CENTRE FOR DECISION RESEARCH & EXPERIMENTAL ECONOMICS

The University of Nottingham
UNITED KINGDOM · CHINA · MALAYSIA

Discussion Paper No. 2017-11

Abigail Barr, Tom Lane &
Daniele Nosenzo

On the social inappropriateness of discrimination

June 2017

CeDEx Discussion Paper Series
ISSN 1749 - 3293

CENTRE FOR DECISION RESEARCH & EXPERIMENTAL ECONOMICS

The Centre for Decision Research and Experimental Economics was founded in 2000, and is based in the School of Economics at the University of Nottingham.

The focus for the Centre is research into individual and strategic decision-making using a combination of theoretical and experimental methods. On the theory side, members of the Centre investigate individual choice under uncertainty, cooperative and non-cooperative game theory, as well as theories of psychology, bounded rationality and evolutionary game theory. Members of the Centre have applied experimental methods in the fields of public economics, individual choice under risk and uncertainty, strategic interaction, and the performance of auctions, markets and other economic institutions. Much of the Centre's research involves collaborative projects with researchers from other departments in the UK and overseas.

Please visit http://www.nottingham.ac.uk/cedex for more information about the Centre or contact

Suzanne Robey
Centre for Decision Research and Experimental Economics
School of Economics
University of Nottingham
University Park
Nottingham
NG7 2RD
Tel: +44 (0)115 95 14763
suzanne.robey@nottingham.ac.uk

The full list of CeDEx Discussion Papers is available at

http://www.nottingham.ac.uk/cedex/publications/discussion-papers/index.aspx

On the social inappropriateness of discrimination

Abigail Barr[a], Tom Lane[b], Daniele Nosenzo[c]

17 June 2017

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

a: Department of Economics, University of Nottingham

b: Corresponding author: Tom Lane, Nottingham University Business School China, University of Nottingham Ningbo China, 199 Taikang East Road, Ningbo, China, 315100
Email address: Tom.Lane@nottingham.edu.cn

c: Department of Economics, University of Nottingham

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.[1] 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.[2] 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.[3]

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

[1] See Elster (1989) and Ostrom (2000) for definitions of social norms.
[2] See also Huang and Wu (1994) and Montgomery (1994) for related approaches.
[3] 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 restrictions on occupation, prohibit inter-caste marriage, and limit or prohibit access to public places and services.

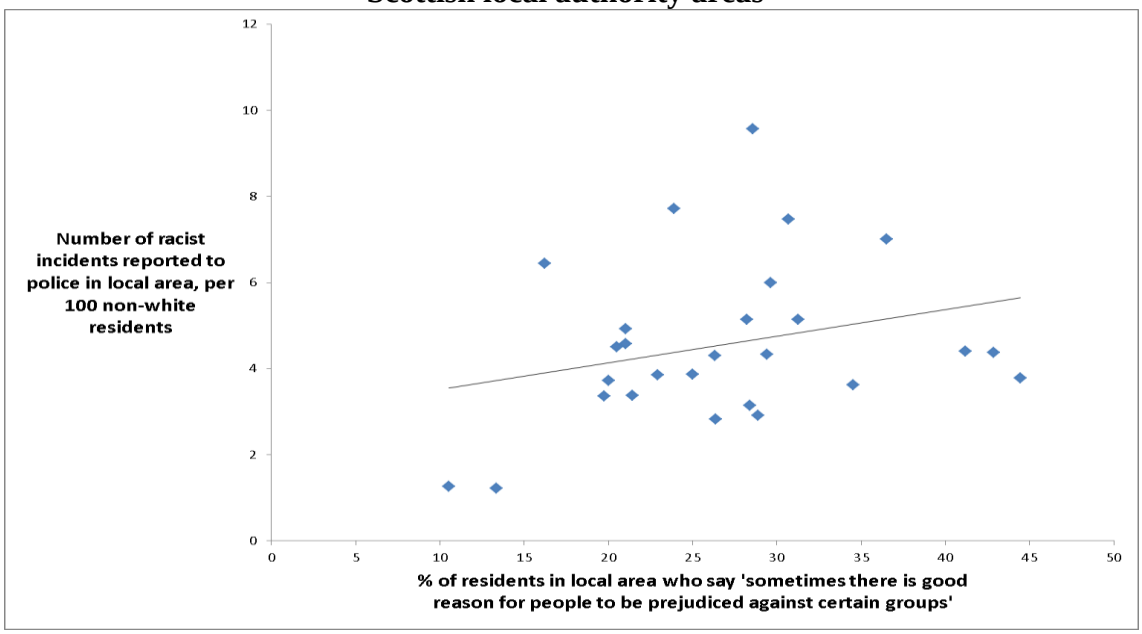
economic behaviour. Following Krupka and Weber (2013), lab and lab-in-the-field experiments have shown that in a variety of economic contexts people are more likely to take an action the more socially appropriate they perceive it to be (e.g. Burks and Krupka, 2012 – corporate ethics; Gachter et al, 2013 – gift-exchange; Krupka et al, 2016 – informal contract enforcement; Banerjee, 2016 – bribery). There is also evidence from econometric research (e.g. Buonanno et al, 2009) and natural field experiments (e.g. Allcott, 2011) suggesting norms drive economically-relevant behaviour. 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.[4]

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 variable against the number of racist incidents[5], per 100 non-white residents,[6] reported to the police in the financial year 2003-4 (Scottish Executive Statistical Bulletin, 2007). A correlation of 0.27 between the two variables suggests 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.

[4] See Krupka and Weber (2013) and Mackie et al. (2015) for a discussion of the difficulties of measuring social norms empirically.
[5] 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)
[6] The contemporaneous proportion of non-white residents in each Scottish local area is taken from the 2001 UK Census (National Records of Scotland, 2011).

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.

