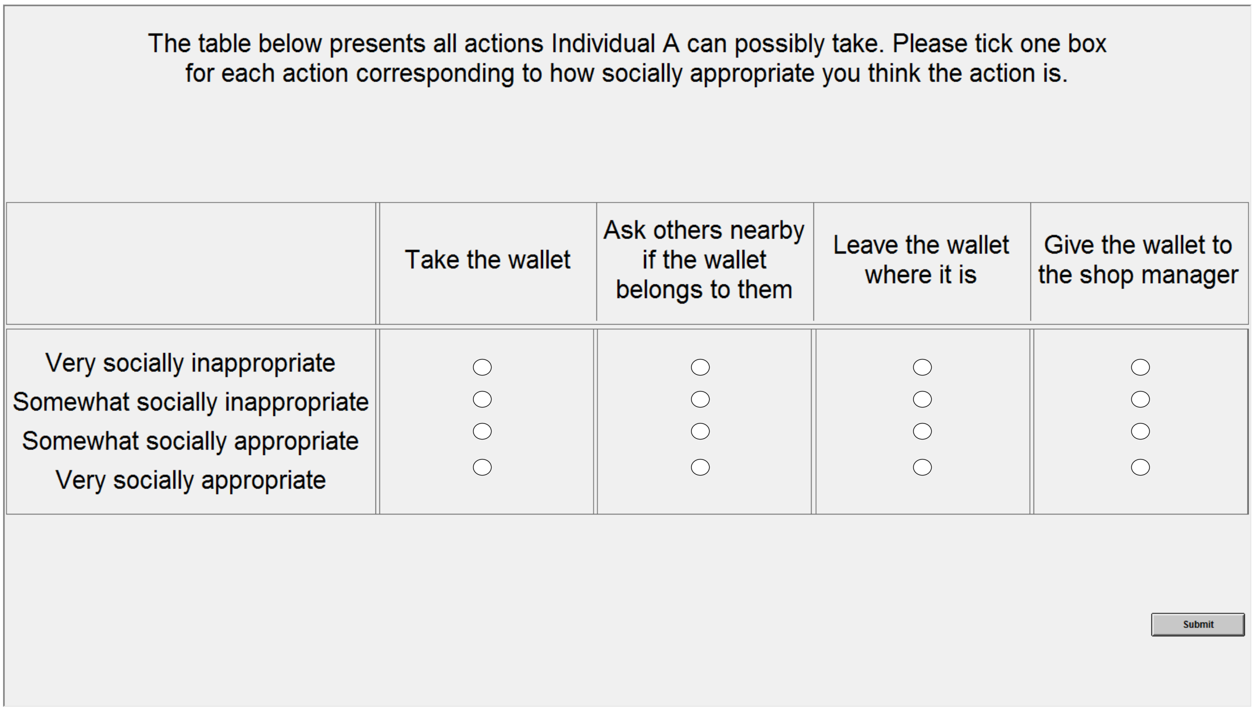
In each of your responses, we would like you to answer as truthfully as possible, based on your opinions of what constitutes socially appropriate or socially inappropriate behaviour.

To give you an idea of how the experiment will proceed, we will go through an example situation and show you how you will indicate your responses.

Example Situation

Individual A is at a local coffee shop near campus. While there, Individual A notices that someone has left a wallet at one of the tables. Individual A must decide what to do. Individual A can choose four possible actions: take the wallet, ask others nearby if the wallet belongs to them, leave the wallet where it is, or give the wallet to the shop manager.

The table below presents the list of the possible actions Individual A can choose. For each of the actions, you would be asked to indicate whether you believe choosing that action is very socially inappropriate, somewhat socially inappropriate, somewhat socially appropriate, or very socially appropriate. To indicate your response, you would click on the corresponding button.

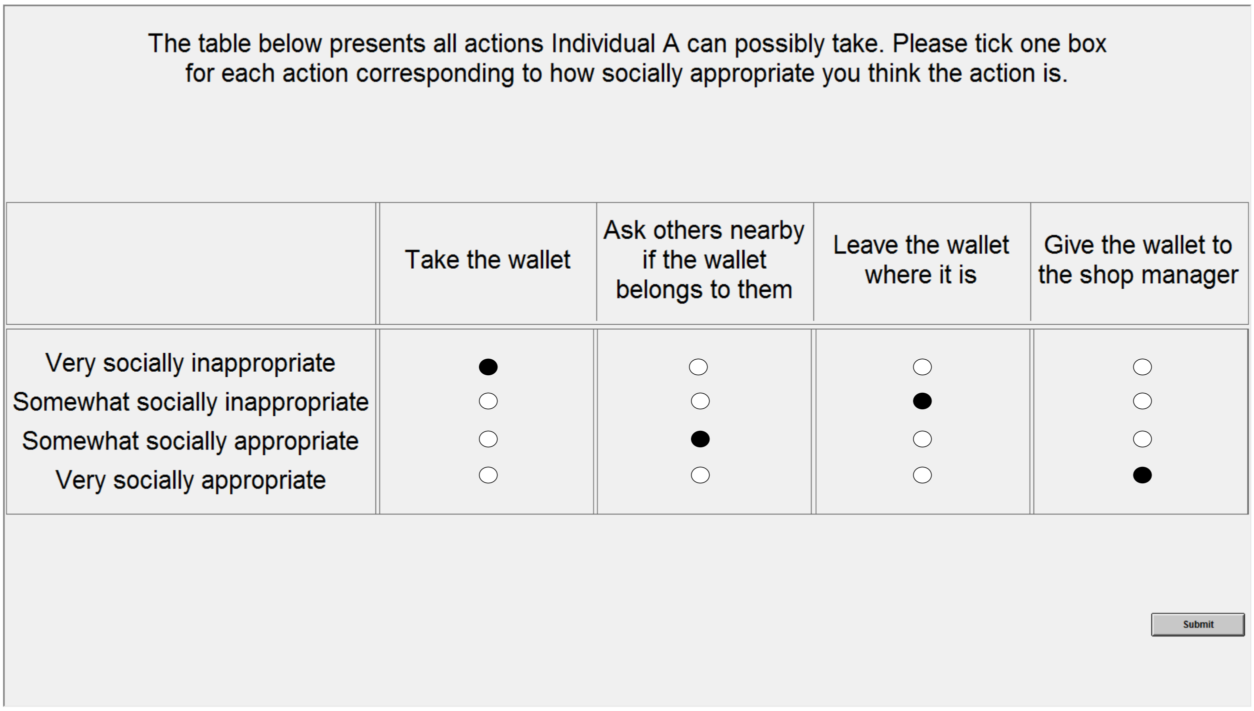


The table below presents all actions Individual A can possibly take. Please tick one box for each action corresponding to how socially appropriate you think the action is.

	Take the wallet	Ask others nearby if the wallet belongs to them	Leave the wallet where it is	Give the wallet to the shop manager
Very socially inappropriate	○	○	○	○
Somewhat socially inappropriate	○	○	○	○
Somewhat socially appropriate	○	○	○	○
Very socially appropriate	○	○	○	○

If this was the situation for this study, you would consider each of the possible actions above and, for that action, indicate the extent to which you believe taking that action would be "socially appropriate" or "socially inappropriate". Recall that by socially appropriate we mean behaviour that most participants of your nationality agree is the "correct" thing to do.

For example, suppose you thought that taking the wallet was very socially inappropriate, asking others nearby if the wallet belongs to them was somewhat socially appropriate, leaving the wallet where it is was somewhat socially inappropriate, and giving the wallet to the shop manager was very socially appropriate. Then you would indicate your responses as follows:



The table below presents all actions Individual A can possibly take. Please tick one box for each action corresponding to how socially appropriate you think the action is.

	Take the wallet	Ask others nearby if the wallet belongs to them	Leave the wallet where it is	Give the wallet to the shop manager
Very socially inappropriate	●	○	○	○
Somewhat socially inappropriate	○	○	●	○
Somewhat socially appropriate	○	●	○	○
Very socially appropriate	○	○	○	●

If you have any questions about this example situation or about how to indicate your responses, please raise your hand now.

You will next be given the description of a situation where Individual A, a participant in an experiment, has to choose between various possible actions. After you read the description, you must consider the possible actions and indicate on your computer screen how socially appropriate these are in a table similar to the one shown above for the example situation.

After this, the computer will randomly select one participant of your nationality (that is, it will select a British participant if you are British, or a Chinese participant if you are Chinese). The computer will then randomly select one action Individual A can choose. Your evaluation of this action will be compared with that of the randomly selected participant of your nationality. If your evaluation is the same as theirs, you will receive £8 for this task; otherwise you will receive zero.

For instance, imagine the example situation above was the actual situation and the possible action "Leave the wallet where it is" was selected by the computer. If your evaluation had been "somewhat socially inappropriate" then your task earnings would be £8 if the person you are matched with also evaluated the action as “somewhat socially inappropriate” and zero otherwise.

<u>The situation</u>

The situation you are asked to evaluate is like the one you participated in in the previous task. Here is a summary.

Individual A is taking part in an experiment in this lab. The room has been divided into two sections on the basis of nationality. On one side everyone is British; on the other side everyone is Chinese. The anonymity of Individual A's decisions in the experiment is guaranteed.

Individual A's task will be to decide how to divide £16 between two other participants in the experiment, one who has the same nationality as Individual A, and another who has a different nationality from Individual A. Individual A may divide the money any way he or she likes so long as the amount allocated to each person is a multiple of two. Individual A may not allocate any of the money to his- or herself. However, Individual A will also receive a payment. This payment might be £6, £8 or £10. This will be randomly decided at the end of the experiment by the computer, which is equally likely to select any of these amounts.

A.2. Instructions for subjects in the Nationality treatment, taking the norm-elicitation task first

<u>Instructions</u>

Welcome to this experiment. This is an experiment about decision-making. During the experiment, we request that you remain quiet and do not attempt to communicate with other participants. Participants not following this request may be asked to leave without receiving payment. If you have any questions, please raise your hand and the experimenter will come to you. For your participation, you will be paid a show-up fee of £4. You may also receive some additional money based on your choices and the choices of others in the tasks described below.

There will be two tasks for all participants to perform. At the end of the experiment, the experimenter will toss a fair coin. If it lands on heads, all participants will receive payment for the first task only; if it lands on tails, all participants will receive payment for the second task only. As you will not know until the end of the experiment which task you will receive payment for, *please make your decisions in each task carefully*. You will not receive feedback on the outcome of any task until the end of the experiment, and your decisions in the first task will have no effect on the nature or outcome of the second task. You will not receive any instructions for or information about the second task until you have completed the first task. After the second task, there will also be a questionnaire. The anonymity of your responses to all parts of all tasks and questions is guaranteed.

Please now answer two questions on your screen, to ensure you understand the process of the experiment.

In this experiment, the room has been divided into two sections on the basis of nationality. On one side everyone is British; on the other side everyone is Chinese. The sign on your desk reminds you whether you are on the British or Chinese side of the room.

Task One

In the first part of this experiment, you will receive a description of a situation. This description corresponds to a situation in which one person, “Individual A,” must decide how to act. You will be given a description of various possible actions Individual A can choose to take.

After you receive the description of the situation, you will be asked to evaluate each of the various possible actions Individual A can choose to take. You must indicate, for each of the possible actions, whether taking that action would be "socially appropriate" or "socially inappropriate". By socially appropriate, we mean behaviour that you think most participants of your nationality would agree is the "correct" thing to do. Another way to think about what we mean is that if Individual A were to select a socially inappropriate action, then another participant of your nationality might be angry at Individual A.

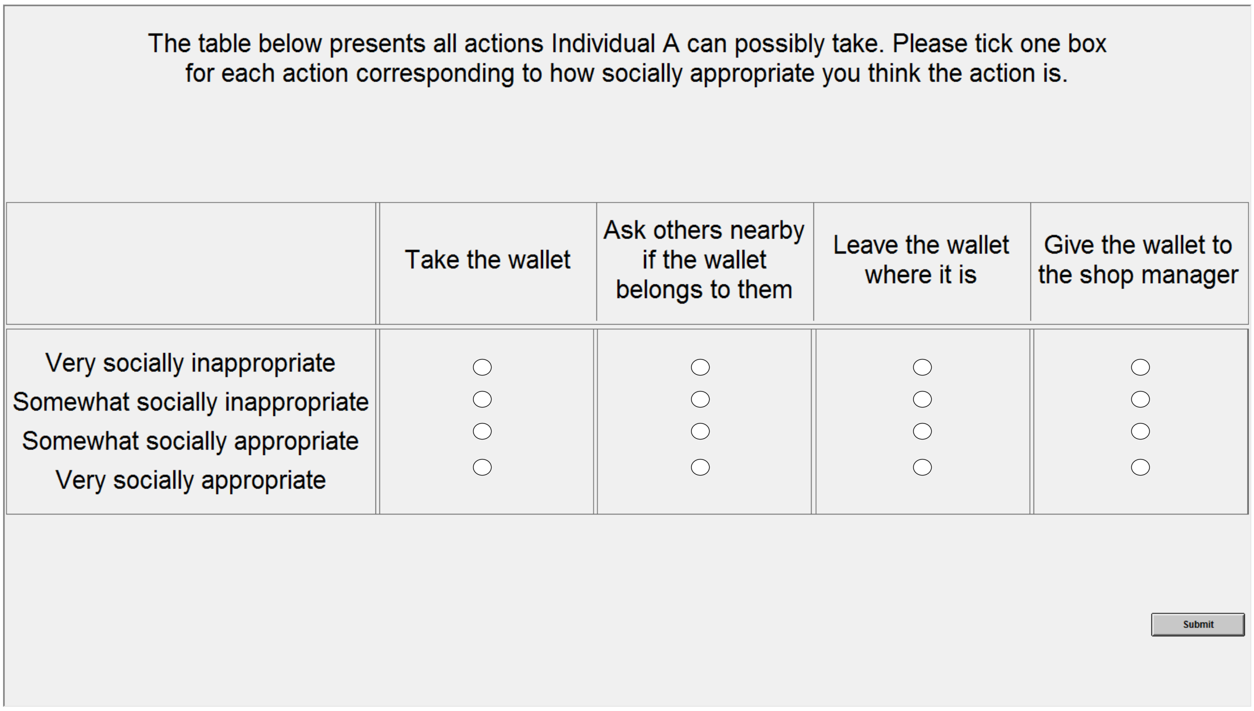
In each of your responses, we would like you to answer as truthfully as possible, based on your opinions of what constitutes socially appropriate or socially inappropriate behaviour.

To give you an idea of how the experiment will proceed, we will go through an example situation and show you how you will indicate your responses.

Example Situation

Individual A is at a local coffee shop near campus. While there, Individual A notices that someone has left a wallet at one of the tables. Individual A must decide what to do. Individual A can choose four possible actions: take the wallet, ask others nearby if the wallet belongs to them, leave the wallet where it is, or give the wallet to the shop manager.

The table below presents the list of the possible actions Individual A can choose. For each of the actions, you would be asked to indicate whether you believe choosing that action is very socially inappropriate, somewhat socially inappropriate, somewhat socially appropriate, or very socially appropriate. To indicate your response, you would click on the corresponding button.

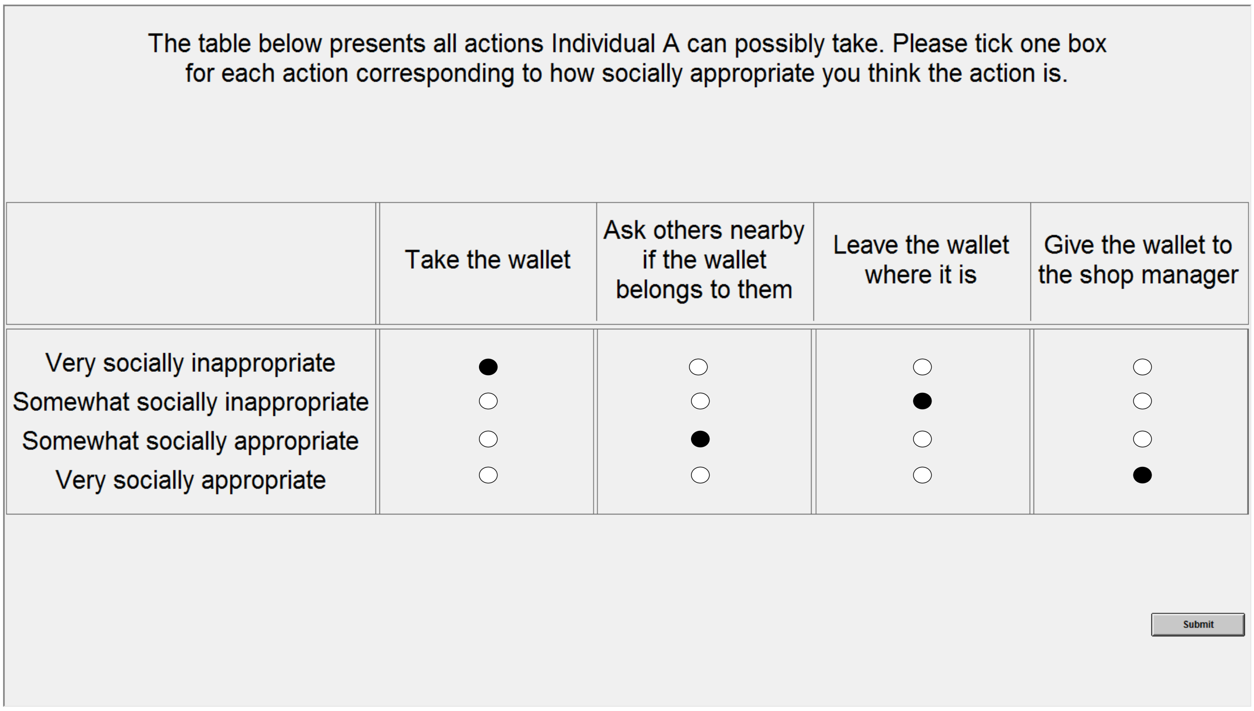


The table below presents all actions Individual A can possibly take. Please tick one box for each action corresponding to how socially appropriate you think the action is.

	Take the wallet	Ask others nearby if the wallet belongs to them	Leave the wallet where it is	Give the wallet to the shop manager
Very socially inappropriate	○	○	○	○
Somewhat socially inappropriate	○	○	○	○
Somewhat socially appropriate	○	○	○	○
Very socially appropriate	○	○	○	○

If this was the situation for this study, you would consider each of the possible actions above and, for that action, indicate the extent to which you believe taking that action would be "socially appropriate" or "socially inappropriate". Recall that by socially appropriate we mean behaviour that most participants of your nationality agree is the "correct" thing to do.

For example, suppose you thought that taking the wallet was very socially inappropriate, asking others nearby if the wallet belongs to them was somewhat socially appropriate, leaving the wallet where it is was somewhat socially inappropriate, and giving the wallet to the shop manager was very socially appropriate. Then you would indicate your responses as follows:



The table below presents all actions Individual A can possibly take. Please tick one box for each action corresponding to how socially appropriate you think the action is.

	Take the wallet	Ask others nearby if the wallet belongs to them	Leave the wallet where it is	Give the wallet to the shop manager
Very socially inappropriate	●	○	○	○
Somewhat socially inappropriate	○	○	●	○
Somewhat socially appropriate	○	●	○	○
Very socially appropriate	○	○	○	●

If you have any questions about this example situation or about how to indicate your responses, please raise your hand now.

You will next be given the description of a situation where Individual A, a participant in an experiment, has to choose between various possible actions. After you read the description, you must consider the possible actions and indicate on your computer screen how socially appropriate these are in a table similar to the one shown above for the example situation.

After this, the computer will randomly select one participant of your nationality (that is, it will select a British participant if you are British, or a Chinese participant if you are Chinese). The computer will then randomly select one action Individual A can choose. Your evaluation of this action will be compared with that of the randomly selected participant of your nationality. If your evaluation is the same as theirs, you will receive £8 for this task; otherwise you will receive zero.

For instance, imagine the example situation above was the actual situation and the possible action "Leave the wallet where it is" was selected by the computer. If your evaluation had been "somewhat socially inappropriate" then your task earnings would be £8 if the person you are matched with also evaluated the action as “somewhat socially inappropriate” and zero otherwise.

The situation

Individual A is taking part in an experiment in this lab. The room has been divided into two sections on the basis of nationality. On one side everyone is British; on the other side everyone is Chinese. The anonymity of Individual A's decisions in the experiment is guaranteed.