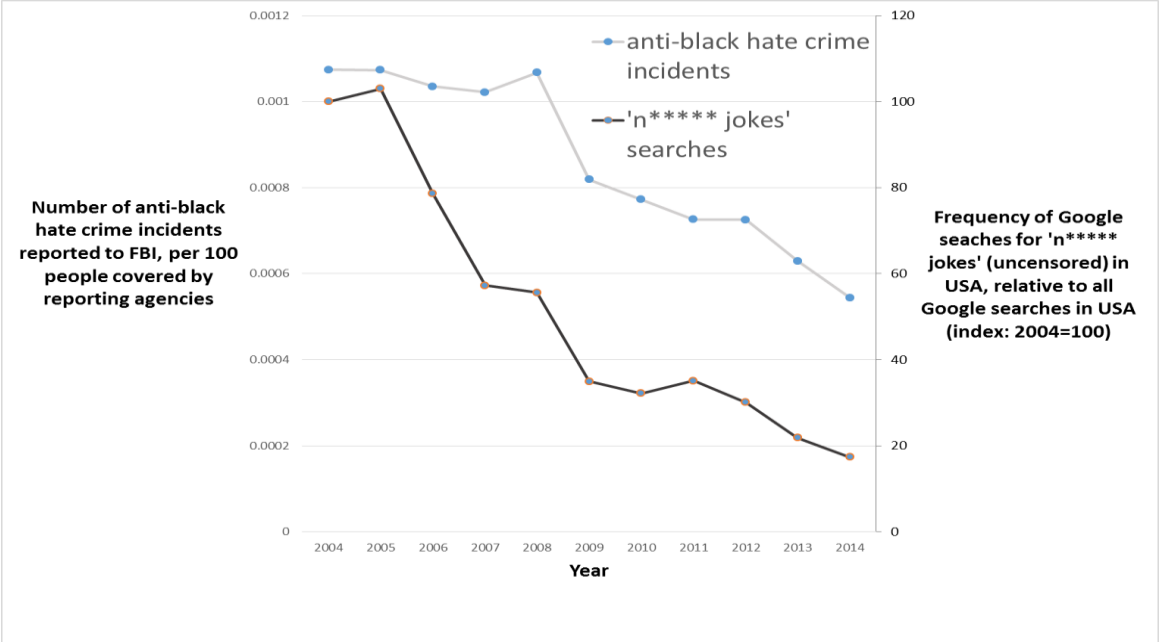
The acceptability of prejudiced-based humour has sometimes been used as a proxy for the normative appropriateness of discrimination (see, e.g., Crandall et al., 2002). Figure 2 plots, over the period 2004 to 2014, the frequency of Google searches in the US for 'N***** jokes' (we apply the censorship for this paper; the original search term was uncensored[7]), 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[8] (United States Department of Justice, 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

[7] 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.

[8] 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.

over time in the social appropriateness of discrimination against black people and the change over time in discriminatory behaviour against black people.

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, crucially, we exogenously 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.[9] 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 of balls randomly picked. If this is the case, our exogenous manipulation varies the strength of the norm relating to discrimination across our treatments and, if discrimination is systematically shaped by norms, we thus 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.

[9] 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.

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.

That discrimination can be observed along a trivial, artificially-induced dimension of identity highlights the strength of the human inclination to discriminate against out-group members, and the ease with which in-group bias can be triggered (Ashburn-Nardo et al, 2001). That we observe weaker discrimination when identity is based upon the more meaningful characteristic of nationality, and that such discrimination is perceived to be more socially inappropriate, suggests that the extent to which human society has been effective in curbing the inclination to discriminate is owing to the development of shared norms proscribing this behaviour.

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 (Democrat 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.[10] 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.

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.

[10] 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.

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, 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 for any given set of actions a.

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 thereby 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.[11] 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.[12] 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

[11] See Supplementary Online Materials A for a copy of the instructions used in the experiments.
[12] 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 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.

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 a 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 depend on what any other player does, so statistical beliefs about other players are irrelevant.[13] Second, to maximise our chances of discerning treatment differences, we wanted a task that reliably produces discriminatory behaviour; in a meta-analysis Lane (2016) 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

[13] 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.

evaluating. In our experiment, in contrast to the original Krupka and Weber method, participants in the norm-elicitation task were the same as those playing the allocator game. This allows us to look at within-individual correlations between norms and actions. To facilitate an investigation into whether this had implications for either the behavioural or normative data, we varied which task came first (participants were unaware of the content of the second task until they had completed the first).[14] 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 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

[14] See also Erkut et al (2015) and D'Adda et al (2016) for further evidence on the robustness of the Krupka and Weber elicitation task to order variation.

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 Netzer and Sutter, 2009; Guillen and Ji, 2011; Goerg et al, 2016). 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.[15] 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.

[15] 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.[16]

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. Goerg et al, 2016; 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 (2016), 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).[17] In our experiment, discrimination against people who randomly drew a ball of a 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

[16] 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).

[17] 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).

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.[18]

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 '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

[18] 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.

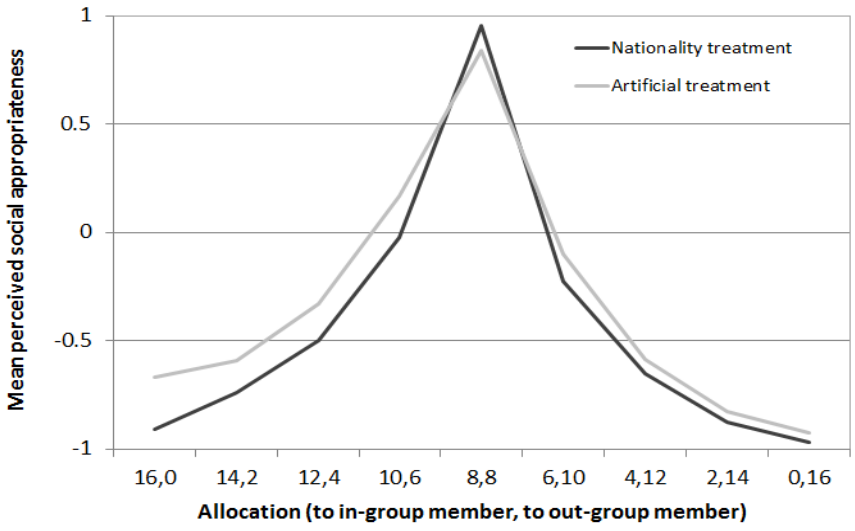
performed (in our case nine, one for each possible allocation) before dividing each by its rank – thus the greatest adjustment is made to smaller p-values.[19]

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.[20]

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 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.[21]

[19] 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.
[20] 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.
[21] In Supplementary Online Materials D we show that these treatment differences in the perceived norms are driven by systematic variations in subjects' 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**

	16,0	14,2	12,4	10,6	8,8	6,10	8,12	2,14	0,16
Allocation (to in-group member, to out-group member)									
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.