Individual A's task will be to decide how to divide £16 between two other participants in the experiment, one who has the same nationality as Individual A, and another who has a different nationality from Individual A. Individual A may divide the money any way he or she likes so long as the amount allocated to each person is a multiple of two. Individual A may not allocate any of the money to his- or herself. However, Individual A will also receive a payment. This payment might be £6, £8 or £10. This will be randomly decided at the end of the experiment by the computer, which is equally likely to select any of these amounts.

Please now answer one question on your screen, to ensure you understand this situation.

Task Two

In this experiment, one third of you will be randomly assigned by the computer into a role entitled 'Individual A'. The decisions made by Individual As during the task will determine the payments from the task received by the other two thirds of participants. Each of you has an equal chance of being an Individual A. Exactly who the Individual As are will not be revealed until the end of the experiment. In the meantime, we ask all participants to make a decision **as if** they are an Individual A.

Please make your decision carefully, as it may be used to determine participants' payments.

Assume for the rest of this paragraph that you are an Individual A. Your task is like the one you evaluated in the previous task. Your task will be to decide how to divide £16 between two other participants in the experiment, one who has the same nationality as you, and another who has a different nationality from you. You may divide the money any way you like so long as the amount allocated to each person is a multiple of two. You may not allocate any of the money to yourself. However, you will also receive a payment. This payment might be £6, £8 or £10. This will be randomly decided at the end of the experiment by the computer, which is equally likely to select any of these amounts.

You will next see one question on your screen. Please answer it to ensure you understand this part of the experiment.

A.3. Instructions for subjects in the Artificial treatment, playing the allocator game first

Instructions

Welcome to this experiment. This is an experiment about decision-making. During the experiment, we request that you remain quiet and do not attempt to communicate with other participants. Participants not following this request may be asked to leave without receiving payment. If you have any questions, please raise your hand and the experimenter will come to you. For your participation, you will be paid a show-up fee of £4. You may also receive some additional money based on your choices and the choices of others in the tasks described below.

There will be two tasks for all participants to perform. At the end of the experiment, the experimenter will toss a fair coin. If it lands on heads, all participants will receive payment for the first task only; if it lands on tails, all participants will receive payment for the second task only. As you will not know until the end of the experiment which task you will receive payment for, *please make your decisions in each task carefully*. You will not receive feedback on the outcome of any task until the end of the experiment, and your decisions in the first task will have no effect on the nature or outcome of the second task. You will not receive any instructions for or information about the second task until you have completed the first task. After the second task, there will also be a questionnaire. The anonymity of your responses to all parts of all tasks and questions is guaranteed.

Please now answer two questions on your screen, to ensure you understand the process of the experiment.

In this experiment, the room has been divided into two sections on the basis of which colour of ball you drew from the bag at the beginning of the experiment. On one side everyone drew a green ball; on the other side everyone drew a yellow ball. The sign on your desk reminds you whether you are on the green or yellow side of the room.

Task One

In this experiment, one third of you will be randomly assigned by the computer into a role entitled 'Individual A'. The decisions made by Individual As during the task will determine the payments from the task received by the other two thirds of participants. Each of you has an equal chance of being an Individual A. Exactly who the Individual As are will not be revealed until the end of the experiment. In the meantime, we ask all participants to make a decision **as if** they are an Individual A.

Please make your decision carefully, as it may be used to determine participants' payments.

Assume for the rest of this paragraph that you are an Individual A. Your task will be to decide how to divide £16 between two other participants in the experiment, one who drew the same ball colour as you, and another who drew a different ball colour from you. You may divide the money any way you like so long as the amount allocated to each person is a multiple of two. You may not allocate any of the money to yourself. However, you will also receive a payment. This payment might be £6, £8 or £10. This will be randomly decided at the end of the experiment by the computer, which is equally likely to select any of these amounts.

Please now answer two questions on your screen, to ensure you understand this part of the experiment.

Task Two

In the second part of this experiment, you will receive a description of a situation. This description corresponds to a situation in which one person, “Individual A,” must decide how to act. You will be given a description of various possible actions Individual A can choose to take.

After you receive the description of the situation, you will be asked to evaluate each of the various possible actions Individual A can choose to take. You must indicate, for each of the possible actions, whether taking that action would be "socially appropriate" or "socially inappropriate". By socially appropriate, we mean behaviour that you think most participants who drew your ball colour would agree is the "correct" thing to do. Another way to think about what we mean is that if Individual A were to select a socially inappropriate action, then another participant who drew your ball colour might be angry at Individual A.

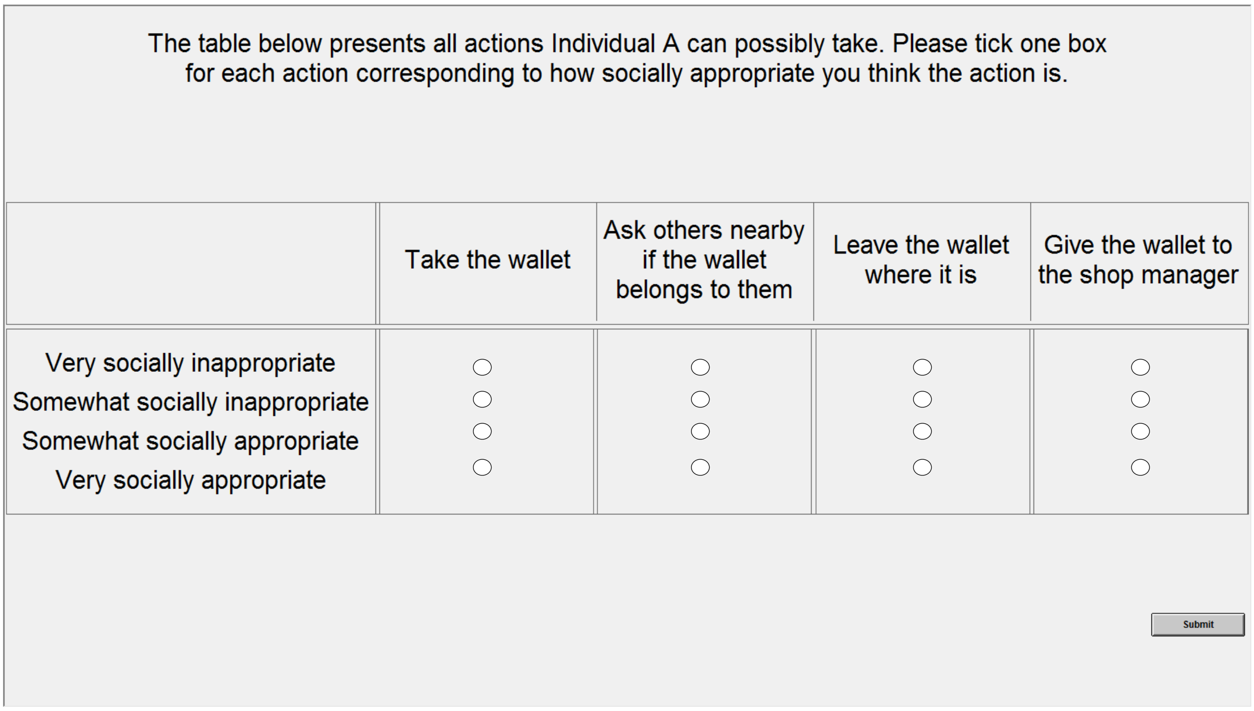
In each of your responses, we would like you to answer as truthfully as possible, based on your opinions of what constitutes socially appropriate or socially inappropriate behaviour.

To give you an idea of how the experiment will proceed, we will go through an example situation and show you how you will indicate your responses.

Example Situation

Individual A is at a local coffee shop near campus. While there, Individual A notices that someone has left a wallet at one of the tables. Individual A must decide what to do. Individual A can choose four possible actions: take the wallet, ask others nearby if the wallet belongs to them, leave the wallet where it is, or give the wallet to the shop manager.

The table below presents the list of the possible actions Individual A can choose. For each of the actions, you would be asked to indicate whether you believe choosing that action is very socially inappropriate, somewhat socially inappropriate, somewhat socially appropriate, or very socially appropriate. To indicate your response, you would click on the corresponding button.

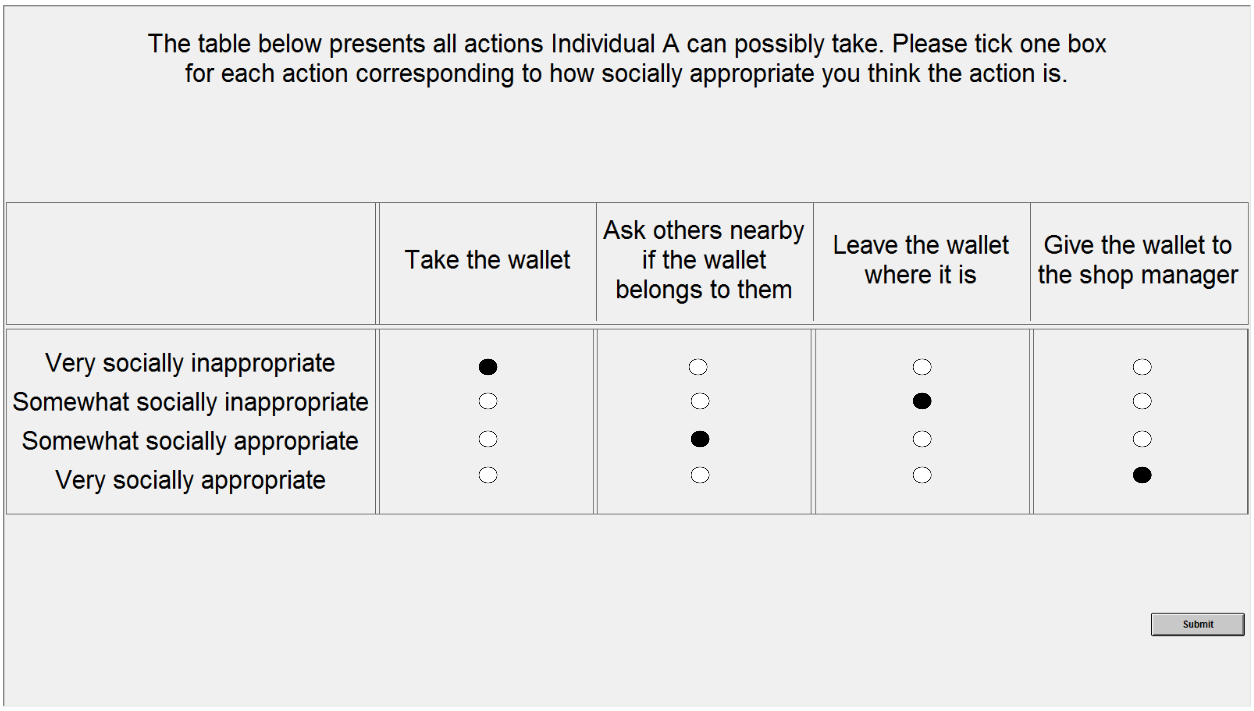


The table below presents all actions Individual A can possibly take. Please tick one box for each action corresponding to how socially appropriate you think the action is.

	Take the wallet	Ask others nearby if the wallet belongs to them	Leave the wallet where it is	Give the wallet to the shop manager
Very socially inappropriate	○	○	○	○
Somewhat socially inappropriate	○	○	○	○
Somewhat socially appropriate	○	○	○	○
Very socially appropriate	○	○	○	○

If this was the situation for this study, you would consider each of the possible actions above and, for that action, indicate the extent to which you believe taking that action would be "socially appropriate" or "socially inappropriate". Recall that by socially appropriate we mean behaviour that most participants who drew your ball colour agree is the "correct" thing to do.

For example, suppose you thought that taking the wallet was very socially inappropriate, asking others nearby if the wallet belongs to them was somewhat socially appropriate, leaving the wallet where it is was somewhat socially inappropriate, and giving the wallet to the shop manager was very socially appropriate. Then you would indicate your responses as follows:



The table below presents all actions Individual A can possibly take. Please tick one box for each action corresponding to how socially appropriate you think the action is.

	Take the wallet	Ask others nearby if the wallet belongs to them	Leave the wallet where it is	Give the wallet to the shop manager
Very socially inappropriate	●	○	○	○
Somewhat socially inappropriate	○	○	●	○
Somewhat socially appropriate	○	●	○	○
Very socially appropriate	○	○	○	●

If you have any questions about this example situation or about how to indicate your responses, please raise your hand now.

You will next be given the description of a situation where Individual A, a participant in an experiment, has to choose between various possible actions. After you read the description, you must consider the possible actions and indicate on your computer screen how socially appropriate these are in a table similar to the one shown above for the example situation.

After this, the computer will randomly select one participant who drew a ball of the same colour as you (that is, it will select a participant who drew a green ball if you drew a green ball, or a participant who drew a yellow ball if you drew a yellow ball). The computer will then randomly select one action Individual A can choose. Your evaluation of this action will be compared with that of the randomly selected participant who drew a ball of the same colour as you. If your evaluation is the same as theirs, you will receive £8 for this task; otherwise you will receive zero.

For instance, imagine the example situation above was the actual situation and the possible action "Leave the wallet where it is" was selected by the computer. If your evaluation had been "somewhat socially inappropriate" then your task earnings would be £8 if the person you are matched with also evaluated the action as “somewhat socially inappropriate” and zero otherwise.

The situation

The situation you are asked to evaluate is like the one you participated in in the previous task. Here is a summary.

Individual A is taking part in an experiment in this lab. The room has been divided into two sections on the basis of which colour of ball participants drew from a bag at the beginning of the experiment. On one side everyone drew a green ball; on the other side everyone drew a yellow ball. The anonymity of Individual A's decisions in the experiment is guaranteed.

Individual A's task will be to decide how to divide £16 between two other participants in the experiment, one who drew the same ball colour as Individual A, and another who drew a different ball colour from Individual A. Individual A may divide the money any way he or she likes so long as the amount allocated to each person is a multiple of two. Individual A may not allocate any of the money to his- or herself. However, Individual A will also receive a payment. This payment might be £6, £8 or £10. This will be randomly decided at the end of the experiment by the computer, which is equally likely to select any of these amounts.

A.4. Instructions for subjects in Artificial treatment, taking norm-elicitation task first

Instructions

Welcome to this experiment. This is an experiment about decision-making. During the experiment, we request that you remain quiet and do not attempt to communicate with other participants. Participants not following this request may be asked to leave without receiving payment. If you have any questions, please raise your hand and the experimenter will come to you. For your participation, you will be paid a show-up fee of £4. You may also receive some additional money based on your choices and the choices of others in the tasks described below.

There will be two tasks for all participants to perform. At the end of the experiment, the experimenter will toss a fair coin. If it lands on heads, all participants will receive payment for the first task only; if it lands on tails, all participants will receive payment for the second task only. As you will not know until the end of the experiment which task you will receive payment for, *please make your decisions in each task carefully*. You will not receive feedback on the outcome of any task until the end of the experiment, and your decisions in the first task will have no effect on the nature or outcome of the second task. You will not receive any instructions for or information about the second task until you have completed the first task. After the second task, there will also be a questionnaire. The anonymity of your responses to all parts of all tasks and questions is guaranteed.

Please now answer two questions on your screen, to ensure you understand the process of the experiment.

In this experiment, the room has been divided into two sections on the basis of which colour of ball you drew from the bag at the beginning of the experiment. On one side everyone drew a green ball; on the other side everyone drew a yellow ball. The sign on your desk reminds you whether you are on the green or yellow side of the room.

Task One

In the first part of this experiment, you will receive a description of a situation. This description corresponds to a situation in which one person, “Individual A,” must decide how to act. You will be given a description of various possible actions Individual A can choose to take.

After you receive the description of the situation, you will be asked to evaluate each of the various possible actions Individual A can choose to take. You must indicate, for each of the possible actions, whether taking that action would be "socially appropriate" or "socially inappropriate". By socially appropriate, we mean behaviour that you think most participants who drew your ball colour would agree is the "correct" thing to do. Another way to think about what we mean is that if Individual A were to select a socially inappropriate action, then another participant who drew your ball colour might be angry at Individual A.

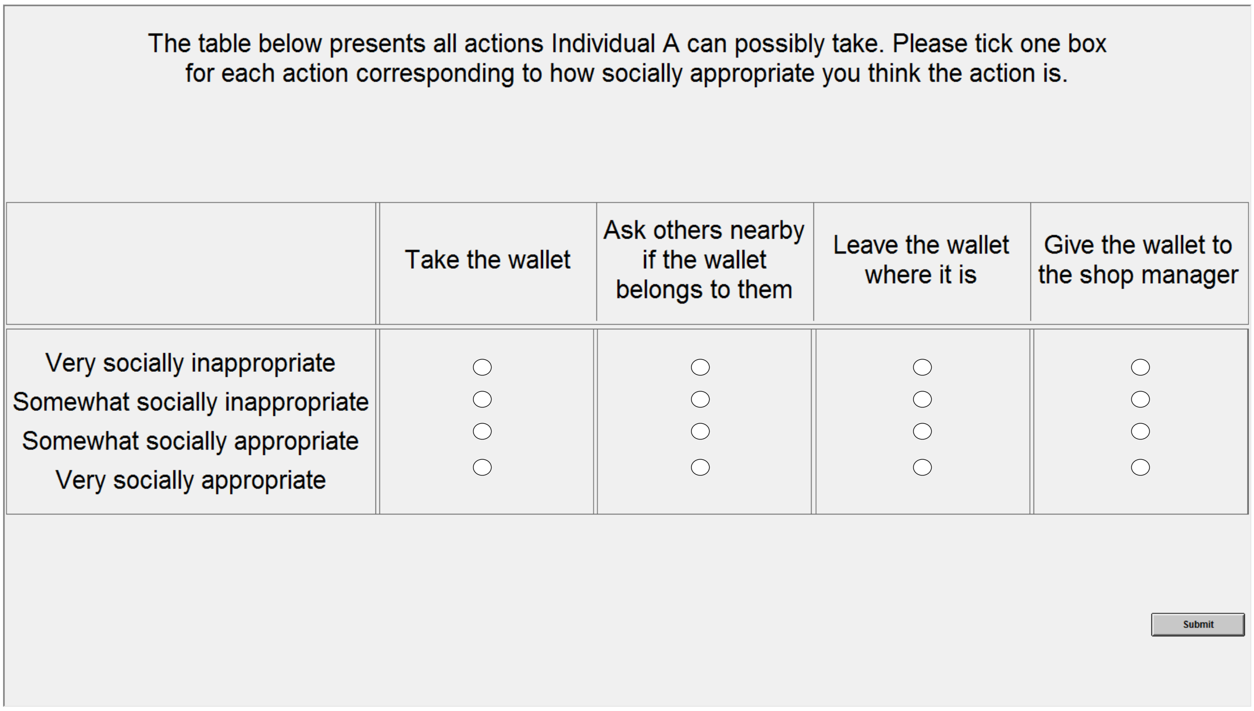
In each of your responses, we would like you to answer as truthfully as possible, based on your opinions of what constitutes socially appropriate or socially inappropriate behaviour.

To give you an idea of how the experiment will proceed, we will go through an example situation and show you how you will indicate your responses.

Example Situation

Individual A is at a local coffee shop near campus. While there, Individual A notices that someone has left a wallet at one of the tables. Individual A must decide what to do. Individual A can choose four possible actions: take the wallet, ask others nearby if the wallet belongs to them, leave the wallet where it is, or give the wallet to the shop manager.

The table below presents the list of the possible actions Individual A can choose. For each of the actions, you would be asked to indicate whether you believe choosing that action is very socially inappropriate, somewhat socially inappropriate, somewhat socially appropriate, or very socially appropriate. To indicate your response, you would click on the corresponding button.

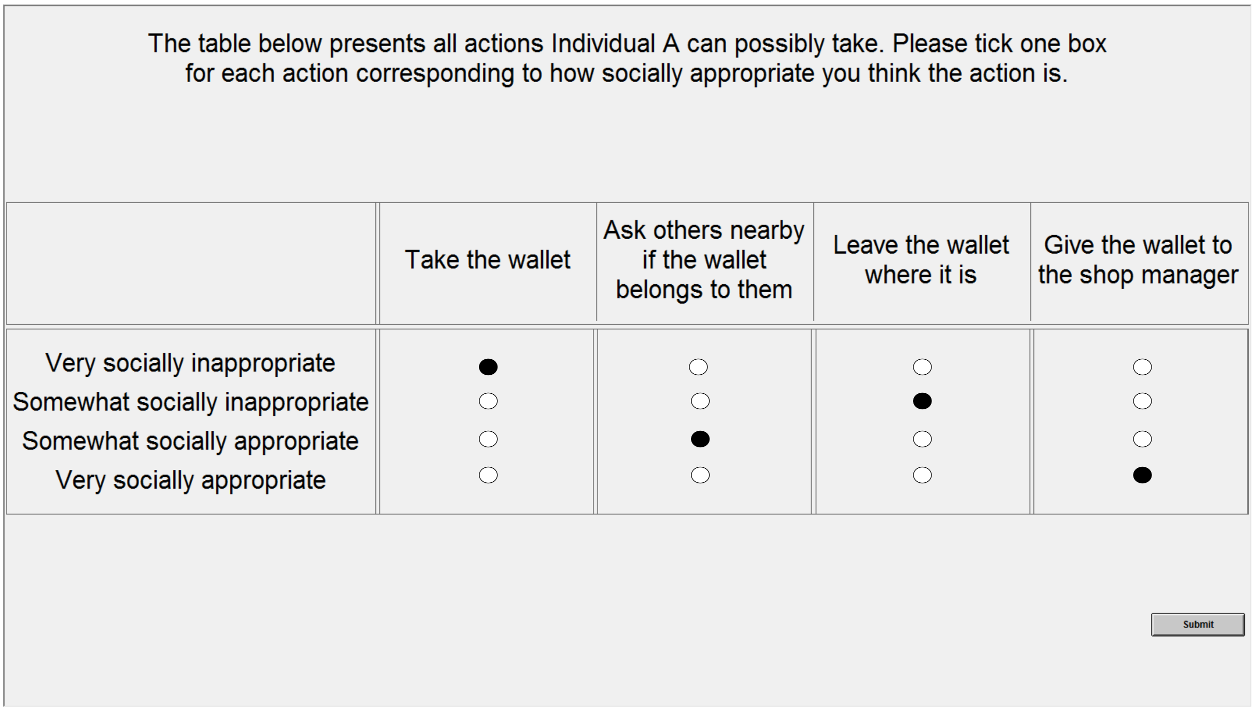


The table below presents all actions Individual A can possibly take. Please tick one box for each action corresponding to how socially appropriate you think the action is.

	Take the wallet	Ask others nearby if the wallet belongs to them	Leave the wallet where it is	Give the wallet to the shop manager
Very socially inappropriate Somewhat socially inappropriate Somewhat socially appropriate Very socially appropriate	○ ○ ○ ○	○ ○ ○ ○	○ ○ ○ ○	○ ○ ○ ○

If this was the situation for this study, you would consider each of the possible actions above and, for that action, indicate the extent to which you believe taking that action would be "socially appropriate" or "socially inappropriate". Recall that by socially appropriate we mean behaviour that most participants who drew your ball colour agree is the "correct" thing to do.

For example, suppose you thought that taking the wallet was very socially inappropriate, asking others nearby if the wallet belongs to them was somewhat socially appropriate, leaving the wallet where it is was somewhat socially inappropriate, and giving the wallet to the shop manager was very socially appropriate. Then you would indicate your responses as follows:



The table below presents all actions Individual A can possibly take. Please tick one box for each action corresponding to how socially appropriate you think the action is.

	Take the wallet	Ask others nearby if the wallet belongs to them	Leave the wallet where it is	Give the wallet to the shop manager
Very socially inappropriate	●	○	○	○
Somewhat socially inappropriate	○	○	●	○
Somewhat socially appropriate	○	●	○	○
Very socially appropriate	○	○	○	●

If you have any questions about this example situation or about how to indicate your responses, please raise your hand now.

You will next be given the description of a situation where Individual A, a participant in an experiment, has to choose between various possible actions. After you read the description, you must consider the possible actions and indicate on your computer screen how socially appropriate these are in a table similar to the one shown above for the example situation.

After this, the computer will randomly select one participant who drew a ball of the same colour as you (that is, it will select a participant who drew a green ball if you drew a green ball, or a participant who drew a yellow ball if you drew a yellow ball). The computer will then randomly select one action Individual A can choose. Your evaluation of this action will be compared with that of the randomly selected participant who drew a ball of the same colour as you. If your evaluation is the same as theirs, you will receive £8 for this task; otherwise you will receive zero.

For instance, imagine the example situation above was the actual situation and the possible action "Leave the wallet where it is" was selected by the computer. If your evaluation had been "somewhat socially inappropriate" then your task earnings would be £8 if the person you are matched with also evaluated the action as “somewhat socially inappropriate” and zero otherwise.

<u>The situation</u>

Individual A is taking part in an experiment in this lab. The room has been divided into two sections on the basis of which colour of ball participants drew from a bag at the beginning of the experiment. On one side everyone drew a green ball; on the other side everyone drew a yellow ball. The anonymity of Individual A's decisions in the experiment is guaranteed.

Individual A's task will be to decide how to divide £16 between two other participants in the experiment, one who drew the same ball colour as Individual A, and another who drew a different ball colour from Individual A. Individual A may divide the money any way he or she likes so long as the amount allocated to each person is a multiple of two. Individual A may not allocate any of the money to his- or herself. However, Individual A will also receive a payment. This payment might be £6, £8 or £10. This will be randomly decided at the end of the experiment by the computer, which is equally likely to select any of these amounts.

Please now answer one question on your screen, to ensure you understand this situation.

Task Two

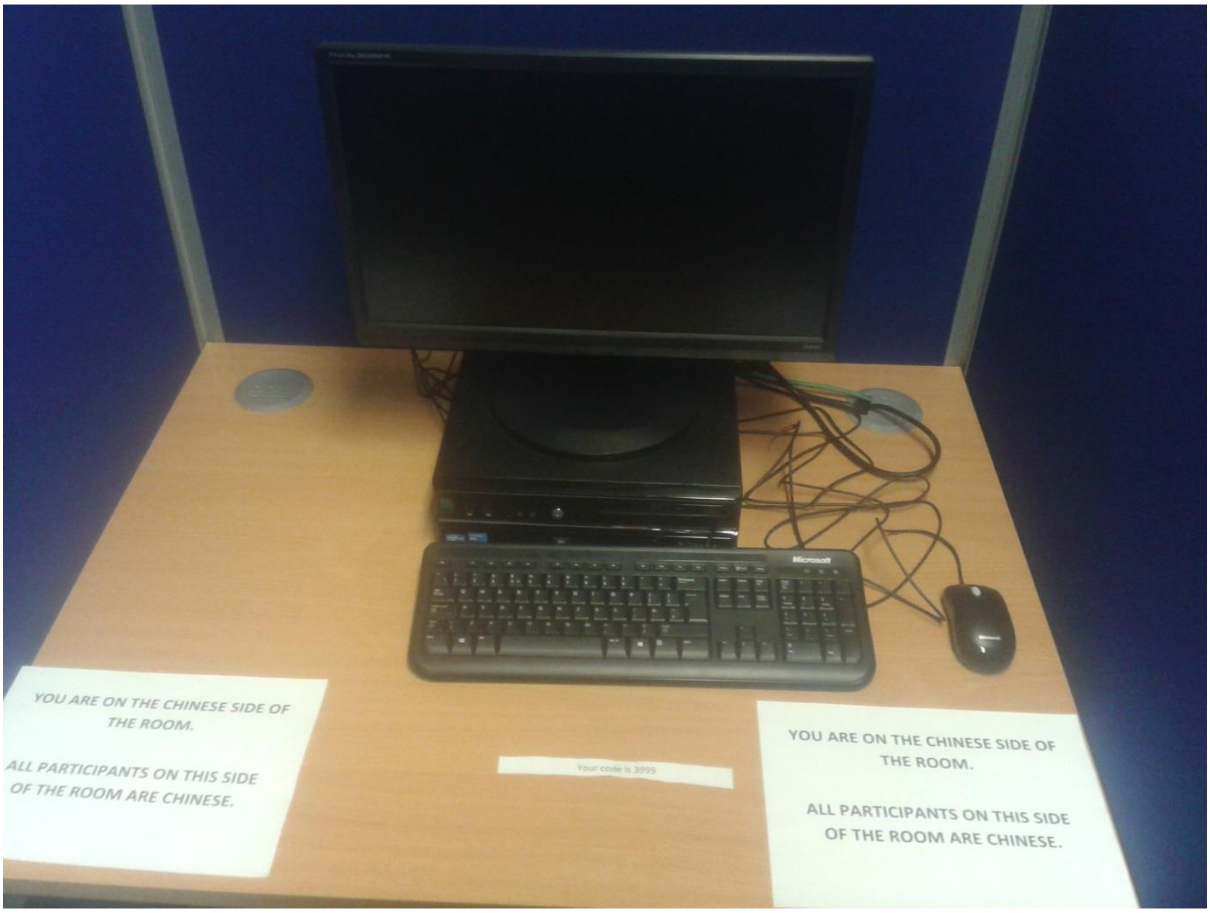
In this experiment, one third of you will be randomly assigned by the computer into a role entitled ‘Individual A’. The decisions made by Individual As during the task will determine the payments from the task received by the other two thirds of participants. Each of you has an equal chance of being an Individual A. Exactly who the Individual As are will not be revealed until the end of the experiment. In the meantime, we ask all participants to make a decision **as if** they are an Individual A.

Please make your decision carefully, as it may be used to determine participants’ payments.

Assume for the rest of this paragraph that you are an Individual A. Your task is like the one you evaluated in the previous task. Your task will be to decide how to divide £16 between two other participants in the experiment, one who drew the same ball colour as you, and another who drew a different ball colour from you. You may divide the money any way you like so long as the amount allocated to each person is a multiple of two. You may not allocate any of the money to yourself. However, you will also receive a payment. This payment might be £6, £8 or £10. This will be randomly decided at the end of the experiment by the computer, which is equally likely to select any of these amounts.

You will next see one question on your screen. Please answer it to ensure you understand this part of the experiment.

B: Photo of sign on desks in computer lab



YOU ARE ON THE CHINESE SIDE OF THE ROOM.

ALL PARTICIPANTS ON THIS SIDE OF THE ROOM ARE CHINESE.

Your code is 3999

YOU ARE ON THE CHINESE SIDE OF THE ROOM.

ALL PARTICIPANTS ON THIS SIDE OF THE ROOM ARE CHINESE.

C: Analysis of the significance of ordering effects

In each of the *Nationality* and *Artificial* treatment we conducted three sessions (72 participants) wherein subjects first played the allocator game and then the Krupka-Weber norm elicitation task, and two sessions (48 participants) where the order of tasks was reversed. We test whether the order in which the tasks are played affects either discrimination behaviour or the perceived appropriateness of discrimination.

Regarding the impact of task order on discrimination behaviour, randomisation tests find the average level of observed discrimination does not significantly differ between participants who play the allocation game first and those who have already undergone the norm elicitation task, in either the *Nationality* treatment ($p = 0.77$) or the *Artificial* treatment ($p = 0.23$). Regarding the impact of task order on the perceptions of the appropriateness of discrimination, none of the evaluations are subject to significant ordering effects in either the *Nationality* treatment (all p-values > 0.229) or *Artificial* treatment (all p-values > 0.309). As in the main text, p-values are corrected using the Benjamini-Hochberg false discovery rate procedure to account for the fact that we are conducting multiple tests.

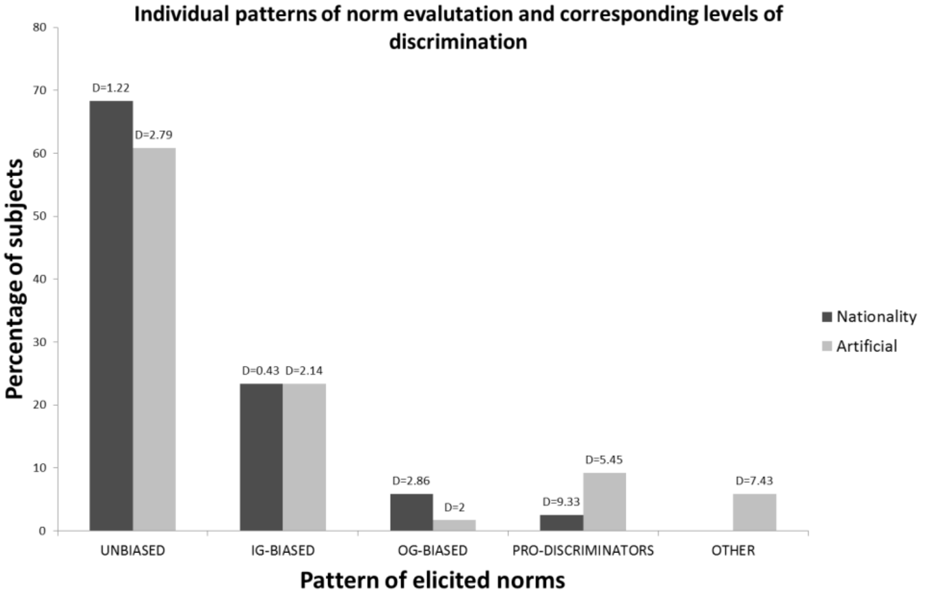
D: Individual patterns of behaviour

We divide participants into five categories, on the basis of their responses to the norm-elicitation task. Most individuals' ratings monotonically increase in appropriateness as allocations move away from the in-group favouring (16,0) towards more equal allocations, until a peak is reached (usually the 8,8 split) after which the individual's ratings monotonically decrease in appropriateness. This means that the individual believes the most appropriate possible action is not extreme discrimination. We subdivide these participants into three types. UNBIASED types perceive discrimination against the in-group member to be of equal appropriateness to discrimination against the out-group member. IG-BIASED types perceive discrimination against the in-group member to be of lesser appropriateness. OG-BIASED types perceive discrimination against the out-group member to be of lesser appropriateness. This categorisation is done by comparing the sum of the ratings the individual assigns to in-group-favouring allocations against the sum of the ratings they assign to out-group-favouring allocations.

Some participants, however, assign their highest rating to the (16,0) allocation and then monotonically decrease the appropriateness of their ratings as more money is given to the out-group member. Such participants are perceiving extreme discrimination against the out-group member to be the social norm. We label them PRO-DISCRIMINATORS. The few participants whose ratings do not follow any of the above patterns are categorised as OTHER.

Figure D1 displays the percentage of participants in each treatment who followed each pattern as well as their average levels of discrimination.

Figure D1



Notes: Figure D1 shows the percentage of participants in each treatment whose evaluations follow each pattern. Above each bar, D=the average level of discrimination against out-group members by participants of the given type in the given treatment – e.g. for UNBIASED participants in the Nationality *treatment, D=1.22 indicates these participants discriminated by an average of £1.22.*

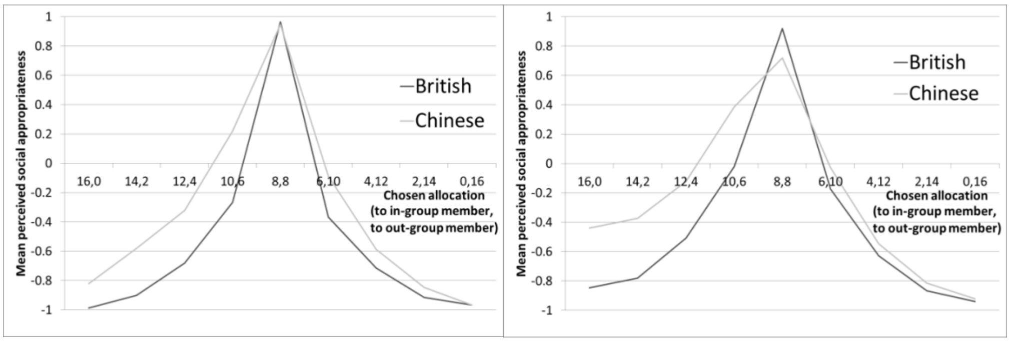
E: Cross-national differences

We conclude with an analysis of cross-national differences in the perceptions of appropriateness of discrimination as well as in discriminatory behaviour. Figure E1 compares, across nationalities, the mean evaluations of social appropriateness in each treatment, while Tables E2a and E2b provide a full breakdown of responses, with tests of significance. In both the *Artificial* and the *Nationality* treatment, Chinese participants perceive discrimination to be more socially appropriate than British participants. The general pattern of norms and most of the modal evaluations for a given allocation is the same for the two nationalities. However, in both treatments all in-group-favouring allocations are given significantly higher appropriateness ratings by the Chinese. Out-group-favouring allocations are also perceived to be more socially appropriate by the Chinese, but the difference is only significant for the (6,10) split in *Nationality*. Finally, in both treatments, the equal split is given a lower appropriateness rating by the Chinese, although the difference is significant only in the *Artificial* treatment (p=0.028).

As the Chinese give relatively favourable evaluations to allocations favouring either the in-group member or the out-group member, one might question whether what we find is actually national differences in the social appropriateness of inequality rather than of discrimination. However, note that the national differences in evaluations are greater for in-group-favouring allocations than out-group-favouring ones. If the national differences existed only for the social appropriateness of inequality, they would be reflected in symmetrical national differences in the evaluation of in-group and out-group favouring allocations.

In Table E3 we confirm this asymmetry is significant. We subtract the rating each individual assigns to an allocation discriminating by a given amount in favour of the out-group member from the rating they assign to the allocation which discriminates by the same amount in favour of the in-group member. This provides a measure of the relative appropriateness an individual ascribes to discriminating in favour of their own group member rather than the other group member. We run randomisation tests on these relative ratings for each level of discrimination, and find that in each case there are significant national differences in the means. This indicates that, in both treatments, the relative appropriateness of discriminating in favour of one’s own group member rather than the other group member is perceived to be significantly higher by the Chinese than by the British.