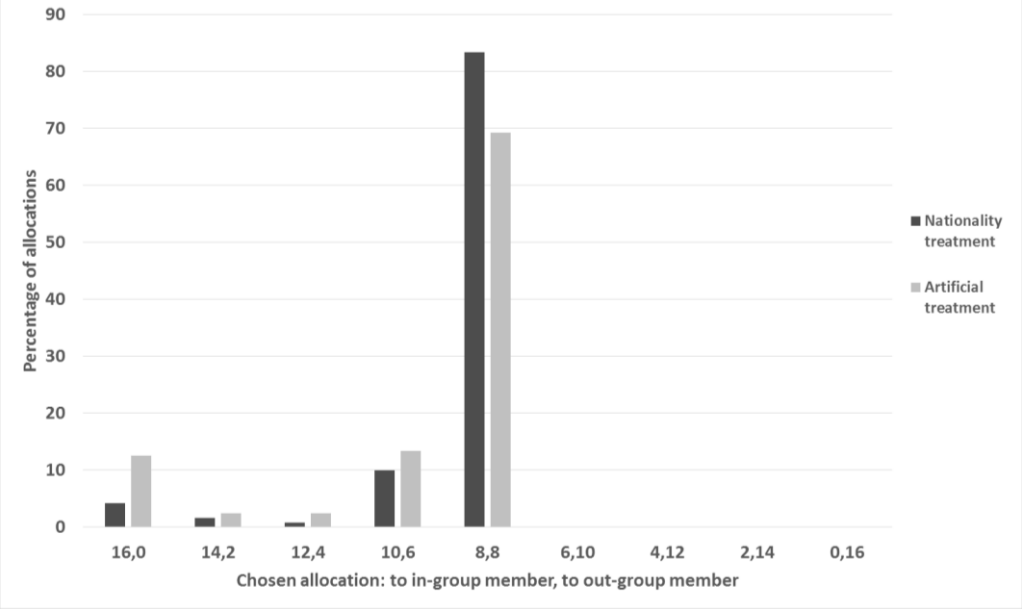
As a consequence, all in-group-favouring allocations are on average perceived to be more appropriate in the *Artificial* treatment, and the differences are statistically significant at the 5% level or lower in three out of four possible cases (the exception being the allocation 14, 2 for which the difference 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.

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.

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.[22],[23]

[22] As can be seen from Table 1, Chinese participants discriminated more than British participants. We also found that Chinese participants perceived discrimination to be less socially inappropriate than British participants did, in both the *Artificial* and *Nationality* treatments. These cross-national differences are more fully explored in Supplementary Online Materials E. One possible reason for the stronger discrimination, and more favourable perception of it by the Chinese in our setting could be their minority status within Britain – there is some evidence that belonging to a minority group may lead one to have stronger discriminatory preferences (Chen et al, 2014; Tanaka and Camerer, 2016)

[23] 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)
$\mathbf{R^2}$	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.[24]

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.[25]

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 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:

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

[24] 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.

[25] 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.

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, in Model (2), 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 the individual's normative evaluation of the action under consideration.

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

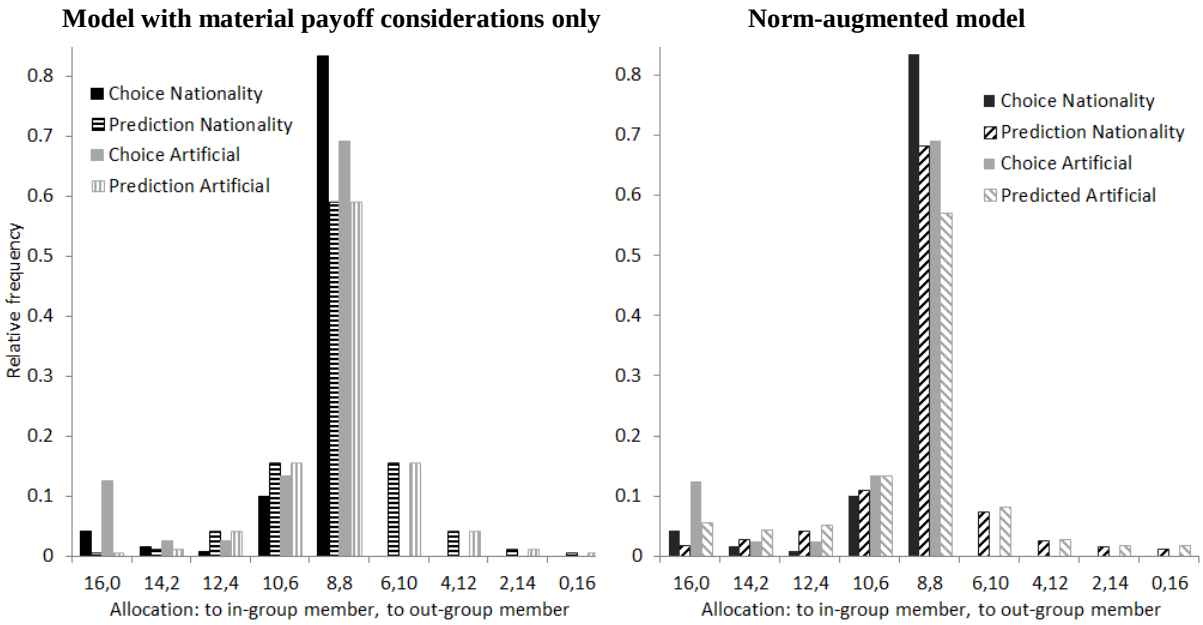
Dependent variable = 1 if action is chosen; 0 otherwise		
Model	(1)	(2)
v (weight on payoff inequality)	-0.338***	-0.111***
	(0.021)	(0.028)
γ (weight on normative prescriptions)		1.081***
		(0.119)
Pseudo R^2	0.424	0.511
Bayesian Information Criterion	615.17	531.02
Number of Observations	2,160	2,160

*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 that actions which yield larger payoff inequality are less likely to be chosen. The significant positive estimate of γ in model (2) indicates that an individual is more likely to choose actions he or she perceives to be more socially appropriate. 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.