Figure E1: Social appropriateness of allocations in the *Nationality* treatment (left) and *Artificial* treatment (right)



Notes: Figure E1 shows the mean ratings participants of each nationality ascribe to each allocation, in the Nationality *treatment (left) and the* Artificial *treatment (right). Mean ratings are constructed by assigning values of 1, 0.33, -0.33 and -1 for the ratings 'very appropriate', 'somewhat appropriate', 'somewhat inappropriate' and 'very inappropriate' respectively, and averaging the values for all participants in a given treatment. Allocations are denoted by the amount given to the in-group member on the left, and the amount given to the out-group member on the right – e.g. (16,0) denotes allocating £16 to the in-group member and £0 to the out-group member.*

Table E2a: Social appropriateness ratings of British and Chinese in Nationality treatment

Allocation	16,0	14,2	12,4	10,6	8,8	6,10	4,12	2,14	0,16
British participants									
v.appropriate	0.0	0.0	0.0	1.7	96.7	0.0	0.0	0.0	0.0
s.appropriate	0.0	0.0	3.3	33.3	1.7	26.7	3.3	1.7	1.7
s.inappropriate	1.7	15.0	41.7	38.3	1.7	41.7	36.7	10.0	1.7
v.inappropriate	98.3	85.0	55.0	26.7	0.0	31.7	60.0	88.3	96.7
Mean rating	**-0.99**	**-0.90**	**-0.68**	**-0.27**	**0.97**	**-0.37**	**-0.71**	**-0.91**	**-0.97**
Chinese participants									
v.appropriate	3.3	5.0	5.0	8.3	93.3	1.7	0.0	0.0	0.0
s.appropriate	6.7	11.7	16.7	70.0	5.0	43.3	5.0	1.7	1.7
s.inappropriate	3.3	25.0	53.3	18.3	1.7	45.0	51.7	20.0	1.7
v.inappropriate	86.7	58.3	25.0	3.3	0.0	10.0	43.3	78.3	96.7
Mean rating	**-0.82**	**-0.58**	**-0.32**	**0.22**	**0.95**	**-0.09**	**-0.59**	**-0.85**	**-0.97**
p-value (Benjamini-Hochberg)	**0.026**	**0.006**	**0.000**	**0.000**	**0.883**	**0.008**	**0.167**	**0.364**	**1.000**

Table E2b: Social appropriateness ratings of British and Chinese in Artificial treatment

Allocation	16,0	14,2	12,4	10,6	8,8	6,10	4,12	2,14	0,16
British participants									
v.appropriate	7.6	6.1	6.1	3.0	92.4	0.0	0.0	1.5	3.0
s.appropriate	0.0	1.5	3.0	53.0	4.6	40.9	6.1	1.5	0.0
s.inappropriate	0.0	12.1	50.0	31.8	1.5	42.2	43.9	12.1	0.0
v.inappropriate	92.4	80.3	40.9	12.1	1.5	16.7	50.0	84.9	97.0
Mean rating	**-0.85**	**-0.78**	**-0.51**	**-0.02**	**0.92**	**-0.17**	**-0.63**	**-0.87**	**-0.94**
Chinese participants									
v.appropriate	20.0	10.0	4.0	16.0	72.0	2.0	2.0	2.0	4.0
s.appropriate	12.0	20.0	36.0	78.0	18.0	52.0	6.0	2.0	0.0
s.inappropriate	0.0	24.0	48.0	4.0	6.0	36.0	50.0	18.0	0.0
v.inappropriate	68.0	46.0	12.0	2.0	4.0	10.0	42.0	78.0	96.0
Mean rating	**-0.44**	**-0.37**	**-0.12**	**0.39**	**0.72**	**-0.03**	**-0.55**	**-0.81**	**-0.92**
P-value (Benjamini-Hochberg)	**0.006**	**0.003**	**0.000**	**0.000**	**0.028**	**0.177**	**0.492**	**0.571**	**1.000**

Notes: Tables E2a and E2b presents the breakdown, by percentage, of social appropriateness ratings assigned to each allocation by participants of each nationality, in the Nationality *treatment (top) and* Artificial *treatment (bottom). Allocations (e.g. 16,0) are denoted by the amount given to the in-group member on the left (£16), and the amount given to the out-group member on the right (£0). Shaded cells represent the modal ratings for each allocation in each treatment. Mean ratings are taken by assigning values of 1, 0.33, -0.33 and -1 for the ratings 'very appropriate', 'somewhat appropriate', 'somewhat inappropriate' and 'very inappropriate' respectively, and averaging the values for all participants in a given treatment. Benjamini-Hochberg-corrected p-values are reported from randomisation tests on the null hypothesis that the mean ratings on a given allocation are statistically indistinguishable by nationality.*

Table E3: Appropriateness bias towards in-group favouritism over out-group favouritism

Level of discrimination	£16	£12	£8	£4
Nationality treatment				
British	-0.02	0.01	0.03	0.10
Chinese	0.15	0.27	0.27	0.31
p-Value (Bonferroni)	0.020	0.004	0.021	0.022
Artificial treatment				
British	0.09	0.09	0.12	0.19
Chinese	0.48	0.44	0.43	0.42
p-Value (Bonferroni)	0.009	0.004	0.007	0.005

Notes: Table E3 shows the extent to which, on average, participants of each nationality perceive discrimination by a given amount to be more appropriate when the beneficiary is the in-group member, in the Nationality *treatment (top) and the* Artificial *treatment (bottom). The measure is constructed by subtracting each individual's rating of an out-group-favouring allocation from their rating of the equivalent in-group-favouring allocation. Benjamini-Hochberg-corrected p-values are reported from randomisation tests on the null hypothesis that the relative perceived appropriateness of discriminating in favour of an in-group member rather than an out-group member does not differ between nationalities.*

Finally, in Table E4 we look at whether the national differences in the perceived social appropriateness of discrimination are reflected in more discriminatory choices actually being made by the Chinese. In the *Nationality* treatment, the answer is no. The distribution of allocations made by each nationality in this treatment is almost identical. The mean difference between allocation to the in-group and out-group is £1.27 by the British and £1.40 by the Chinese, amounts which are statistically indistinguishable. However, in the *Artificial* treatment, the Chinese are more discriminatory; 81.8% of British participants distribute the money equally compared to only 52% of the Chinese, and the mean differences in allocations are significantly different (p=0.007), at £1.88 and £4.72 for the British and Chinese participants respectively. An OLS regression (Table E5) indicates that this finding is robust to the inclusion of various controls.

Table E4: Discrimination in the allocator game – national comparisons

Allocation	(16,0)	(14,2)	(12,4)	(10,6)	(8,8)	(6,10)	(4,12)	(2,14)	(0,16)
	Nationality treatment								
British	3.3	1.7	1.7	10.0	83.3	0.0	0.0	0.0	0.0

Chinese	5.0	1.7	0.0	10.0	83.3	0.0	0.0	0.0	0.0
	Artificial treatment								
British	9.1	0.0	1.5	7.6	81.8	0.0	0.0	0.0	0.0
Chinese	18.0	6.0	4.0	20.0	52.0	0.0	0.0	0.0	0.0

Notes: Table E4 shows the percentage of participants of each nationality who choose each allocation, in the Nationality *treatment (top) and the* Artificial *treatment (bottom). Allocations are denoted by the amount given to the in-group member on the left, and the amount given to the out-group member on the right – e.g. (16,0) denotes allocating £16 to the in-group member and £0 to the out-group member.*

Chapter Four: Nudging the electorate: what works and why?[56]

Abstract

We report results from two studies designed to test the effectiveness of different interventions to raise voter registration rates, and to probe the mechanisms underlying observed behavior change. In the first study, in a natural field experiment ahead of the 2015 UK General Election, we varied the content of a postcard sent by Oxford City Council to 7,679 unregistered student voters encouraging them to register onto the electoral roll. We find that relative to a baseline condition, emphasising negative monetary incentives (i.e. the possibility of being fined) significantly increases registration rates, while positive monetary incentives (being entered into a lottery if you register) may have some tendency to reduce registration rates. A third class of purely non-monetary nudges have no overall effect on registration rates. In the second study, we show that these differences can be explained, at least in part, by social norms.

JEL classifications: D72: Political Processes: Rent-Seeking, Lobbying, Elections, Legislatures, and Voting Behavior; D03: Behavioral Microeconomics: Underlying Principles; C93: Field Experiments; H83: Public Administration.

Keywords: Voter Registration; Voting; Field Experiment; Nudging; Social Norms; Fines; Rewards.

[56] This chapter was co-authored with Felix Kölle, Daniele Nosenzo and Chris Starmer

1. Introduction

Behavioural science is increasingly being used to advance low-cost interventions across a growing spectrum of public policy areas. The associated body of research is being built by – and is of interest to – both academics (e.g. Chetty, 2015; Hallsworth et al., 2014) and applied policy units (e.g. Behavioural Insights Team, 2010, 2011, 2012, 2016). Interventions often take the form of 'nudges' – manipulations of the 'choice architecture' which aim to systematically change people's behaviour at low cost for the policymaker.[57] This paper explores aspects of the mechanics of nudging by using a combination of field and online experiments. Our natural field experiment explores the extent to which a particular decision can be influenced by different types of nudge; with our online experiment we investigate the underlying mechanisms that lead to the success or failure of different nudges.

We do this in the context of a particular policy area: voter registration. In many countries – including the United Kingdom, where our study is conducted – any citizen wishing to vote must first register on the electoral roll. Registration in the UK is technically mandatory, with non-registration punishable with an £80 fine, although in the past two decades a substantial gap has emerged between the numbers of eligible and registered voters (Bite the Ballot, 2016), one which was further increased by the implementation of a legislative change in 2014: previously all members of a household could be registered collectively, but the law now requires each person to register individually (Electoral Registration and Administration Act 2013). Besides any intrinsic benefits of wide democratic participation, high registration rates serve the government's interest insofar as the electoral roll has secondary uses such as fraud-detection and jury recruitment. Given its very low cost nature, nudging – if it can be shown to be effective – would be an attractive strategy for such organisations to employ in pursuit of this goal. Our study therefore explores interventions that can

[57] We are using the label 'nudge' to describe any type of low-cost intervention that is designed with the objective of changing individuals' behaviour. Thaler and Sunstein (2008) refer to 'nudges' to describe interventions that leave unaffected the underlying economic incentives: some of the manipulations that we use in our experiment have this property, whereas others induce small changes in the underlying incentives and so, strictly speaking, would not qualify as nudges according to the Thaler and Sunstein's definition.

be applied, at minimal financial expense, to nudge citizens to register to vote in elections.

In the first part of this paper, we report the results of a natural field experiment run ahead of the 2015 UK General Election in partnership with one such interested party, Oxford City Council, who sent postcards to students living in university accommodation, encouraging them to register. Councils have a particular interest in discovering successful ways of targeting such students, as they represent a segment of society whose registration rates have been particularly affected by the recent legal change – previously universities could register en masse all accommodated students, but students are now required to register themselves individually. While all the postcards urged recipients not to miss their chance to vote, we systematically varied the precise content of their messages in order to test the effects of different nudges on registration rates.

We investigated two broad types of nudges, relying on monetary incentives in the form of either a small *gain* or a small *loss*. There is a large and diverse literature in economics showing that interventions based on negative incentives, such as the threat of monetary loss, may produce stronger responses than those relying on positive incentives, such as the promise of monetary gains, even when the interventions involve identical financial incentives and the only difference is in the way these incentives are described to the individual (for reviews of these literatures, see, e.g., Balliet et al., 2011; van Lange et al., 2014; Nosenzo, 2016). In the context of policy interventions aimed at reinforcing civic duties, as in the case of voter registration, positive and negative incentives may produce different behaviours because of the way they interact with the very notion of 'civic duty': while negative incentives may reinforce the perception that registering to vote is what one ought to do, positive incentives may have the opposite effect. In the second part of our paper, we investigate this hypothesis by examining the effects that positive and negative incentives have on the perception that registering to vote is a normative obligation.

We also investigated another class of nudges, which do not involve financial incentives but rely on purely psychological mechanisms to affect behaviour. We refer to these as our *non-monetary* nudges. Non-monetary nudges are sometimes regarded as preferable to monetary ones, especially in areas of public policy, such as voter registration programs, where intrinsic motivations may play an important role in producing the desired behaviour: the concern here is that the use of extrinsic (monetary) incentives may crowd-out these intrinsic motivations (see, e.g., Gneezy et al., 2011; Bowles and Polania-Reyes, 2012). Another advantage of non-monetary interventions, compared to monetary ones, is that the former typically involve smaller financial costs than the latter.

In our field experiment, we implement these nudges across six treatments. In a baseline treatment, the postcard that was sent to unregistered student voters simply encouraged them to register, without any additional message. We investigate the effectiveness of monetary losses in a second treatment by adding a message highlighting the existence of the potential £80 fine for those who fail to register. Emphasising the possibility of encountering financially costly legal action has previously been found to exert a substantial positive effect in other domains of policy intervention, such as the enforcement of TV license registration (Fellner et al., 2013), credit card debt repayment (Bursztyn et al., 2015), or traffic violations (Lu et al., 2016).[58] Kleven et al. (2011) similarly found that the threat of audits raised tax returns, although mixed evidence was found in this domain by Blumenthal et al. (2001) and Slemrod et al. (2001).[59]

To test the effectiveness of monetary gains, we designed two further treatments where students were offered entry into a lottery to win small cash prizes (of the value of £80) for those who registered early. Financial inducements have previously been found to raise voter registration (John et al., 2015) and voter turnout (Panagopoulos, 2013), although in the latter case only when the inducements were sufficiently large. Moreover, in John et al. (2015) the lottery involved large financial incentives (between £1000 and £5000) and produced only a small positive effect (an increase of two percentage points in registration rates).

[58] However, Lu et al. (2016) found that messages were effective only when they contained personalised information about own past traffic violations, and not when they communicated the mere existence of fines.

[59] See Hallsworth (2014) for a recent review of field experiments on tax compliance interventions.

Finally, we investigate the effectiveness of non-monetary nudges in two treatments where we included in the postcard messages that invoked purely psychological mechanisms. Specifically, we attempted to harness the so called ‘foot-in-the-door’ mechanism. First identified by Freedman and Fraser (1966), this is a psychological effect – which has been replicated across a wide array of circumstances (see e.g. Burger, 1999) – wherein people are more likely to comply with a large request, oriented towards a particular goal, after they have first complied with a small request. In our case, the small requests took the form of asking students to provide their phone number (so that they could be sent a reminder to register); or simply to report by text whether they intended to register.

Our results show that emphasising the possibility of being fined yielded a large positive effect, with subjects exposed to this intervention being around 1.7 times more likely to register than those exposed to the baseline intervention. In contrast, the prospect of financial gain had no overall effect on registration – and once the deadline for entry into the lottery had passed, subjects exposed to this intervention were significantly less likely to register than those in the baseline condition. Meanwhile, the non-monetary interventions had no discernible effect.

In the second part of the study reported in this paper, we investigated a possible mechanism underlying the effectiveness of the negative monetary incentives and the ineffectiveness of positive incentives: their contrasting effects on the perception of what constitutes socially appropriate behaviour in the context of voter registration. A growing body of recent economic research (e.g., Burks and Krupka, 2012; Gächter et al., 2013; Krupka and Weber, 2013; Banerjee, 2016; Barr et al., 2016; Krupka et al., 2016) suggests that social norms and norm-compliance drive behaviour in a wide range of contexts. We hypothesised that the fine and lottery treatments may have divergent effects on the perceptions of normative appropriateness of registering to vote: emphasising that failing to register is punishable by law may strengthen the perception that one ought to register, while offering monetary inducements for registering may weaken the perception that doing so is an action already expected within society.

Our results support these hypotheses. Using the incentivised norm-elicitation method introduced by Krupka and Weber (2013), we found that exposing individuals to the fine nudge strengthened their perception that failing to register to vote was socially inappropriate behaviour, while exposing them to the lottery nudge weakened the perception that registering to vote was socially appropriate behaviour. Consequently, we propose that strengthening/weakening the social norm relating to registration had a direct effect on respondents' likelihood of registering, and that this partly explains why the fine intervention was successful while the lottery intervention was not. Indeed, a possible interpretation of our results is that– just as in some previous research (e.g. Frey and Oberholzer-Gee, 1997; Ariely et al., 2009; Gneezy et al., 2011; Bowles and Polania-Reyes, 2012) – the introduction of monetary incentives crowded out individuals' intrinsic motivation to engage in socially constructive behaviour (especially once the deadline for the lottery had passed), and the adverse effect of the lottery intervention on the perceived social norm of registering may be partly behind this effect.

Our study contributes to the literature regarding behavioural insights and nudging in public policy. We identify a low-cost strategy – emphasising the possibility of a fine for not registering – that governments can use to substantially increase registrations. However, we also identify cases where nudges can fail to work or perhaps even backfire. We show offering monetary rewards for registering can have neutral or adverse effects, while attempting purely psychological interventions may not yield any desired positive effects. One advantage of our study is that our investigation into the relative effectiveness of non-monetary and monetary nudges (and, within the latter, of positive and negative incentives) is done within a unified experimental design and in the context of a large-scale field intervention (N = 7,679).

Moreover, another distinctive feature of our paper is that our focus is not only on identifying which intervention is most effective to encourage voter registration, but also on understanding the underlying behavioural and psychological mechanisms that make specific interventions more successful than others. In particular, we believe that we are the first to show that policy interventions that rely on positive incentives may weaken, whereas those relying on negative incentives strengthen, the perceptions of normative appropriateness of the target behaviour. These differences in normative perceptions may explain the relative success of the positive and negative monetary interventions that we found in our field experiment. In this sense, our paper also adds to the recent economic literature on the importance of social norms for understanding human social behaviour (Krupka and Weber, 2013; Gächter et al., 2013; Kimbrough and Vostroknutov, 2016).

2. Study I: Field experiment on voter registration

2.1 Experimental design

Our field experiment was designed to test the effectiveness of six different low cost, nudge-style, interventions for raising voter registration rates ahead of the 2015 UK General Election. These six interventions fall into three broad categories: negative monetary nudges; positive monetary nudges; and non-monetary nudges. All of these interventions were implemented via adjustments to a message sent in a bulk, randomised, mail out (details below) to students living in the UK City of Oxford.

We implemented one negative monetary intervention invoking the threat of a *monetary loss* for failing to register. We did this by highlighting to subjects the truthful fact that they could be fined £80 if they did not register. This penalty is specified in UK law, and it is referred to in standard materials that Oxford City Council (OCC) use to promote voter registration. Despite the power existing in law, however, it is rarely used by councils and it is unclear how aware a typical unregistered voter would be of the possibility of being fined for failure to register.

We implemented two interventions involving the prospect of *monetary gain* for registering. We did this with two treatments offering entry into a lottery to win cash prizes of the value of £80 for those who registered by a specific deadline. The two treatments differed only in that one attempted also to harness regret aversion (Loomes and Sugden, 1982), by telling recipients that those who did not register would still be entered into the lottery and informed if they won, but would be unable to claim their prize. Regret aversion has previously been shown to affect entry decisions into lotteries (Zeelenberg and Pieters, 2004; Gneezy, 2014; Imas et al., 2016).

We also implemented two interventions involving *non-monetary* mechanisms.[60, 61] These were motivated by the so called foot-in-the-door effect (Freedman and Fraser, 1966). More specifically, we tested whether message recipients were more likely to register (i.e., comply with a large request) if they had first complied with a small request related to voter registration: in both variants, subjects were asked to send a one-word text message to OCC; in one case subjects were asked to text the word 'myvote' to indicate that they planned to register; in the other, subjects could text the word 'reminder' to register to receive a free reminder text to prompt them to register nearer the deadline for voters to register in the 2015 UK General Election.[62] Foot-in-the-door mechanisms have previously been tested in the context of voter turnout, with mixed results: asking subjects if they intended to vote was found to have a large, positive impact on turnout by Greenwald et al. (1987, 1988), but not by Smith et al. (2003).

Our nudges were transmitted via postcards, which OCC mailed to all unregistered voters living in student accommodation belonging to the University of Oxford and Oxford Brookes University on March 9-10 2015, ahead of the April 20 deadline for

[60] There are a number of non-monetary interventions that have been shown to positively affect voter *turnout*, e.g. personalised get-out-the-vote contact (e.g. Gerber and Green, 2000; Ramirez, 2005), priming one's identity as a voter (Bryan et al., 2011), or applying social pressure on people to vote (e.g. Gerber et al., 2008, 2010; Davenport, 2010; Panagopoulos, 2010). See Rogers et al. (2013) for a recent review of this literature. While one might expect that what works for voter turnout should also work for voter *registration*, ultimately this is an empirical question. Nevertheless, the fact that voter turnout appears quite susceptible to psychological processes provides a strong motive for exploring non-monetary nudges as a policy tool regarding registration.

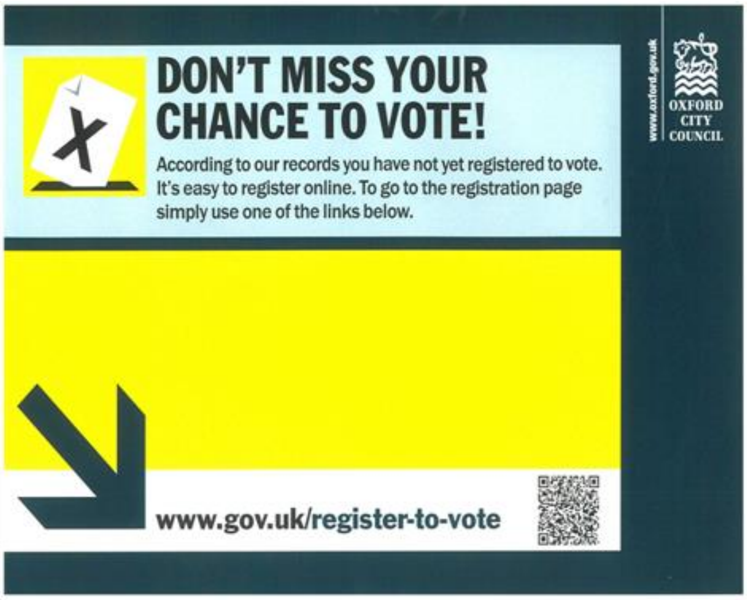
[61] We also considered designing a 'peer pressure' treatment, where we would inform students of the past registration rates of their peers in order to induce them to register. However, at the time we designed the experiment, the registration rate of students was very low (about 8-9%) and we thought that informing subjects of such low rates may actually have the opposite effect of discouraging registrations.