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.

The model in which participants are identity-blind and care only about inequality fails to capture two 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-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.

Consistent with longstanding results from the minimal group literature, our study shows how remarkably easy it is to trigger discrimination between groups whose identities are based on artificial, trivial characteristics. That we find weaker discrimination on the basis of more meaningful identity characteristics such as nationalities, and that discrimination is perceived to be more socially inappropriate in that setting, suggests that shared norms opposing discrimination help moderate this most natural of human inclinations. A possible implication of this would be that if society allows such prohibitive social norms to be eroded by whatever means, discrimination will increase. This would be consistent, arguably, with the recent backlash in various western countries against 'political correctness', spurred on by leaders

promoting nationalism and identity politics, and the apparent rise in hostility towards immigrants and ethnic minorities in these countries.

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.

Allcott, H. (2011). Social norms and energy conservation. *Journal of Public Economics*, *95*(9–10): 1082-1095.

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.

Ashburn-Nardo, L., Voils, C. I., & Monteith, M. J. (2001). Implicit associations as the seeds of intergroup bias: How easily do they take root?. *Journal of personality and social psychology*, *81*(5), 789-799.

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.

Chen, Y., Li, S. X., Liu, T. X., & Shih, M. (2014). Which hat to wear? Impact of natural identities on coordination and cooperation. *Games and Economic Behavior*, *84*, 58-86.

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." *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.

Goerg, S. J., Hennig-Schmidt, H., Walkowitz, G., & Winter, E. (2016). In Wrong Anticipation-Miscalibrated Beliefs between Germans, Israelis, and Palestinians. *PLoS One*, *11*(6), e0156998.

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. (2016). Discrimination in the laboratory: a meta-analysis of economics experiments. *European Economic Review, 90, 375-402.*

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.

Tanaka, T., & Camerer, C. F. (2016). Trait perceptions influence economic out-group bias: lab and field evidence from Vietnam. *Experimental Economics*, *19*(3), 513-534.

United States Department of Justice, Federal Bureau of Investigation. (2015). *Hate Crime Statistics, 2004-2014*. Retrieved (8th June, 2016), from (https://ucr.fbi.gov/hate-crime/).

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.

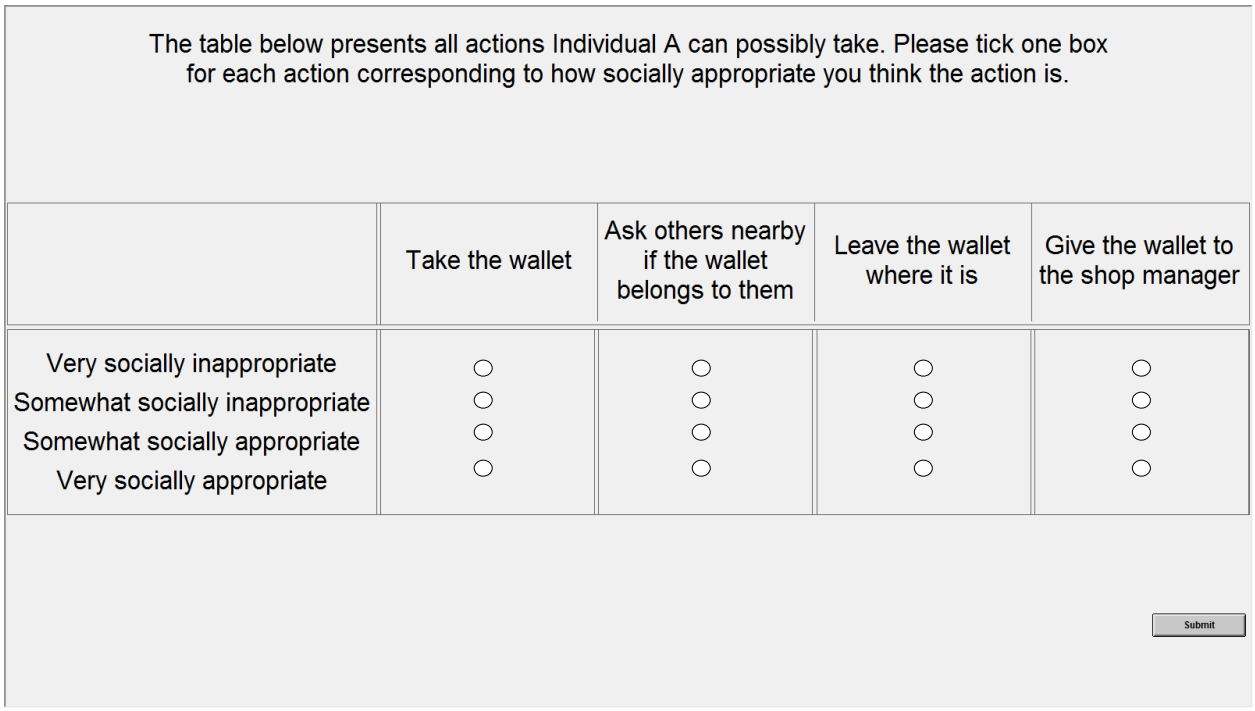
In each of your responses, we would like you to answer as truthfully as possible, based on your opinions of what constitutes socially appropriate or socially inappropriate behaviour.

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.