[62] Our primary interest in relation to this treatment was in assessing whether the *act of registering* for an automated text reminder (which was designed to be very quick and easy) would act as a foot in the door, enhancing the propensity to subsequently register to vote. We were not especially interested in the later effect that a subsequent automated reminder might have (we do not have controlled comparisons for benchmarking any such effect). Of course all of our treatments can be considered as sending slightly different forms of reminder and, more broadly, reminders have been shown to be effective nudges in relation to voter registration (Bennion and Nickerson, 2011), as well as in other contexts (e.g. Altmann and Traxler, 2014)

voters to register in the General Election. We collaborated with the Council to engineer the content of these postcards. While all postcards encouraged their recipients to register, the content of the messages they contained varied, allowing us to test the effects of different nudges on registration rates.

All postcards were double sided (see Figure 1 for a copy of the postcard used in our baseline condition and Appendix A for copies of the other postcards). The back simply contained the message, '*IMPORTANT INFORMATION ABOUT YOUR RIGHT TO VOTE, OVERLEAF*.' The front featured the heading, '*DON'T MISS YOUR CHANCE TO VOTE! According to our records you have not yet registered to vote. It's easy to register online. To go to the registration page simply use one of the links below*.' The bottom of this side contained the address of the government web page for registering to vote, and a QR code which would take recipients to the same page. These features were held constant across treatments. The postcards differed by treatment only according to the text included in a box below the heading on the front side.

Figure 1: Postcard used in the Baseline treatment



DON'T MISS YOUR CHANCE TO VOTE!

According to our records you have not yet registered to vote. It's easy to register online. To go to the registration page simply use one of the links below.

www.gov.uk/register-to-vote

X
www.oxford.gov.uk
OXFORD
CITY
COUNCIL

2.2 Treatments

In the **Baseline** treatment, the box was left blank (Figure 1). This treatment therefore serves as a basis for comparison against the other treatments.

In the **Monetary Loss** treatment, the box contained the message: '*If you don't register you could be fined £80.*'.

In treatment **Monetary Gain 1**, the box contained the message: '*If you register by 27 March 2015 you will be entered into a lottery to receive one of ten £80 prizes. Winning students will be notified in June 2015.*'. In treatment **Monetary Gain 2**, the box contained the message: '*You have been entered into a lottery to receive one of ten £80 prizes. Winners will be notified in June 2015 but you will only be able to claim your prize if you were already registered by 27 March 2015. If not your prize will go to another student.*'.

In treatment **Non-Monetary 1**, the box contained the message: '*We'd like to know if you are intending to register. If you are, please text 'myvote' to 60886.*'. In treatment **Non-Monetary 2**, the box contained the message: '*Would you like us to send you a text reminder? If you do, please text 'reminder' to 60886.*'. In both cases, texts were free of charge and this was clearly stated in the postcard.

2.3 Assignment to treatment

The postcards were sent out on March 9-10 2015 to 7,679 voters who were still unregistered at the time and who lived in student accommodation buildings belonging to the University of Oxford and Oxford Brookes. In order to minimise the likelihood of subjects seeing postcards belonging to treatments other than the one they were assigned to, we randomised assignment to treatment at the building level: all students living in a single building were assigned to the same treatment. For the University of Oxford, all students living in a single college were assigned to the same treatment. For Oxford Brookes University, all students living in a single hall of residence were assigned to the same treatment, with the exception that two very large halls were split into several geographically distinct units of assignment. This was to ensure balance between treatments in the proportion of subjects attending each university – we considered this important, given large demographic (in particular, socioeconomic) differences between the student populations of each university. We further balanced treatment assignment by residence size (small and large) and to the age of college (ancient and modern) to account for other potential unobserved characteristics. The resulting sample sizes were as follows: Baseline (n = 1193); Monetary Loss (n = 1357); Monetary Gain 1 (n = 1250); Monetary Gain 2 (n = 1317); Non-Monetary 1 (n = 1262); and Non-Monetary 2 (n = 1300). See Appendix B for further details on the assignment procedure and for a full breakdown of the colleges and halls assigned to each treatment.

2.4 The dataset

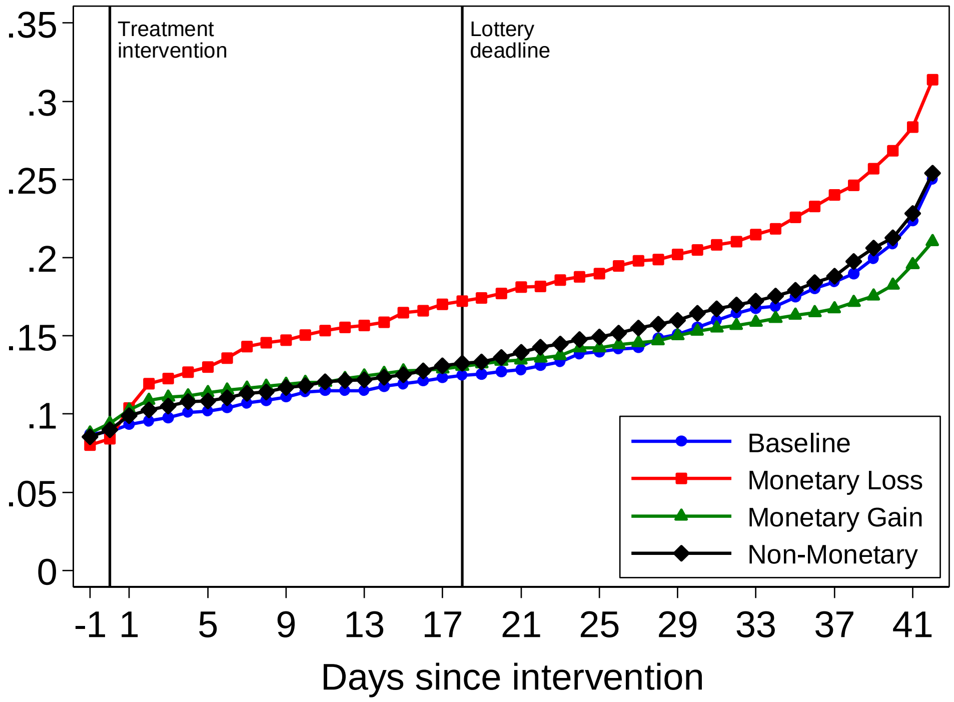
OCC provided us with anonymised data on registration rates amongst students residing in each college and hall at various points in time between January and April 2015. In particular, for each individual in our dataset, our data specify whether or not they were registered on: January 2, March 8 (the day before the postcards of our experiment were sent out), and any subsequent day between March 9 and April 20 (the formal deadline to register to vote for the General Election). The data also contain information on the treatment each individual was assigned to, their university affiliation (University of Oxford or Oxford Brookes University), and their hall or college. Other demographic data such as gender, age, etc. was not available to the Council.

3. Results

In the following analysis we pool data from the two Monetary Gain treatments. This is because we found there were no significant differences between the effects of Monetary Gain 1 and Monetary Gain 2 (see Table C1 in the Appendix). Similarly, we found no significant differences between the two Non-Monetary treatments (see Table C1) and thus, for ease of exposition, we combine these two treatments in our analysis. Thus, in the remainder of the section, our analysis is based on the following four conditions: Baseline (n = 1193); Monetary Loss (n = 1357); Monetary Gain (n = 2567); and Non-Monetary (n = 2562).

Figure 2 shows how registration rates differ between treatments over the entire period between the intervention and the deadline for registering. On a daily basis between March 8 and April 20, it displays the cumulative fraction, by treatment, of registered subjects amongst those who were unregistered on January 2.

Figure 2: Cumulative registration rates by treatment



Notes: Figure 2 shows, on a daily basis between March 8 and April 20, the amount of registered students in the treated buildings as a fraction of all students in these buildings who had been unregistered on January 2. The first vertical line represents the day of our treatment intervention (March 9) and the second line represents the day of the lottery deadline in the Monetary Gain treatments (March 27).

Pre-intervention registration rates (i.e. in the period January 2 – March 8) are very similar across all treatments; the fraction of registered subjects ranges between 0.080 and 0.088, showing no significant differences across treatments (see also Table 1 below). This suggests that later treatment differences are unlikely to be driven by pre-existing differences between the subjects assigned to each intervention.

After our intervention (i.e. in the period March 9 – April 20), substantial differences emerge in the registration rates across treatments. On April 20 (the day of the registration deadline), the fraction of registered students amounts to 0.25 in Baseline, 0.31 in Monetary Loss, 0.21 in Monetary Gain, and 0.25 in the Non-Monetary treatment. Hence, compared to the case of a simple postcard, only the emphasis of (potential) negative monetary consequences had a positive effect on registration; registration rates in Monetary Loss are 24% higher than in Baseline. The emphasis of (potential) positive monetary consequences, in contrast, had no positive effect on registration rates. While registration rates are initially somewhat above the ones in

Baseline, this effect flips towards the end of our observation period, after monetary incentives have been removed again. Overall, the registration rate in our Monetary Gain treatments is 16% lower than in Baseline. For the Non-Monetary treatments, we find almost identical registration rates as in Baseline, suggesting that the effects of the non-monetary nudges were minimal in our setup. Given how few engaged with the invitation to send a text message (only 4 students requested a text reminder, while a mere 9 texted their intention to register) it is not surprising that we record no discernible foot-in-the-door effect.

To further explore the observed treatment differences in registration rates, we run logistic regressions to model the individual level registration decision. Our dependent variable is 1 or 0 depending on whether or not an individual has registered in a given period. As independent variables, we use dummies to represent the treatment an individual was assigned to. Given the very different nature of the two universities that were included in our study, we also include a dummy variable for whether a student was affiliated with either Oxford Brookes University or University of Oxford. As further controls we include: which of the two Oxford voting areas (General Election constituencies) a given student resides in; the size of the residence unit they live in; and an ‘ancient’ dummy, which takes value 1 if the college or hall in which they live is older than 100 years, and value 0 otherwise.[63] To correct for heteroscedasticity and potential dependency of observations within halls, we cluster standard errors by residence unit.[64] The results of these regressions are reported in Table 1.

Model (1) reports, for each treatment relative to Baseline (the omitted category), the factor changes in the odds of registering in the period before our intervention (i.e. between January 2 and March 8). The sample includes all students in treated buildings who were unregistered on January 2. In the Baseline treatment, the odds of registering in the pre-intervention period are 0.174, i.e. there are expected to be approximately 6 unregistered students for each registered student in our benchmark

[63] We included this dummy because we conjectured that there could be some difference in ethos or culture which could be relevant to the registration decision, comparing older and more newly established colleges.

[64] That is, each college, hall, residence block, and/or cross-college accommodation is treated as providing a cluster of observations (see Appendix B for further details).

condition. The odds of registering are very similar in the other treatments: in all cases the factor changes in the odds are close to 1 and none of the treatment variables are significant (all p-values > 0.532). This confirms that registration rates are indistinguishable across treatments in the pre-intervention period.

Table 1: The effects of treatments on registration rates

Dependent variable	Registered (1 if yes, 0 otherwise)	
	(1)	(2)
	Before Intervention (Jan 2 – Mar 8)	After Intervention (Mar 9 – Apr 20)
Monetary Loss	0.968 (0.250)	1.739*** (0.348)
Monetary Gain	1.123 (0.209)	0.749 (0.171)
Non-Monetary	0.930 (0.195)	1.009 (0.180)
Odds of registering in Baseline	*0.174**** *(0.064)*	*0.250**** *(0.072)*
Controls	Yes	Yes
N	8397	7679

Notes: The table reports odds ratios from logistic regressions. Note that a ratio greater than 1 implies a positive effect, whereas a ratio smaller than 1 implies a negative effect. The dependent variable indicates whether an individual registered within a given period. Robust standard errors clustered at the residence unit are reported in parentheses. Significance levels: *** p<0.01, ** p<0.05, * p<0.1

In model (2), we look at the effect of our different nudges after their implementation. The dependent variable now is whether an individual registered or not during the period between March 9 and April 20, the day of the registration deadline. The sample includes all students in treated buildings who were still unregistered on March 8 (i.e., we drop those who registered before the intervention, since they did not receive the postcards that were sent out on March 9). The treatment dummies therefore represent treatment differences in registration rates after the intervention. The odds of registering in Baseline are now 0.25: there are expected to be 4 unregistered students for each registered student in Baseline. The higher odds of registering in the post-intervention period relative to the pre-intervention period may reflect an impact of sending the postcard per se, or a natural increasing trend in registrations as the deadline for the General Election draws nearer.[65]

The Monetary Loss treatment increases the Baseline odds by a factor of 1.7 and the effect is significant at the 1% level. This implies an expected ratio of almost 2:1 between unregistered and registered students in the Monetary Loss treatment (the odds of registering are 0.25 x 1.739 = 0.43). In contrast, the Monetary Gain treatment reduces the odds of registering relative to Baseline by a factor of 0.749, although the effect does not reach statistical significance (p = 0.205). The Non-Monetary treatment has virtually no impact on the odds of registering, with a factor change close to 1 and very far from statistical significance (p = 0.958).[66]

Another interesting feature of the effect of the Monetary Gain Treatments is apparent in Figure 2. While at the beginning registration rates under the Monetary Gain treatments are very similar to the ones in Baseline, in the former they experience a less pronounced uptick towards the end of the post-intervention period, after the deadline for eligibility to enter the lottery has passed (on March 27).

To quantify the negative post-deadline effect of the Monetary Gain treatments, we split our data into two time periods. The first period spans from the day of our intervention until the lottery deadline (i.e. between March 8 and March 27). The second period starts from the first day after the lottery deadline until the registration deadline (i.e. between March 28 and April 20). As before, we apply logistic regression analysis. Note that we only use data from Baseline and Monetary Gain treatments, as the March 27 deadline is inconsequential in the other treatments. The results of this analysis are presented in Table 2. Model (1) reports the factor changes in the odds of registering in Monetary Gain relative to Baseline in the first period. The sample includes all students in treated buildings who were unregistered on March 8. In Model (2), we instead only look at the period after the deadline for the lottery

[65] We cannot distinguish between these two explanations because, in designing our treatments in collaboration with OCC, we agreed not to have a treatment where no postcard was sent.

[66] To test the robustness of these results, for the post-intervention period we also apply duration analysis to examine the time it takes an individual to register after having received our postcards. The results of the duration analysis are consistent with those of the logistic regressions, as we show in Appendix C, Table C2 (model (1)).

has passed already. The sample now includes all students in treated buildings who were still unregistered on March 28, i.e., we drop those subjects who registered before the lottery deadline. As before, we cluster standard errors by unit of residence..

Table 2: The effect of the lottery deadline on registrations relative to Baseline

Dependent variable:	Registered (1 if yes, 0 otherwise)	
	(1) Before Deadline (Mar 8– Mar 27)	(2) After Deadline (Mar 28 – Apr 20)
Monetary Gain	1.286 (0.463)	0.657* (0.425)
Monetary Gain x Post deadline		
Odds of registering in Baseline	*0.070**** *(0.047)*	*0.096**** *(0.026)*
Controls	*Yes*	*Yes*
N	3760	3579

Notes: The table reports odds ratios from logistic regressions. Note that a ratio greater than 1 implies a positive effect, whereas a ratio smaller than 1 implies a negative effect. The dependent variable indicates whether an individual registered within a given period. Only data from Monetary Gain and Baseline are included. Robust standard errors clustered at the hall level are in parentheses. Significance levels: *** p<0.01, ** p<0.05, * p<0.1

The results of model (1) reveal that before the end of the lottery deadline being subject to the Monetary Gain treatment slightly increases the odds of registering relative to being subject to the Baseline treatment, by a factor of 1.286, but the effect does not reach statistical significance (p = 0.485). In contrast, once the deadline for the lottery has passed, i.e., monetary incentives for registering have been removed again, being subject to the Monetary Gain treatment reduces the odds of registering relative to being subject to the Baseline treatment by a factor of 0.657, and this effect is statistically significant (p = 0.057).[67] This suggests that implementing monetary incentives has no positive effect while in place, but removing the monetary incentives

[67] To test the robustness of this result, we again use duration analysis. As we show in Appendix C, Table C2 (model (2)), the result of the duration analysis is consistent with that of the logistic regression analysis.

afterwards has a negative effect on registration rates. This would be in line with the literature on motivational crowding out (e.g. Deci And Ryan, 1985; Frey and Oberholzer-Gee, 1997; Meier, 2007), However, it is also possible that the low post-deadline registration rate amongst subjects in the Monetary Gain treatments, relative to those in Baseline, was the result of individuals who otherwise would have registered during this period having been motivated by the lottery to already register before the deadline passed.

To summarise our main findings so far, we find that emphasising the possibility of being fined significantly raises registration rates, while attempting to invoke the foot-in-the-door effect makes no significant impact. Furthermore, using positive monetary incentives may actually backfire and even lower registration rates, especially after the incentives are removed.

4. Study II: The effects of the monetary nudges on social norms

The main finding of our field experiment is that the threat of a monetary fine is more effective in encouraging registrations than the chance of a monetary gain (which has initially no effect and later, once the possibility of gain is removed, a negative effect), or foot-in-the-door type interventions (which have no effect). There are a number of possible mechanisms that could explain why the threat of a fine is *more effective* than the chance of a gain overall (e.g., loss aversion; or a bias in beliefs whereby individuals assess the probability of being fined as higher than the probability of winning the lottery). While we do not rule out such mechanisms being at work in our data, other explanations would be required for how the offer of a monetary gain in the form of a lottery could have a negative effect on registration, even compared to Baseline. In this section, we explore a possible explanation that might organize the contrasting impacts of money gains and money losses in our design. Specifically, we investigate the potentially different effect of these interventions on *social norms*, i.e. collectively recognised rules of behaviour that define which actions are viewed as socially appropriate (Elster, 1989; Ostrom, 2000). We considered this to be a promising channel to pursue because, as we now explain, there are plausible and contrasting effects that our fine and lottery nudges could have had on the perceived social appropriateness of registration behaviour.

We conjecture that the Monetary Loss treatment may *strengthen* a pre-existing social norm that registering to vote is what one ought to do: emphasising that failing to register is against the law may solidify one's perception that such behaviour is socially inappropriate.[68] In contrast, the Monetary Gain treatments may *weaken* that same social norm – the offer of money for registering may suggest to recipients that registering is not something already unconditionally demanded of them by society. If social norms influence registration behaviour, such alterations of subjects' perceptions of them could directly affect their decisions over registration. Indeed, the failure of the Monetary Gain treatments is reminiscent of previous research showing the introduction of economic incentives can crowd out people's intrinsic motivation to behave pro-socially (e.g. Frey and Oberholzer-Gee, 1997; Ariely et al., 2009; Gneezy et al., 2011; Bowles and Polania-Reyes, 2012). A plausible mechanism behind the offer of financial reward crowding out intrinsic motivation to register could be the lottery's weakening effect on the social norm of registering.

4.1 Experimental design and procedures

To investigate this, we ran an online study, employing the social norm elicitation task pioneered by Krupka and Weber (2013). In this study, we first described to subjects the setting of our field experiment. We then exposed each subject to the postcard used in one of three treatments – Baseline, Monetary Loss and Monetary Gain 1[69] – and in each case measured the social norms they perceived pertaining to registration behaviour.

This study was run in June 2016, with subjects who were students at the University of Nottingham, recruited through ORSEE (Greiner, 2015), an online database of

[68] There is some debate over whether social norms and laws are substitutes, or whether laws directly shape norms. See, for instance, Posner (2009) and Benabou and Tirole (2011).

[69] We chose to focus only on the Monetary Loss and Gain treatments because in the field experiment the monetary interventions produced the most interesting effects on registration rates relative to Baseline. The Non-Monetary treatments had no significant effects on registration rates and were indeed highly ineffective in engaging subjects. We only focus on one version of the Monetary Gain treatments because in the field experiment the Monetary Gain 2 treatment was statistically indistinguishable from Monetary Gain 1 and we thought that Monetary Gain 1 was easier to describe to subjects.

experimental participants. Thus, the subjects would have been demographically similar to those in the field experiment, but would not have been previously exposed to the postcards. In total 189 subjects participated in the study: 65 were shown the Baseline postcard, 61 the Monetary Loss postcard, and 63 the Monetary Gain 1 postcard. The study was conducted using Qualtrics (Qualtrics, 2016), an online survey platform.

At the beginning of the experiment, subjects were told to '*Imagine that the date is March 8, 2015. There is an upcoming General Election on May 7, and a local council wants to encourage people to register to vote before the deadline on April 20.*'. They were further informed that the council is considering strategies to raise registration amongst students in university accommodation, where rates have been particularly low. They were then told that the council decides to send a card to every unregistered student living in university accommodation, and were presented with a picture of one of three cards. These were replicas of the postcards sent out to students in the Baseline, Monetary Loss and Monetary Gain 1 treatments (the only difference was that the cards were cropped to cut off the OCC logo).