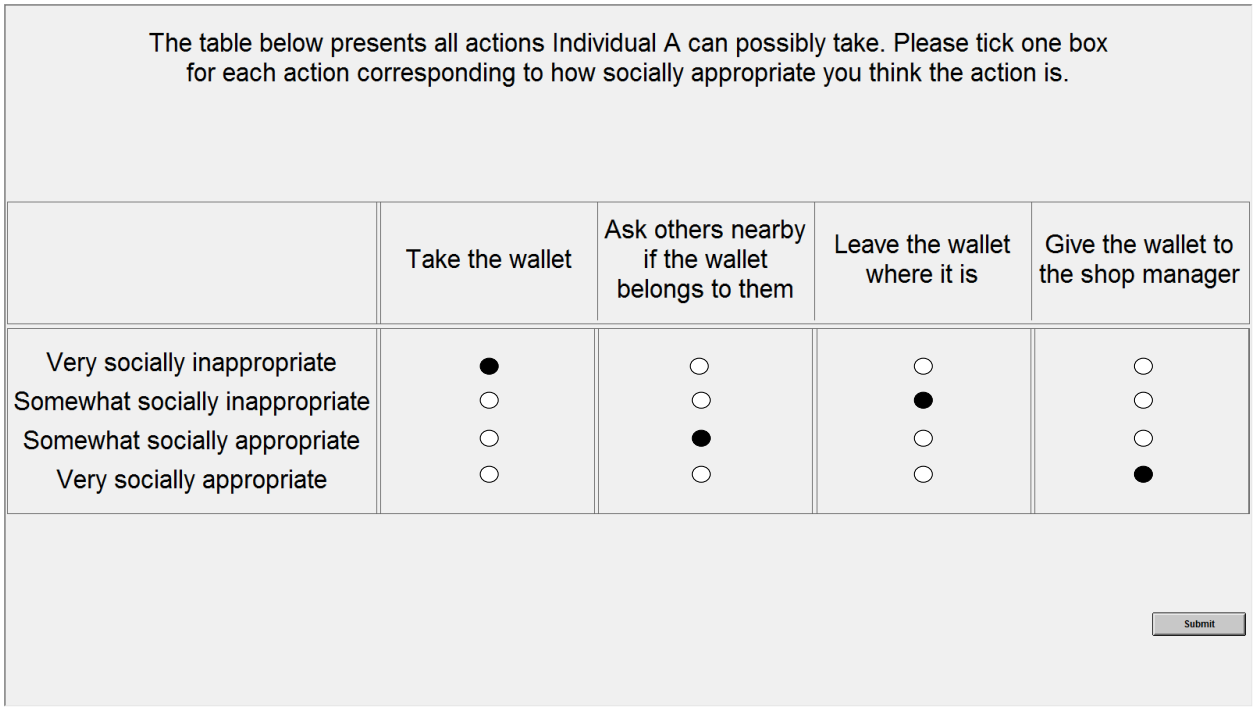
The table below presents all actions Individual A can possibly take. Please tick one box for each action corresponding to how socially appropriate you think the action is.

	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	○	○	○	○

Submit

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:



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

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

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 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.

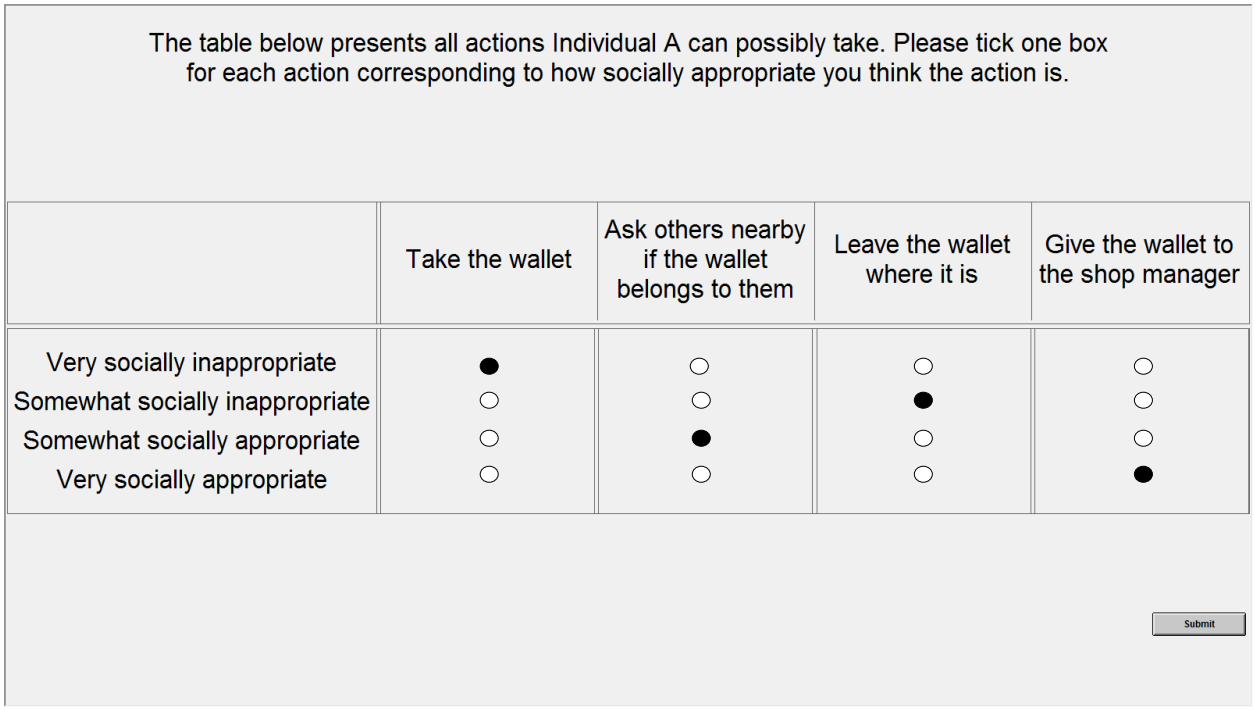
The table below presents all actions Individual A can possibly take. Please tick one box for each action corresponding to how socially appropriate you think the action is.

	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	○	○	○	○

Submit

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:



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.

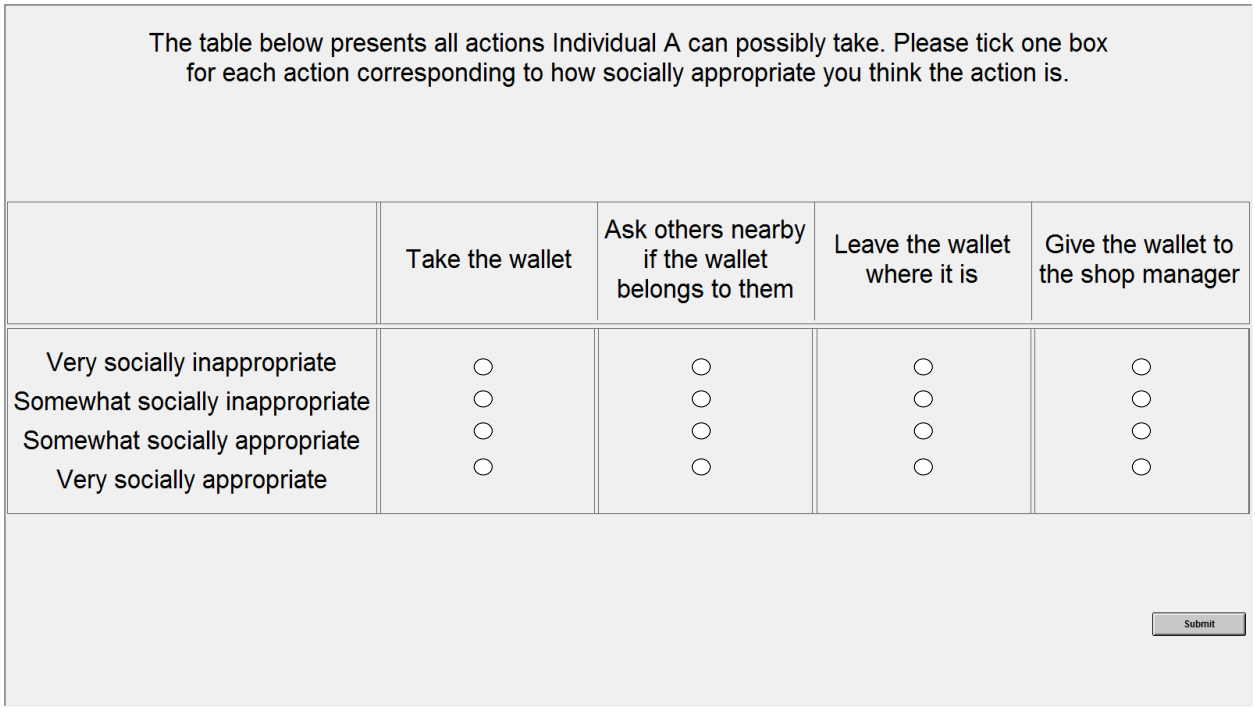
In each of your responses, we would like you to answer as truthfully as possible, based on your opinions of what constitutes socially appropriate or socially inappropriate behaviour.

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.

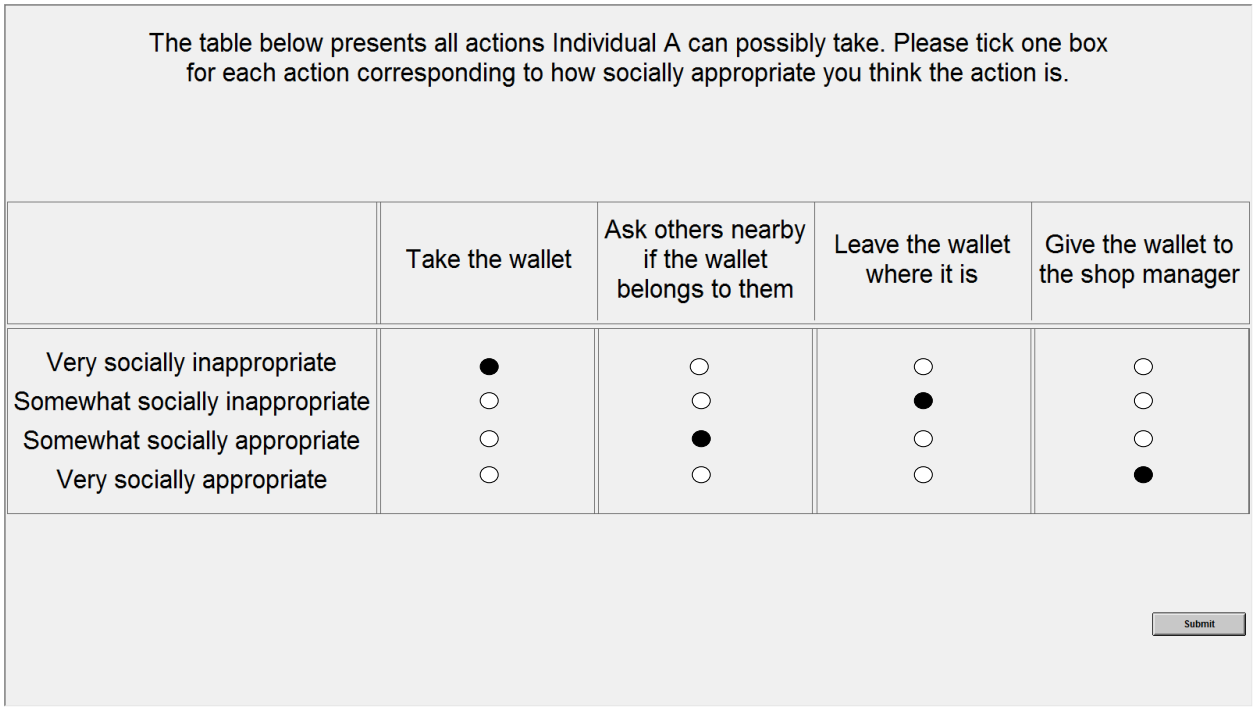
The table below presents the list of the possible actions Individual A can choose. For each of the actions, you would be asked to indicate whether you believe choosing that action is very socially inappropriate, somewhat socially inappropriate, somewhat socially appropriate, or very socially appropriate.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



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:



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.