Subjects were then asked to evaluate '*how socially appropriate most people would think it would be for a student, having received this card, to register to vote or not to register to vote.*'. Earlier in the instructions, we had defined social appropriateness as '*behaviour that you think most people would agree is the "correct" thing to do. Another way to think about what we mean is that if someone were to behave in a socially inappropriate way, then other people might be angry at them.*'.[70]

We then asked subjects to evaluate the social appropriateness of each action (register to vote, not register to vote) on a four-point scale, encompassing 'very socially appropriate', 'somewhat socially appropriate', 'somewhat socially inappropriate', and 'very socially inappropriate'. These evaluations were incentivised such that subjects were encouraged to coordinate on the social norm: we told subjects we would randomly select one of the two actions, and for this action, they would be eligible to

[70] This follows the experimental instructions introduced by Krupka and Weber (2013). See Appendix D for the experimental instructions and screenshots of the online survey.

receive a cash prize if their evaluation of its social appropriateness was the same as that chosen by the most other subjects.[71]

4.2 Results

To analyse the data, we follow Krupka and Weber (2013) in transforming the evaluations into numerical values. We assign evenly-spaced values of -1 for the rating 'very socially inappropriate', -0.33 for the rating 'somewhat socially inappropriate', 0.33 for the rating 'somewhat socially appropriate' and 1 for the rating 'very socially appropriate.' We then calculate the mean ratings for each action by subjects exposed to each treatment. The results are displayed in Figure 3.

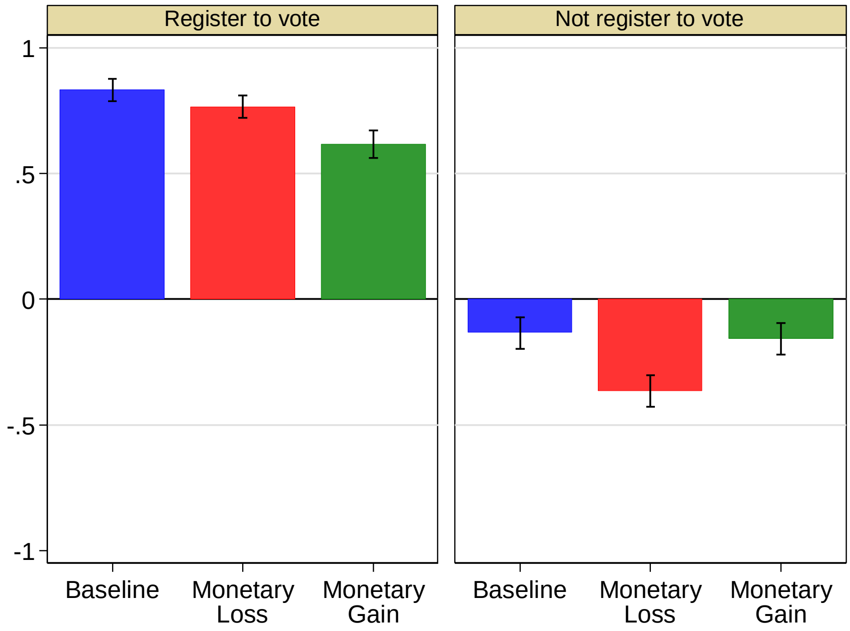
We find that regardless of the treatment subjects are exposed to, registering to vote tends to be seen as highly appropriate behaviour, while failing to register is generally seen as inappropriate. However, there are also subtle but significant treatment differences in peoples' appropriateness judgments. In particular, subjects exposed to the Monetary Gain treatment perceived registering to vote to be less appropriate than did those exposed to the Baseline treatment (two tailed Fisher Randomisation Test, $p = 0.012$).[72] Moreover, failing to register to vote was perceived to be more inappropriate by subjects exposed to the Monetary Loss treatment than it was by those exposed to the Baseline treatment ($p = 0.016$). In contrast, we find no significant differences in the perception of appropriateness of registering between

[71] As the study was very short and conducted online, we paid only one out of every eight subjects, determined retrospectively by a random lottery (subjects were informed about this at the beginning of the experiment). Those chosen for payment received an automatic £10, plus a further £30 if their evaluation matched that of the most other subjects in their treatment. Although most subjects would not be paid, the study was still incentivised to a conventional level: all subjects had a 1/8 chance of receiving between £10 and £40 for an approximately five-minute task.

[72] See Moir (1998) for a discussion of the randomisation test, and Kaiser and Lacy (2009) for information on the Stata command used to apply it. In Study II, we correct our p-values for the fact that we are testing multiple hypotheses relating to two interrelated dependent variables (the appropriateness of registering to vote and not registering to vote). For each dependent variable, we test Monetary Gain vs. Baseline and Monetary Loss vs. Baseline, a total of four tests. The correction method we use is that of Benjamini and Hochberg (1995) and corrected p-values are displayed in the text.

Monetary Loss and Baseline (*p* = 0.543), or the inappropriateness of non-registering between Monetary Gain and Baseline (*p* = 0.718).

Figure 3: Social appropriateness of registration behaviour by treatment



Notes: Figure 3 shows the mean appropriateness ratings assigned to each action (registering to vote, not registering to vote) by subjects exposed to the Monetary Loss, Baseline and Monetary Gain 1 postcards. Mean ratings are taken by assigning values of -1, -0.33, 0.33 and 1 for the ratings 'very inappropriate', 'somewhat inappropriate', 'somewhat appropriate' and 'very appropriate' respectively, and averaging the values for each action for all participants exposed to a given treatment. Error bars indicate standard errors of the mean.

Table 3 sheds light on how these treatment differences arise. It presents, for each treatment, the distribution of subjects' evaluations of the social appropriateness of each possible action. It shows that the lower perceived social appropriateness of registering to vote under Monetary Gain is driven by fewer subjects regarding registering as 'very socially appropriate' relative to Baseline (50.8% versus 78.1%). It can also be seen that the higher perceived social inappropriateness of failing to register under the Monetary Loss treatment is driven by more subjects regarding non-registering as 'very socially inappropriate' relative to Baseline (25.0% versus 10.9%), and by fewer regarding it as 'somewhat socially appropriate' (13.3% versus 28.1%).

To summarise the results of Study II, we find that subjects exposed to the Monetary Loss treatment perceived a relatively strong social norm against failing to register, while subjects exposed to the Monetary Gain treatment perceived a relatively weak social norm in favour of registering. This suggests that emphasising the fine for failing to register strengthens the social norm against such behaviour, while offering monetary incentives for successfully registering weakens the social norm demanding such behaviour. Given the strong evidence from previous studies (see introduction) that social norms influence economic behaviour, we propose that these normative effects explain at least part of the success of the Monetary Loss treatment and ineffectiveness of the Monetary Gain treatments.

Table 3: Distribution of social appropriateness evaluations

Appropriateness of registering to vote

	Very socially inappropriate	Somewhat socially inappropriate	Somewhat socially appropriate	Very socially appropriate
Baseline	1.6	0	20.3	78.1
Monetary Loss	0	1.7	31.7	66.7
Monetary Gain	0	8.2	41.0	50.8

Appropriateness of not registering to vote

	Very socially inappropriate	Somewhat socially inappropriate	Somewhat socially appropriate	Very socially appropriate
Baseline	10.9	54.7	28.1	6.3
Monetary Loss	25.0	58.3	13.3	3.3
Monetary Gain	9.8	60.7	23.0	6.6

Notes: Table 3 displays, by treatment, the percentage of subjects who evaluated registering to vote (top) or not registering to vote (bottom) as very socially inappropriate, somewhat socially inappropriate, somewhat socially appropriate, or very socially appropriate.

5. Conclusion

We investigated the effectiveness of different nudges aimed at raising voter registration rates. A unique feature of our study is that it combines two types of experiment: a natural field experiment to measure which nudge is most effective in raising registrations, and an online experiment to investigate possible reasons why different nudges may trigger different behavioural responses.

Our field experiment shows that highlighting to citizens the possibility of being fined for failing to register is an effective strategy for public bodies to use. The effect we identified by doing this was not only statistically significant, but also of a substantial magnitude: having the fine emphasised made subjects around 1.5 times more likely to register.

In contrast, we do not find evidence that offering financial inducements for registration is an effective strategy, nor do we find an effect of non-monetary interventions based on the foot-in-the-door technique (i.e., asking people to comply with small requests in order to get them agree to a subsequent larger request). The ineffectiveness of the latter can be explained by the fact that, of 3562 people who were asked to agree with our small requests, only 13 did so. The effect had no chance of taking hold, as we were unable to get our foot through the door in the first place.

The lack of success of our Monetary Gain treatments may represent another case of economic incentives crowding out people's intrinsic motivation to behave in a socially constructive way, to add to those uncovered in previous research (e.g. Frey and Oberholzer-Gee, 1997; Ariely et al., 2009; Gneezy et al., 2011; Bowles and Polania-Reyes, 2012). We note that, in this respect, our study's results differ from those of John et al. (2015), who also offered entry into a cash lottery as a reward for registering to vote in the UK. They found a small (approximately 4%) but significant positive effect of monetary rewards on registration rates. We conjecture that a plausible explanation for the contrast in results is that the maximum winnings offered by John et al. (2015) were much larger than ours (between £1000 and £5000). Their large material incentives may well have been enough to produce a positive effect, even if they had to overcome a crowding out effect (Bowles and Polania-Reyes, 2012). A tentative conclusion could then be that, when offering cash for registration or similar behaviours where there is danger of crowding out intrinsic motivations, one must 'pay enough or don't pay at all' (Gneezy and Rustichini, 2000). This would also be consistent with Panagopoulos (2013), who found that financial inducements raised voter turnout but only once they were sufficiently large. In our study registration rates amongst those offered financial incentives were particularly low once the opportunity for financial gain had passed; this may suggest– if our results generalise – that would-be nudgers seeking to use monetary incentives to promote specific behaviours, in environments where instrinsic motives can be crowded out, would be well warned against subsequently turning off those financial incentives.

Finally, our online experiment sheds light on a possible explanation for the contrasting effects that positive and negative monetary incentives have on voter registration rates. We propose that a plausible explanation for these effects may be via their influence on social norms. Our online experiment shows that emphasising the fine strengthens the perception that failing to register is socially inappropriate, while offering money for registering weakens the perception that registering is socially appropriate. We interpret this as evidence that social norms are a significant factor determining voter registration, a finding which is consistent with other recent experimental literature pointing to the importance of social norms as drivers of a wide range of behaviours (e.g. Burks and Krupka, 2012; Gachter et al., 2013; Krupka and Weber, 2013; Banerjee, 2016; Barr et al., 2016; Kimbrough and Vostroknutov, 2016; Krupka et al., 2016). A key novel contribution of our study in relation to this literature is in identifying the potentially important role of norms in determining the effectiveness of different types of nudge intervention.

Acknowledgements

This research was funded by the Economic and Social Research Council via the Network for Integrated Behavioural Sciences (Award No. ES/K002201/1), and by the Centre for Decision Research and Experimental Economics within the University of Nottingham's School of Economics. The research was done in collaboration with Oxford City Council's Electoral Services team, in particular Suzette Starmer and Martin John; and we received important assistance in Nottingham from Suzanne Robey. We received helpful input during the course of the project, and after presentations at Nottingham, from, amongst others, Abigail Barr, Robin Cubitt, Matthias Dahm, Gianni De Fraja, Markus Eberhardt and Konstantinos Tatsiramos.

References

Altmann, S., & Traxler, C. (2014). Nudges at the Dentist. *European Economic Review*, *72*, 19-38.

Ariely, D., Bracha, A., & Meier, S. (2009). Doing good or doing well? Image motivation and monetary incentives in behaving prosocially. *The American Economic Review*, *99*(1), 544-555.

Balliet, D., Mulder, L. B., & Van Lange, P. A. (2011). Reward, punishment, and cooperation: a meta-analysis. *Psychological bulletin*, *137*(4), 594.

Banerjee, R. (2016). On the interpretation of bribery in a laboratory corruption game: Moral frames and social norms. *Experimental Economics*, *19*(1), 240-267.

Barr, A., Lane, T., & Nosenzo, D. (2015). On the social appropriateness of discrimination. *Manuscript, University of Nottingham*.

Behavioural Insights Team (2012). Applying behavioural insights to reduce fraud, error and debt. *London: Cabinet Office.*

Behavioural Insights Team (2016). Applying behavioural insights to regulated markets. *London: Cabinet Office*.

Behavioural Insights Team (2010). Applying behavioural insight to health. *London: Cabinet Office*.

Behavioural Insights Team (2011). Behaviour change and energy use. *London: Cabinet Office*.

Benabou, R., & Tirole, J. (2011). *Laws and norms* (No. w17579). National Bureau of Economic Research.

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)*, 289-300.

Bennion, E. A., & Nickerson, D. W. (2011). The cost of convenience: An experiment showing e-mail outreach decreases voter registration. *Political Research Quarterly*, *64*(4), 858-869.

Bite the Ballot (2016). *Getting the missing millions back on the electoral register. A vision of voter registration reform in the United Kingdom. Draft Report.* Accessed (July 2016): https://drive.google.com/file/d/0B8L8l8Sw8aKVaG5RR1V3Rmw4Sk0/view?pref=2&pli=1

Blumenthal, M., Christian, C., Slemrod, J., & Smith, M. G. (2001). Do normative appeals affect tax compliance? Evidence from a controlled experiment in Minnesota. *National Tax Journal*, 125-138.

Bowles, S., & Polania-Reyes, S. (2012). Economic incentives and social preferences: substitutes or complements? *Journal of Economic Literature*, *50*(2), 368-425.

Bryan, C. J., Walton, G. M., Rogers, T., & Dweck, C. S. (2011). Motivating voter turnout by invoking the self. *Proceedings of the National Academy of Sciences*, *108*(31), 12653-12656.

Burger, J. M. (1999). The foot-in-the-door compliance procedure: A multiple-process analysis and review. *Personality and Social Psychology Review*, *3*(4), 303-325.

Burks, S. V., & Krupka, E. L. (2012). A multimethod approach to identifying norms and normative expectations within a corporate hierarchy: Evidence from the financial services industry. *Management Science*, *58*(1), 203-217.

Bursztyn L., Fiorin, S., Gottlieb, D., & Kanz, M. (2015). Moral Incentives: Experimental Evidence from Repayments of an Islamic Credit Card. *NBER Working Paper No. 21611*.

Chetty, R. (2015). Behavioral economics and public policy: A pragmatic perspective. *The American Economic Review*, *105*(5), 1-33.

Davenport, T. C. (2010). Public accountability and political participation: Effects of a face-to-face feedback intervention on voter turnout of public housing residents. *Political Behavior*, *32*(3), 337-368.

Elster, J. 1989. Social Norms and Economic Theory. *The Journal of Economic Perspectives* 3(4), 99–117.

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

Freedman, J. L., & Fraser, S. C. (1966). Compliance without pressure: the foot-in-the-door technique. *Journal of Personality and Social Psychology*, *4*(2), 195.

Frey, B. S., & Oberholzer-Gee, F. (1997). The cost of price incentives: An empirical analysis of motivation crowding-out. *The American Economic Review*, *87*(4), 746-755.

Gächter, S., Nosenzo, D., & Sefton, M. (2013). Peer effects in pro-social behaviour: Social norms or social preferences? *Journal of the European Economic Association*, *11*(3), 548-573.

Gerber, A. S. & Green, D. P. (2000). The effects of canvassing, telephone calls, and direct mail on voter turnout: A field experiment. *American Political Science Review, 94(3),* 653.

Gerber, A. S., Green D. P., & Larimer C.W. (2008). Social pressure and voter turnout: Evidence from a large-scale field experiment. *American Political Science Review. 102*: 33–48.

Gerber, A. S., Green, D. P., & Larimer, C. W. (2010). An experiment testing the relative effectiveness of encouraging voter participation by inducing feelings of pride or shame. *Political Behavior*, *32*(3), 409-422.

Gneezy, U. (2014). Incentives and behavior change - Put your money to work. Conference Presentation. http://www.behaviourworksaustralia.org/V2/wp-content/uploads/2014/11/Incentives-and-behavior-change-sep-14-13-short.pdf

Gneezy, U., Meier, S., & Rey-Biel, P. (2011). When and why incentives (don't) work to modify behavior. *The Journal of Economic Perspectives*, *25*(4), 191-209.

Gneezy, U., & Rustichini, A. (2000). Pay enough or don't pay at all. *The Quarterly Journal of Economics*, *115*(3), 791-810.

Greenwald, A. G., Carnot, C. G., Beach, R., & Young, B. (1987). Increasing voting-behaviour by asking people if they expect to vote. *Journal of Applied Psychology, 72(2),* 315-318.

Greenwald, A. G., Klinger, M. R., Vande Kamp, M. E., & Kerr, K. L. (1988). The self-prophecy effect: Increasing voter turnout by vanity-assisted consciousness raising. Unpublished manuscript, University of Washington.

Greiner, B. (2015) Subject pool recruitment procedures: organizing experiments with ORSEE. *Journal of the Economic Science Association*, 1(1), 1-12.

Hallsworth, M. (2014). The use of field experiments to increase tax compliance. *Oxford Review of Economic Policy*, 30(4), 658–679.

Hallsworth, M., List, J., Metcalfe, R., & Vlaev, I. (2014). *The behavioralist as tax collector: Using natural field experiments to enhance tax compliance* (No. w20007). National Bureau of Economic Research.

Imas, A., Lam, D., & Wilson, A. J. (2016). The anticipation and realization of regret: Differential effects for lottery incentives in one-shot and repeated settings. *Working paper*

John, P., MacDonald, E., & Sanders, M. (2015). Targeting voter registration with incentives: A randomized controlled trial of a lottery in a London borough. *Electoral Studies*, *40*, 170-175.

Kaiser, J., & Lacy, M. G. (2009). A general-purpose method for two-group randomization tests. *Stata Journal*, *9*(1), 70.

Kimbrough, E.O., and A. Vostroknutov. 2016. Norms Make Preferences Social. *Journal of the European Economic Association* 14, 608–638.

Kleven, H. J., Knudsen, M. B., Kreiner, C. T., Pedersen, S., & Saez, E. (2011). Unwilling or unable to cheat? Evidence from a tax audit experiment in Denmark. *Econometrica*, *79*(3), 651-692.

Krupka, E. L., Leider, S., & Jiang, M. (2016). A meeting of the minds: informal agreements and social norms. *Management Science*.

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.

Loomes, G., & Sugden, R. (1982). Regret theory: An alternative theory of rational choice under uncertainty. *The economic journal*, *92*(368), 805-824.

Lu F., Zhang, J., & Perloff, J. (2016). General and specific information in deterring traffic violations: Evidence from a randomized experiment. *Journal of Economic Behavior & Organization*, 123, 97–107.

Meier, S. (2007). Do subsidies increase charitable giving in the long run? Matching donations in a field experiment. *Journal of the European Economic Association*, *5*(6), 1203-1222.

Moir, R. (1998). A Monte Carlo analysis of the Fisher randomization technique: reviving randomization for experimental economists. *Experimental Economics*, *1*(1), 87-100.

Nosenzo, D. (2016). Employee incentives: Bonuses or penalties? *IZA World of Labor*.

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

Panagopoulos, C. (2010). Affect, social pressure and prosocial motivation: Field experimental evidence of the mobilizing effects of pride, shame and publicizing voting behavior. *Political Behavior*, *32*(3), 369-386.

Panagopoulos, C. (2013). Extrinsic rewards, intrinsic motivation and voting. *The Journal of Politics*, *75*(01), 266-280.

Posner, E. A. (2009). *Law and social norms*. Harvard university press.

Qualtrics (2016). Data for this paper was generated using Qualtrics software. Copyright © [2016] Qualtrics. Qualtrics and all other Qualtrics product or service names are registered trademarks or trademarks of Qualtrics, Provo, UT, USA. http://www.qualtrics.com

Ramirez, R. (2005). Giving voice to Latino voters: A field experiment on the effectiveness of a national nonpartisan mobilization effort. *The Annals of the American Academy of Political and Social Science*, *601*(1), 66-84.

Rogers, T., Fox, C. R., & Gerber, A. S. (2013). Rethinking why people vote. *The behavioral foundations of public policy*, 91.