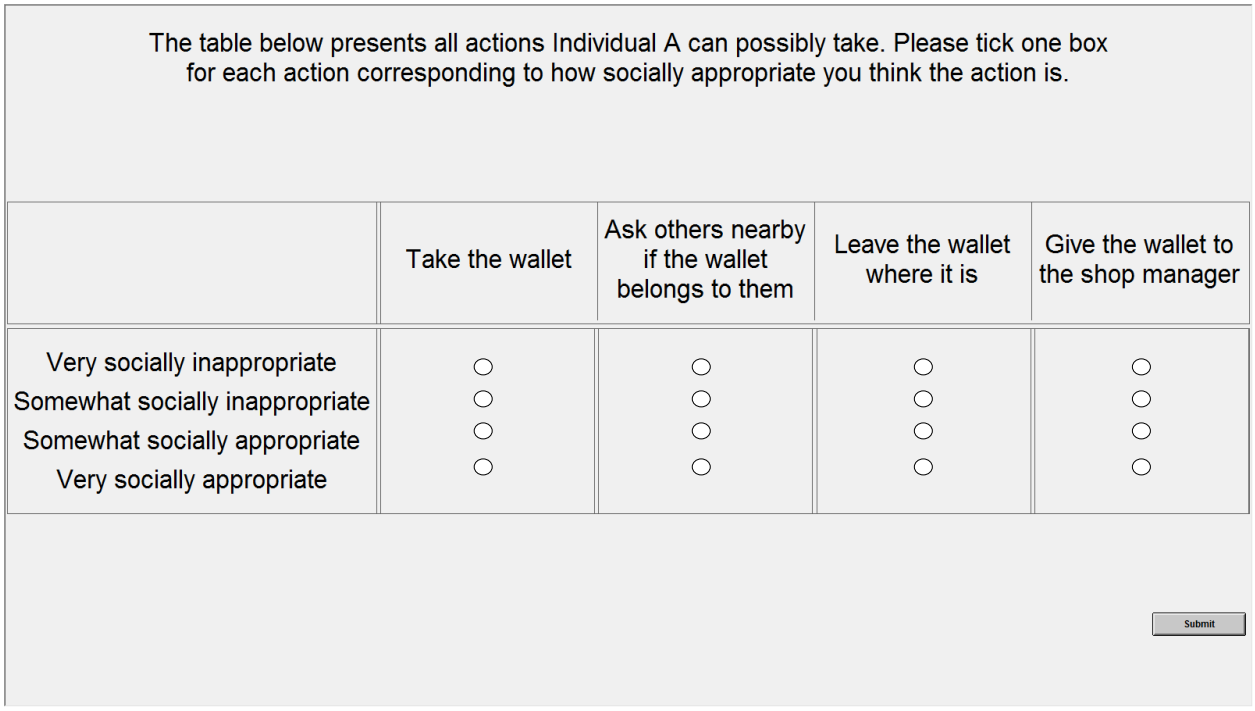
In each of your responses, we would like you to answer as truthfully as possible, based on your opinions of what constitutes socially appropriate or socially inappropriate behaviour.

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.

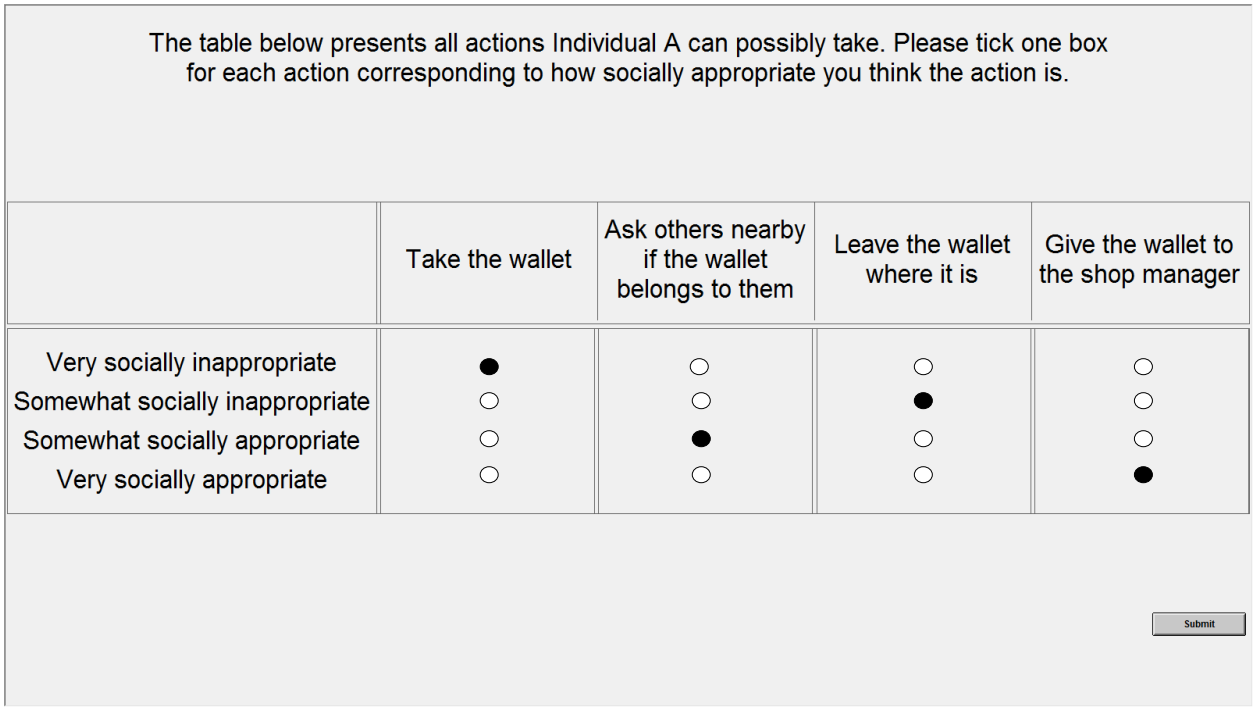
The table below presents the list of the possible actions Individual A can choose. For each of the actions, you would be asked to indicate whether you believe choosing that action is very socially inappropriate, somewhat socially inappropriate, somewhat socially appropriate, or very socially appropriate.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



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:



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

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

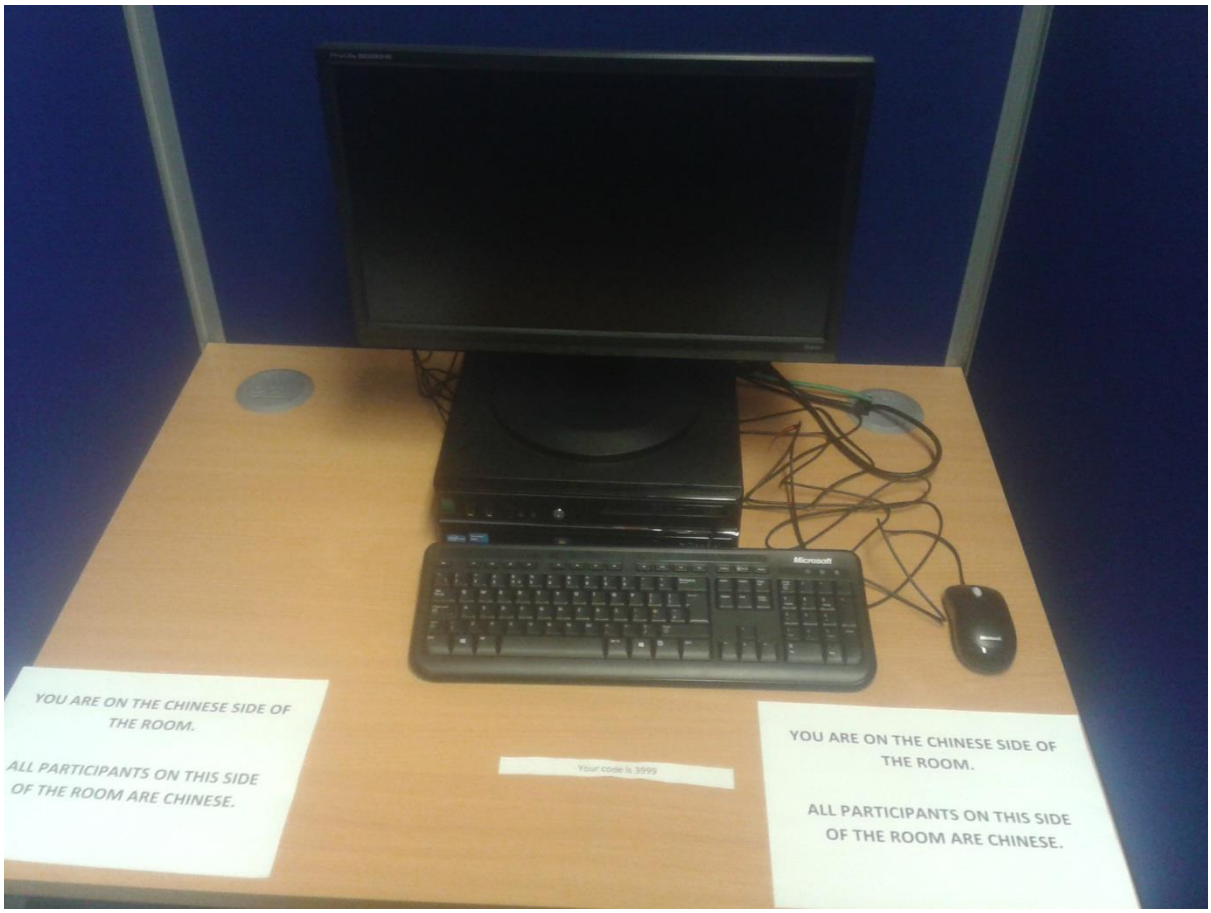
In this experiment, one third of you will be randomly assigned by the computer into a role entitled 'Individual A'. The decisions made by Individual As during the task will determine the payments from the task received by the other two thirds of participants. Each of you has an equal chance of being 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



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.

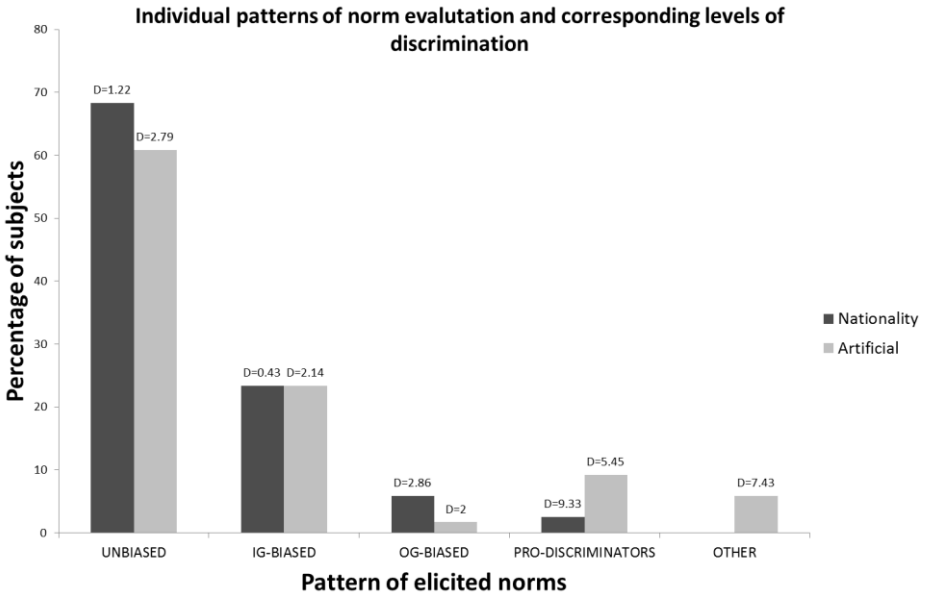
D: Individual patterns of behaviour

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



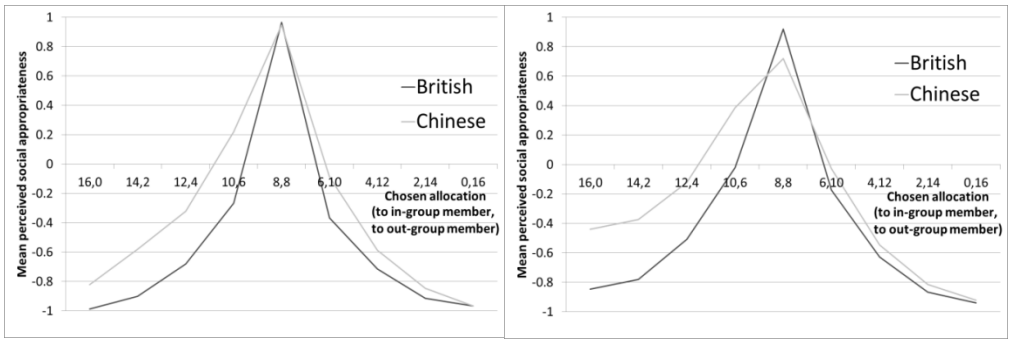
Notes: Figure D1 shows the percentage of participants in each treatment whose evaluations follow each pattern. Above each bar, D=the average level of discrimination against out-group members by participants of the given type in the given treatment – e.g. for UNBIASED participants in the Nationality *treatment,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

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.*

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.

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.*