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

Smith, J. K., Gerber, A. S., & Orlich, A. (2003). Self-Prophecy Effects and Voter Turnout: An Experimental Replication. *Political Psychology, 24(3)*, 593-604.

Thaler, R. H., & Sunstein, C. R. (2008). *Nudge: Improving Decisions About Health, Wealth, and Happiness.* Yale University Press.

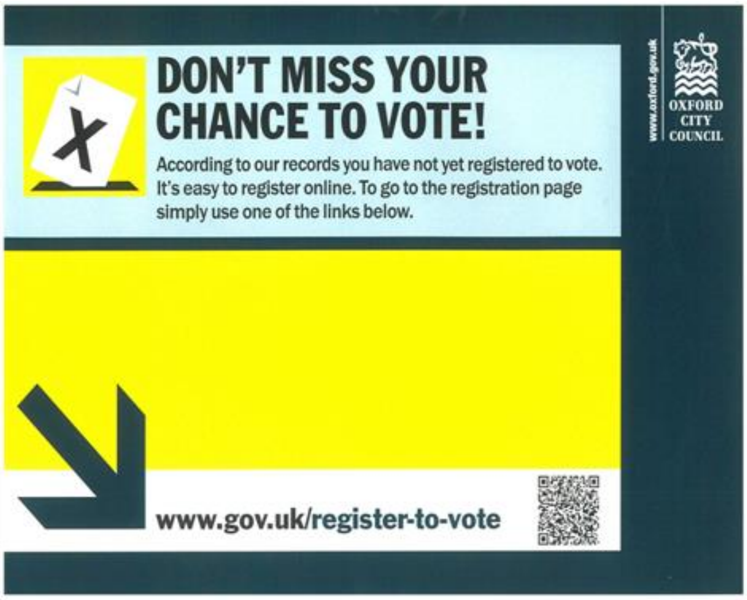
Van Lange, P. A., Rockenbach, B., & Yamagishi, T. (Eds.). (2014). *Reward and punishment in social dilemmas*. Oxford University Press.

Zeelenberg, M., & Pieters, R. (2004). Consequences of regret aversion in real life: The case of the Dutch postcode lottery. *Organizational Behavior and Human Decision Processes*, *93*(2), 155-168.

Online Appendices

Appendix A: Postcard Intervention in Study I

<u>Baseline Treatment</u>



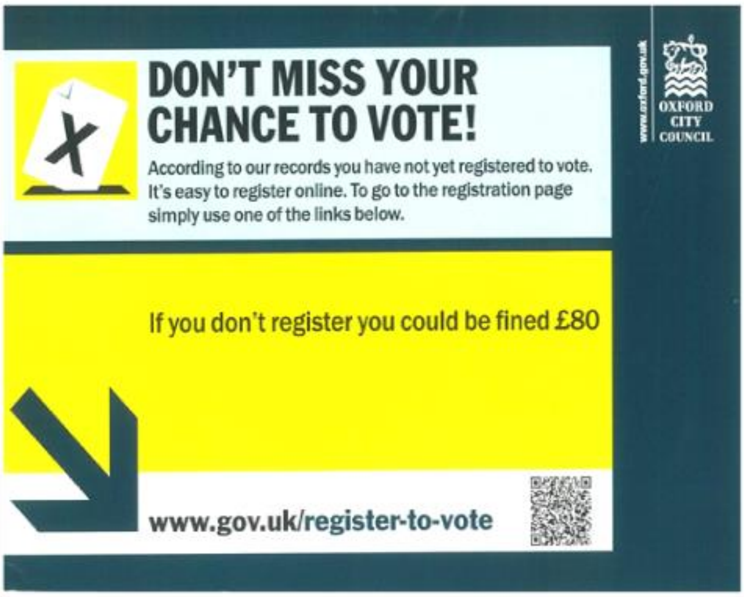
DON'T MISS YOUR CHANCE TO VOTE!

According to our records you have not yet registered to vote. It's easy to register online. To go to the registration page simply use one of the links below.

www.gov.uk/register-to-vote

www.oxford.gov.uk
OXFORD CITY COUNCIL

<u>Monetary Loss treatment</u>



DON'T MISS YOUR CHANCE TO VOTE!

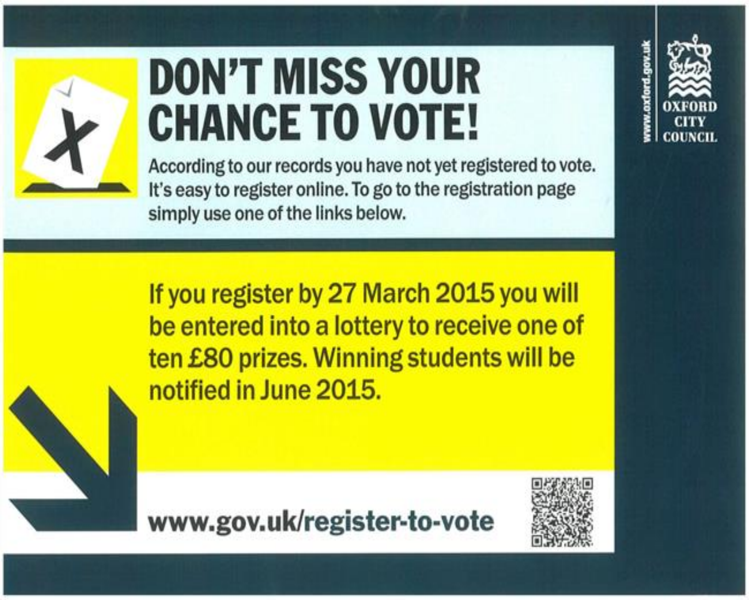
According to our records you have not yet registered to vote. It's easy to register online. To go to the registration page simply use one of the links below.

If you don't register you could be fined £80

www.gov.uk/register-to-vote

www.oxford.gov.uk
OXFORD CITY COUNCIL

Monetary Gain Treatment 1



DON'T MISS YOUR CHANCE TO VOTE!

According to our records you have not yet registered to vote. It's easy to register online. To go to the registration page simply use one of the links below.

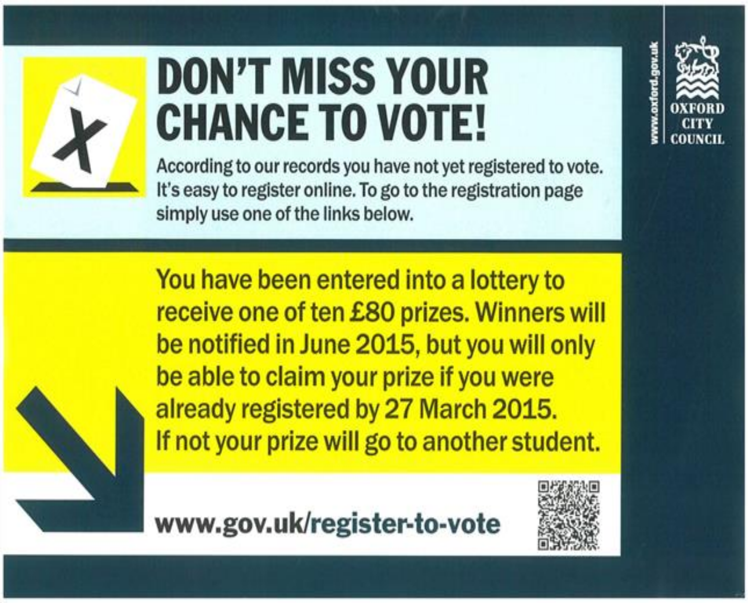
If you register by 27 March 2015 you will be entered into a lottery to receive one of ten £80 prizes. Winning students will be notified in June 2015.

www.gov.uk/register-to-vote

www.oxford.gov.uk

OXFORD CITY COUNCIL

Monetary Gain Treatment 2



DON'T MISS YOUR CHANCE TO VOTE!

According to our records you have not yet registered to vote. It's easy to register online. To go to the registration page simply use one of the links below.

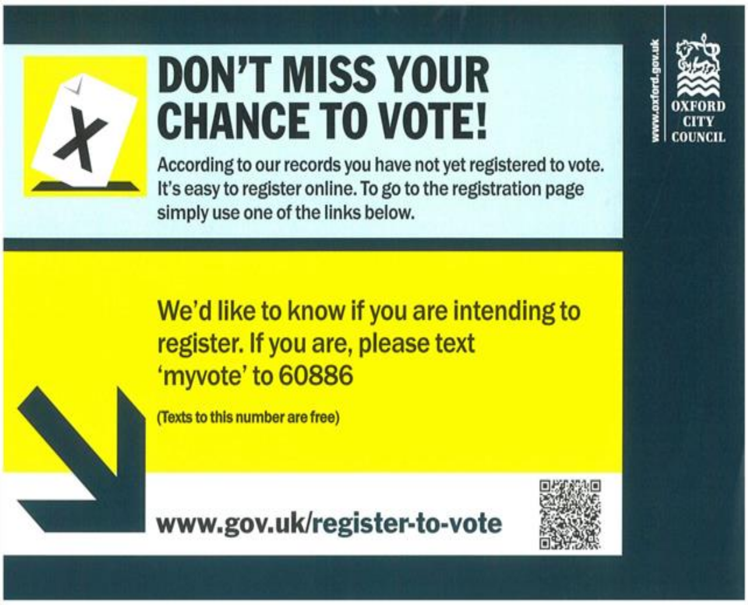
You have been entered into a lottery to receive one of ten £80 prizes. Winners will be notified in June 2015, but you will only be able to claim your prize if you were already registered by 27 March 2015. If not your prize will go to another student.

www.gov.uk/register-to-vote

www.oxford.gov.uk

OXFORD CITY COUNCIL

Non-monetary Treatment 1



DON'T MISS YOUR CHANCE TO VOTE!

According to our records you have not yet registered to vote. It's easy to register online. To go to the registration page simply use one of the links below.

We'd like to know if you are intending to register. If you are, please text 'myvote' to 60886

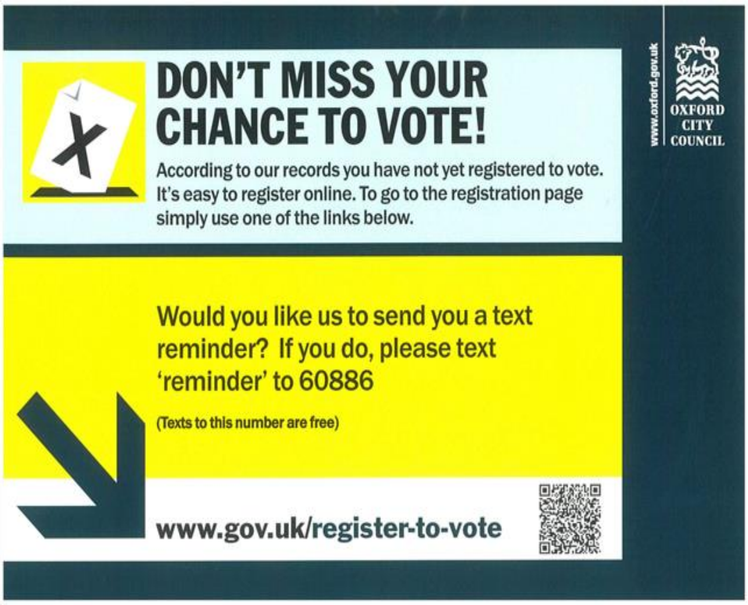
(Texts to this number are free)

www.gov.uk/register-to-vote

www.oxford.gov.uk

OXFORD CITY COUNCIL

Non-monetary Treatment 2



DON'T MISS YOUR CHANCE TO VOTE!

According to our records you have not yet registered to vote. It's easy to register online. To go to the registration page simply use one of the links below.

Would you like us to send you a text reminder? If you do, please text 'reminder' to 60886

(Texts to this number are free)

www.gov.uk/register-to-vote

www.oxford.gov.uk

OXFORD CITY COUNCIL

Appendix B: Details on assignment to treatment

Table B1 presents a breakdown, by university and place of residence, of the subjects assigned to each treatment. As stated in Section 2.3, we randomised assignment to treatment at the building level. For the University of Oxford, all students living in a single college were assigned to the same treatment. Two university-owned buildings (140 Walton St. and Castle Mill) which house students from various colleges were treated as if they were colleges, with common treatment assignment within each building. Two colleges, St. Cross and Brasenose, share a residence building (the St. Cross/Brasenose Annex). St. Cross and Brasenose were assigned to the same treatment. For the purposes of clustering in our regression analysis, 140 Walton St., Castle Mill and the St. Cross/Brasenose Annex are all treated as independent units of residence.

For Oxford Brookes University, all students living in a single hall of residence were assigned to the same treatment, with the exception that two very large halls, Cheney and Clive Booth, which were split into several units of assignment. Cheney and Clive Booth are naturally split into discrete residence blocks, so we subdivided them on this basis, ensuring maximum geographical distance between the subjects in these halls assigned to different treatments. For the purposes of clustering in our regression analysis, each subdivision that we split Cheney and Clive Booth into is treated as one unit of residence.

Beyond balancing assignment at the university level, the randomisation was also subject to the following constraints. Each treatment had to feature a mixture of large and small residence units; we ensured this by splitting the residence units into pools based on size, and assigned one unit from each pool to each treatment. Each treatment also had to feature a mixture of ancient and modern University of Oxford colleges, to the extent that the average college age in all treatments had to be within 200 years. Finally, the total number of subjects assigned to the largest treatment had to be no more than 15% greater than the number assigned to the smallest. We repeated the randomisation until it produced an assignment which met all the above criteria.

Our randomisation strategy was imperfectly implemented. Mailing errors resulted in some subjects being assigned to treatments we did not intend them to be (in italics in

Table B1). This explains why 16 students in Lincoln College were assigned to the Monetary Loss treatment, while the other 132 were assigned to the Baseline treatment; and why 6 Kellogg College students were assigned to Non-Monetary Treatment 2, with the other 11 were assigned to the Monetary Loss Treatment. As a robustness check, we re-ran our analysis excluding Lincoln and Kellogg from our dataset. All our results are robust and change only marginally.

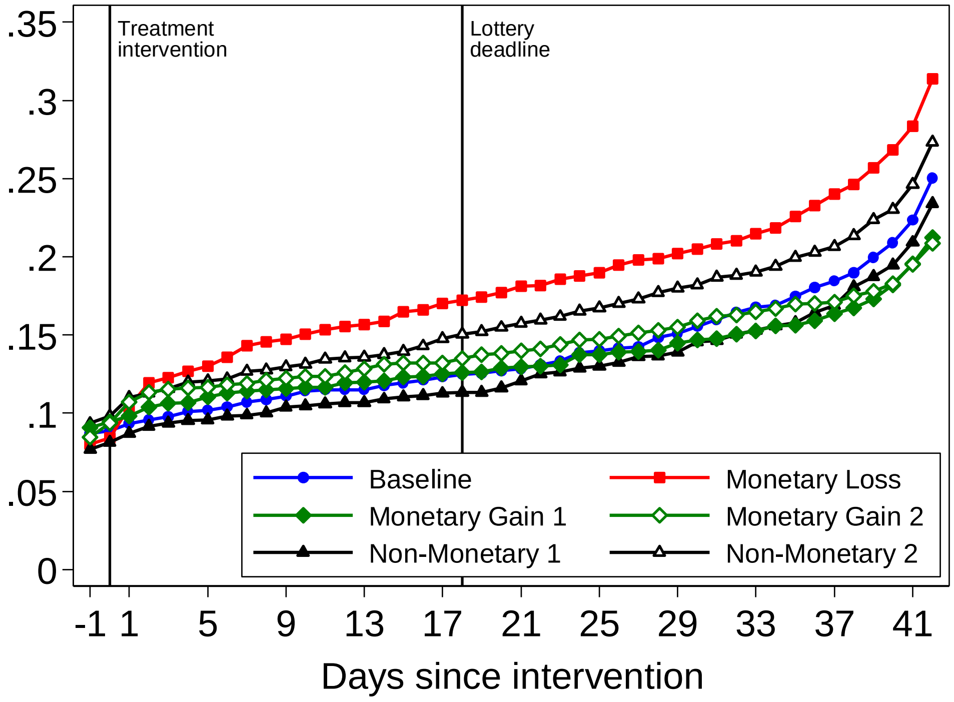
Table B1: Breakdown of treatment assignment by university and college/hall

Treatment	College/hall	Number assigned	Percentage of total assigned to treatment
Baseline	**University of Oxford**	**889**	**74.52**
	Keble	206	17.27
	Magdalen	201	16.85
	St Hugh's	200	16.76
	Lincoln	132	11.06
	Corpus Christi	120	10.06
	Harris Manchester	26	2.18
	All Souls	4	0.34
	Oxford Brookes	**304**	**25.48**
	Clive Booth	153	12.82
	Paul Kent	151	12.66
	Total	**1193**	**100.00**
Monetary Loss	**University of Oxford**	**1026**	**75.61**
	St. Catherine's	254	18.72
	Jesus	188	13.85
	Pembroke	188	13.85
	University College	147	10.83
	St. Peter's	146	10.76
	Green Templeton	71	5.23
	Lincoln	*16*	*1.18*
	Kellogg	11	0.81
	140 Walton Street	5	0.37
	Oxford Brookes	**331**	**24.39**
	Clive Booth	187	13.78
	Warneford	144	10.61
	Total	**1357**	**100.00**
Monetary Gain 1	**University of Oxford**	**940**	**75.20**
	Worcester	311	24.88
	Hertford	236	18.88
	Trinity	154	12.32
	Oriel	140	11.20
	Mansfield	74	5.92
	Linacre	25	2.00
	Oxford Brookes	**310**	**24.80**
	Clive Booth	162	14.56
	Cheney	128	10.24

	Total	**1250**	**100.00**
Monetary Gain 2	**University of Oxford**	**1044**	**79.27**
	St. Edmund	398	30.22
	Merton	253	19.21
	The Queen’s	209	15.87
	St. Hilda’s	156	11.85
	Nuffield	14	1.06
	Wycliffe	14	1.06
	Oxford Brookes	**273**	**20.73**
	Clive Booth	168	12.76
	Cheney	105	7.97
	Total	**1317**	**100.00**
Non-monetary 1	**University of Oxford**	**947**	**75.04**
	Christ Church	199	15.77
	New College	195	15.45
	Wadham	182	14.42
	Castle Mill	160	12.68
	Exeter	80	6.34
	Brasenose	74	5.86
	St. Cross/Brasenose	29	2.30
	St. Cross	28	2.22
	Oxford Brookes	**315**	**24.96**
	Crescent	212	16.80
	Cheney	103	8.16
	Total	**1262**	**100.00**
Non-monetary 2	**University of Oxford**	**979**	**75.31**
	St. Anne’s	216	16.62
	Balliol	192	14.77
	St. John’s	167	12.85
	Somerville	164	12.62
	Lady Margaret Hall	162	12.46
	Wolfson	72	5.54
	Kellogg	*6*	*0.46*
	Oxford Brookes	**321**	**24.69**
	Clive Booth	182	14.00
	Cheney	139	10.69
	Total	**1300**	**100.00**

Appendix C: Additional Analyses

Figure C1: Cumulative registration rates by treatment (uncombined)



Notes: Figure C1 shows, on a daily basis between the start of the intervention on March 9 and the registration deadline on April 20, the amount of registered students in the treated buildings, as a fraction of all students in these buildings who had been unregistered on January 2.

Table C1: The effects of treatments on registration rates (uncombined)

Dependent variable = 1 if registered, 0 if not	Logistic Regression	
	Before Intervention (Jan 8 – March 8)	After Intervention (March 9 – April 20)
Monetary Loss	0.901 (0.223)	1.565** (0.286)
Monetary Gain 1	1.037 (0.246)	0.707 (0.233)
Monetary Gain 2	0.935 (0.251)	0.706 (0.216)
Non-Monetary 1	0.869 (0.298)	0.943 (0.168)
Non-Monetary 2	1.068 (0.228)	1.136 (0.243)
Brookes Student	0.394*** (0.064)	0.671*** (0.084)
Constant	0.113*** (0.020)	0.238*** (0.039)
Wald-tests (p-values):		
Monetary Gain 1 = Monetary Gain 2	0.682	0.996
Non-Monetary 1 = Non-Monetary 2	0.514	0.226
N	8397	7679

Notes: Reported are odds ratios. Robust standard errors clustered at the residence unit are in parentheses. *** p<0.01, ** p<0.05, * p<0.1. The table further includes the p-values from Wald tests of the hypothesis that the coefficients differ between the two types of Monetary Gain treatments and the two types of the Non-Monetary treatments.

Duration analysis

To test the robustness of the results in Table 1, for the post-intervention period we apply duration analysis to examine the data underlying Figure 2 by taking into account the time it takes an individual to register after having received our postcards. Specifically, we use a Cox proportional hazard approach to model the duration until an individual registers. Estimated hazard ratios are reported in model (1) in Table C2. In line with the results from the logistic regression, we find a positive and significant effect of the Monetary Loss dummy. The hazard ratio of 1.656 indicates that - for those in the Monetary Loss treatment - the probability of registering on a certain date (conditional on not having been registered before) is 65.6% larger than in Baseline. The corresponding hazard ratios for the Monetary Gain and Non-Monetary dummies, in contrast, are not significantly different from 1, indicating no difference in the conditional probability of registering relative to Baseline for these two treatments.

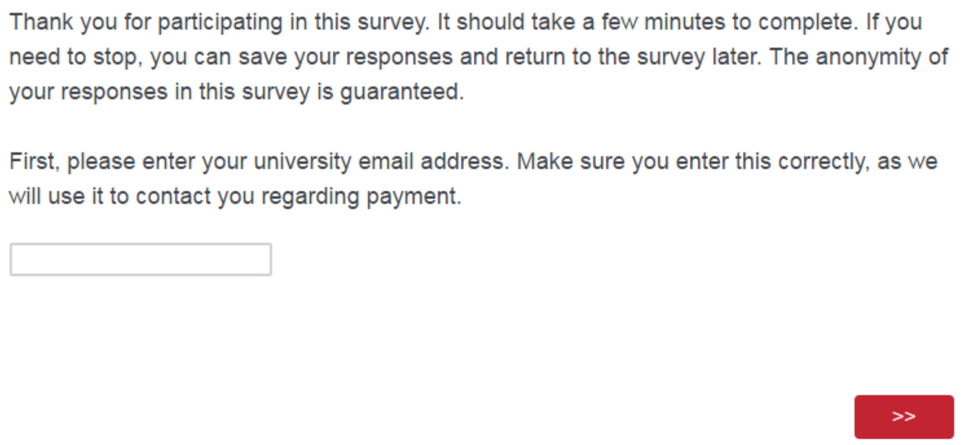
The robustness of the result in Table 2 is also tested using a Cox proportional hazard approach (model (2) in Table C2), which models the duration until an individual registers within the two post-intervention time periods, before and after the lottery deadline has passed. In line with the result from model (1) in Table 2, we find that the Monetary Gain dummy is insignificant (with a hazard ratio of 1.264), indicating that there are no pronounced differences in the conditional probability of registering between Baseline and Monetary Gain in the period where monetary incentives for registering are still in place. Furthermore, in line with the results from model (2) in Table 2 we find a significant interaction effect between the Monetary Gain dummy and a dummy for the post lottery deadline period. This indicates that the probability of registering (conditional on not having been registered before) in Monetary Gain relative to Baseline reduces once the deadline for entry into the lottery has passed.

Table C2: Duration analysis

Dependent variable	Duration to registration (1)	 (2)
	After Intervention (Mar 9 – Apr 20)	
Monetary Loss	1.656^{**} (0.302)	
Monetary Gain	0.782 (0.167)	1.264 (0.425)
Non-Monetary	1.019 (0.165)	
Monetary Gain x Post deadline		0.549** (0.147)
Controls	Yes	Yes
N	7679	3760

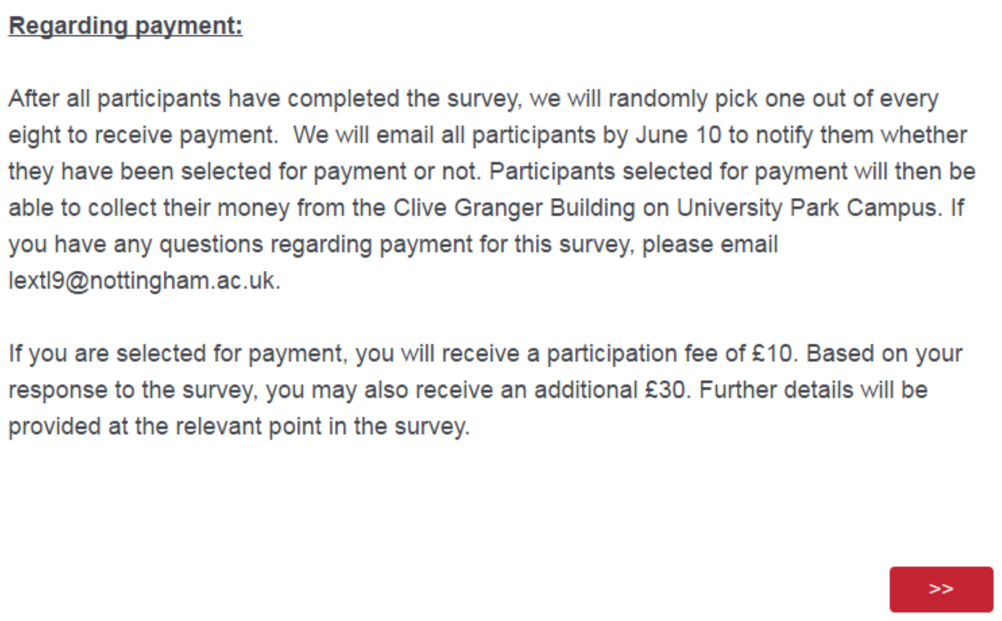
Notes: The table reports hazard ratios from estimations based on Cox proportional hazard models. Note that a ratio greater than 1 implies a positive effect, whereas a ratio smaller than 1 implies a negative effect. The dependent variable is the duration until an individual registers (i.e. the time between our treatment intervention and the date at which an individual registers). In model (2) only data from Monetary Gain and Baseline are included. Robust standard errors clustered at the residence unit are reported in parentheses. Significance levels: *** p<0.01, ** p<0.05, * p<0.1

Appendix D: Screenshots of Study II



Thank you for participating in this survey. It should take a few minutes to complete. If you need to stop, you can save your responses and return to the survey later. The anonymity of your responses in this survey is guaranteed.

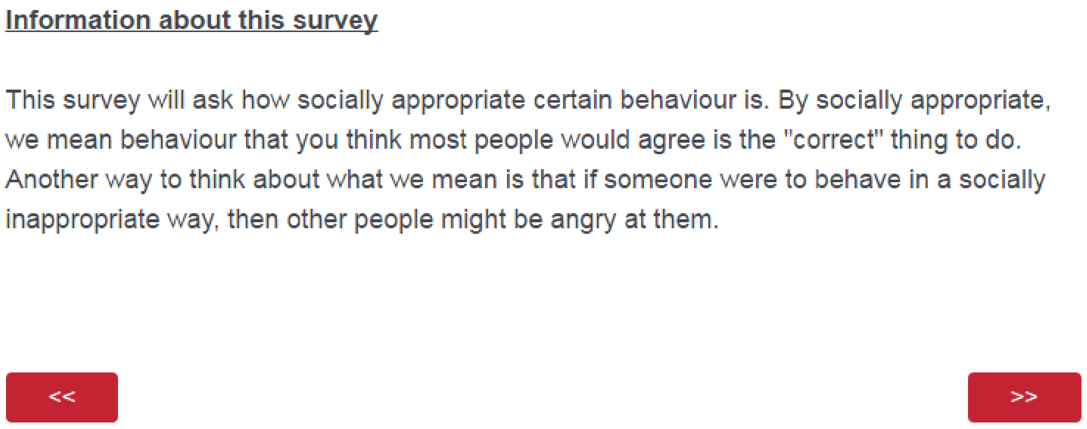
First, please enter your university email address. Make sure you enter this correctly, as we will use it to contact you regarding payment.



Regarding payment:

After all participants have completed the survey, we will randomly pick one out of every eight to receive payment. We will email all participants by June 10 to notify them whether they have been selected for payment or not. Participants selected for payment will then be able to collect their money from the Clive Granger Building on University Park Campus. If you have any questions regarding payment for this survey, please email lextl9@nottingham.ac.uk.

If you are selected for payment, you will receive a participation fee of £10. Based on your response to the survey, you may also receive an additional £30. Further details will be provided at the relevant point in the survey.

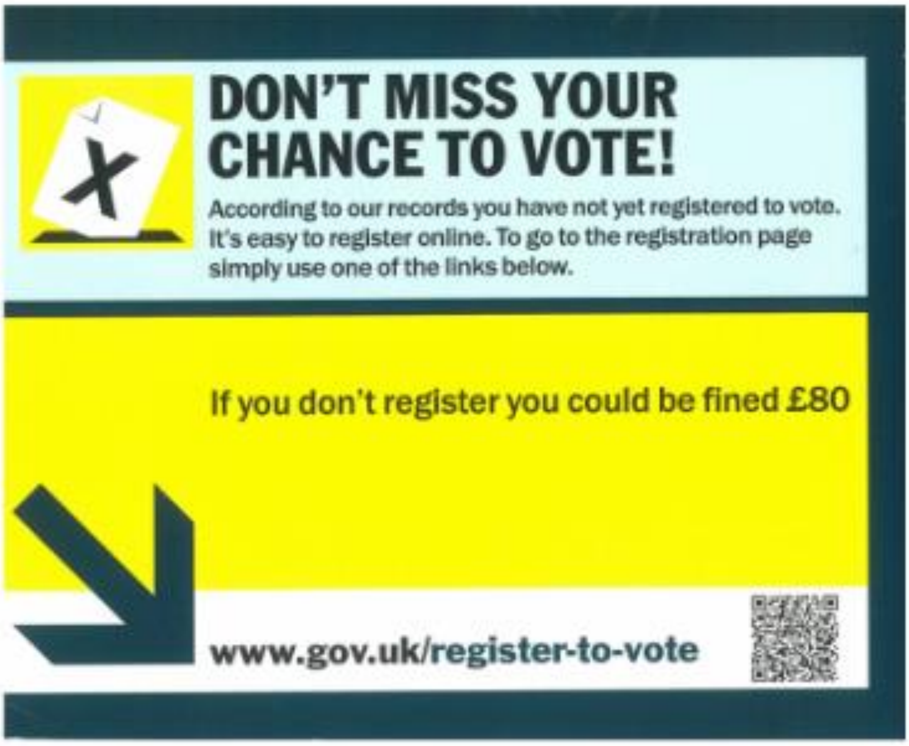


Information about this survey

This survey will ask how socially appropriate certain behaviour is. By socially appropriate, we mean behaviour that you think most people would agree is the "correct" thing to do. Another way to think about what we mean is that if someone were to behave in a socially inappropriate way, then other people might be angry at them.

Imagine that the date is March 8, 2015. There is an upcoming General Election on May 7, and a local council wants to encourage people to register to vote before the deadline on April 20. Registration rates are particularly low amongst students living in university accommodation, and the council is considering various strategies it can use to attempt to raise registration rates amongst these students.

The council decides to send every unregistered student living in university accommodation this card which includes a message urging students not to miss their chance to vote. The card also warns students that those who do not register may be fined £80.



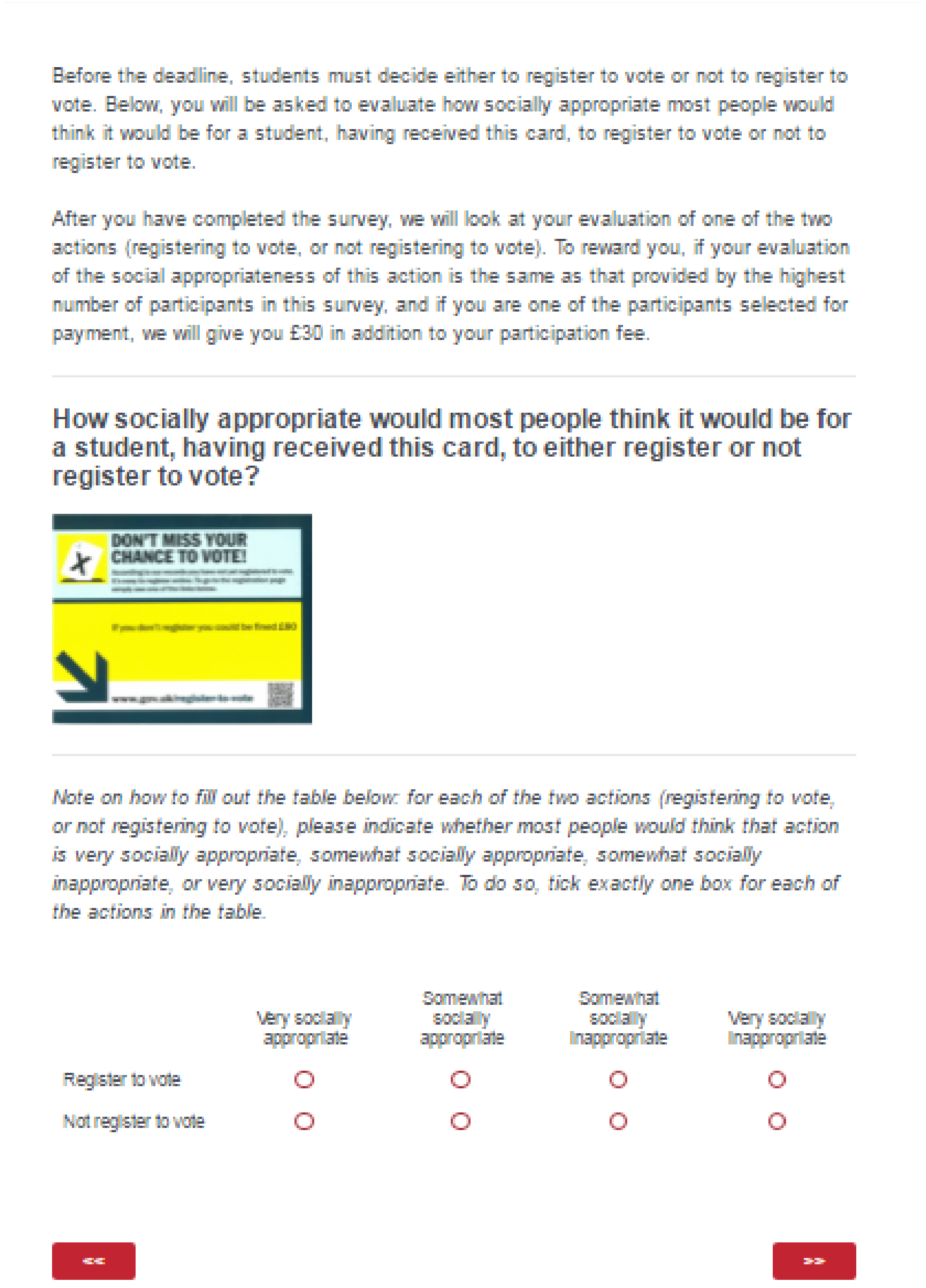
DON'T MISS YOUR CHANCE TO VOTE!

According to our records you have not yet registered to vote. It's easy to register online. To go to the registration page simply use one of the links below.

If you don't register you could be fined £80

www.gov.uk/register-to-vote

<< >>



Before the deadline, students must decide either to register to vote or not to register to vote. Below, you will be asked to evaluate how socially appropriate most people would think it would be for a student, having received this card, to register to vote or not to register to vote.

After you have completed the survey, we will look at your evaluation of one of the two actions (registering to vote, or not registering to vote). To reward you, if your evaluation of the social appropriateness of this action is the same as that provided by the highest number of participants in this survey, and if you are one of the participants selected for payment, we will give you £30 in addition to your participation fee.

How socially appropriate would most people think it would be for a student, having received this card, to either register or not register to vote?

DON'T MISS YOUR CHANCE TO VOTE!
If you don't register you could be fined £80
www.gov.uk/register-to-vote

Note on how to fill out the table below: for each of the two actions (registering to vote, or not registering to vote), please indicate whether most people would think that action is very socially appropriate, somewhat socially appropriate, somewhat socially inappropriate, or very socially inappropriate. To do so, tick exactly one box for each of the actions in the table.

	Very socially appropriate	Somewhat socially appropriate	Somewhat socially inappropriate	Very socially inappropriate
Register to vote	○	○	○	○
Not register to vote	○	○	○	○

Chapter Five: Conclusion

The first two substantive chapters of this dissertation (Chapters 2 and 3) are primarily focused on how and why the strength of discrimination varies across contexts. The most important results of Chapter 2 were: that lab experiments have found stronger discrimination between some types of identity groups than others; and that, in particular, minimal group experiments find stronger discrimination than one might expect. This finding was the central focus of investigation in Chapter 3. Here, we proposed and found support for one possible explanation for the result: that discrimination is perceived to be less socially inappropriate in some circumstances than in others, and that in those circumstances where it is less socially inappropriate people are more willing to engage in discrimination.

There is certainly scope for more future research on the relationship between discrimination and social norms. Field experiments could, for instance, investigate the effects on discrimination of nudges designed to reduce the social appropriateness of discriminating. There is also scope for more research investigating other possible reasons for contextual differences in levels of discrimination – in particular, why lab experiments yield relatively high levels of discrimination across minimal groups but relatively low levels across ethnic, national or religious groups. As I discuss in the first and second chapters, these phenomena may suggest lab experiments face difficulties in estimating levels of discrimination that are generalizable to the outside world. The external generalizability of lab estimates of discrimination could be studied by designing experiments to measure discrimination in both lab and field contexts while carefully holding other features of the decision-making context constant (see Stoop et al, 2012, and Stoop, 2014, for experiments following this approach in other research areas).

Chapter 4 of the dissertation departs from the topic of discrimination: the main focus here is on nudging in public policy. However, it also maintains an emphasis on social norms. Both Chapters 2 and 3 provide evidence to support the notion that social norms play a role in determining behavioural regularities. The norm-elicitation method we use in both cases – that of Krupka and Weber (2013) – is being widely adopted by experimental economists and there is, no doubt, scope for future research into the reliability of the method itself, as well as into other methods of measuring social norms. In particular, the Krupka-Weber method's use of monetary incentives to encourage coordination on the social norm has received some attention (e.g. Vesely, 2015) but further research could provide greater clarity on the effect of this design feature.

Chapter 4 provides a case where the context dependence of social norms can be exploited to achieve a policy goal. With current interest in social norms high, policymakers are increasingly aware of the possibilities offered by nudges harnessing the effects of social norms (e.g. Ayres et al, 2012; Behavioural Insights Team, 2011). Such social nudging could be implemented across a very wide range of policy domains. For instance, governments interested in reducing discrimination may find campaigns aimed at strengthening the taboo nature of such behaviour to be effective tools. Of course, we should have no presumptions about the preferences of policymakers. Indeed, those who wish to see discrimination eliminated may have as much reason to fear the potential for policymakers to influence social norms as to be thankful for it; it is, for instance, quite plausible that the apparent efforts of current prominent politicians around the world to normalise certain forms of discrimination could lead to an upsurge in their prevalence.

References

Ayres, I., Raseman, S., & Shih, A. (2012). Evidence from two large field experiments that peer comparison feedback can reduce residential energy usage. *Journal of Law, Economics, and Organization*, ews020.

Behavioural Insights Team (2011). Behaviour change and energy use. *London: Cabinet Office*.

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.

Stoop, J. (2014). From the lab to the field: envelopes, dictators and manners. *Experimental Economics*, *17*(2), 304-313.

Stoop, J., Noussair, C. N., & Van Soest, D. (2012). From the lab to the field: Cooperation among fishermen. *Journal of Political Economy*, *120*(6), 1027-1056.

Veselý, S. (2015). Elicitation of normative and fairness judgments: Do incentives matter?. *Judgment and Decision Making*, *10*(2), 191.

Acknowledgements

I would like to finally conclude this dissertation by thanking my PhD supervisors, Abigail Barr and Daniele Nosenzo. I am incredibly grateful for all your tireless work in supporting me throughout the course of my studies in Nottingham. I could not have asked for more wonderful supervisors. I am also very grateful to Felix Kölle and Chris Starmer, who co-authored the third chapter. My PhD was funded by the Network for Integrated Behavioural Science. I am particularly grateful to Robin Cubitt, who assisted me in my application for this grant. Finally, I am indebted to all staff and fellow students in the School of Economics, every one of whom has contributed to my learning here.