On the social inappropriateness of discrimination

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

[a] *School of Economics, University of Nottingham, United Kingdom of Great Britain and Northern Ireland*
[b] *School of Economics, University of Nottingham Ningbo China, 199 Taikang East Road, Ningbo 315100, China*

ARTICLE INFO

ABSTRACT

We experimentally investigate the relationship between discriminatory behaviour and the perceived social inappropriateness of discrimination. We conjecture that discrimination will be weaker when social norms oppose it. Our results support this prediction. Using a Krupka-Weber social norm elicitation task, we find participants perceive it to be more socially inappropriate to discriminate on the basis of nationality than on the basis of social identities artificially induced using a trivial minimal group technique. 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.

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). The focus of our paper is on this second form of discrimination, *taste-based discrimination*, and in particular on the psychological foundations of the tastes or preferences for discrimination, which have received remarkably little attention in the literature.

Specifically, in this paper we use experimental methods to investigate whether tastes for discrimination are systematically associated with *social*

norms, i.e. collectively recognised rules of behaviour that define which actions are viewed as socially appropriate within a specific social group.[1] As we discuss further below, there may be a host of factors that shape the tastes for discrimination, including direct altruism towards members of one's own social group. The key contribution of our paper is to provide evidence that one important taste-shaping factor is a norm-based mechanism that regulates the extent to which actions that favour one's own group relative to others are regarded as permissible and appropriate. Uncovering this normative component is an important step towards understanding how patterns of taste-based discrimination are shaped.

If social norms moderate the taste for discrimination, the incidence of discriminatory behaviour should positively correlate with beliefs about the appropriateness of discrimination. Similar correlations have been found in relation to other types of 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; Gächter et al., 2013 –

---

[1] See Elster (1989) and Ostrom (2000) for definitions of social norms. See Akerlof and Kranton (2000, 2005) for a discussion of the importance of norms for discriminatory behaviour.

* Corresponding author.
E-mail address: Tom.Lane@nottingham.edu.cn (T. Lane).

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 behaviour outside the lab. Thus, in driving behaviour, social norms may effectively substitute for laws (e.g. Huang and Wu, 1994), or may complement them (e.g. Sunstein, 1990; Kübler, 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 document empirically using naturally occurring data, not least of all because of the difficulties associated with accurately measuring such beliefs.[2]

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 responses to this question as a proxy for the perceived social appropriateness of discrimination. Using this interpretation, we calculated the percentage of residents in each local authority area of Scotland who agreed with the statement. For each area, Fig. 1 plots this variable against the number of racist incidents,[3] per 100 non-white residents,[4] reported to the police in the financial year 2003–4 (Scottish Executive Statistical Bulletin, 2007). A correlation coefficient 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 notion that norms moderate the taste of discrimination.

The acceptability of prejudice-based humour has sometimes been used as a proxy for the normative appropriateness of discrimination (see, e.g., Crandall et al., 2002). Fig. 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[5]), 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. Fig. 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 (United States Department of Justice, 2015).[6] Both the frequency of anti-black joke searches and the rate of anti-black hate crime incidents declined considerably over the period. This is suggestive of a positive relationship in the US between the change over time in the social appropriateness of discrimination against black people and the change over time in discriminatory behaviour against black people.

In spite of these examples, the paucity of useful naturally occurring data with which to investigate the empirical relevance of norms for discriminatory behaviour 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 a process of social identification 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 ('in-group'), the other an individual not sharing their primed identity ('out-group'). 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. We do this across two treatments that we designed to vary the perceived appropriateness of discriminating in favour of the in-group and against the out-group, while holding other aspects of the decision-making context constant.[7] 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 norms 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. Indeed, when identity groups are artificially formed, participants have no directly relevant social norm to which to refer for guidance about the social appropriateness of discrimination. 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 allocator game and are asked to evaluate the social appropriateness of each and every possible action available to the allocator. We use this norm-elicitation task to construct an incentivized measure of the extent to which participants' perceptions of the appropriateness of discrimination vary across our two treatments and to examine the extent to which these differences in perceived appropriateness translate into differences in discriminatory behaviour in the allocation task.

Our results show that, in both treatments, discriminatory actions are viewed as socially inappropriate. However, as expected, discrimination is perceived to be significantly less appropriate in the nationality treatment compared to the artificial identity treatment. The results of the decision task correlate with 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 notion that the perceived social appropriateness of discrimination varies according to the way identity groups are defined, and this corresponds with 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

---

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

[3] 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)

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

[5] 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.

[6] 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.

[7] To illustrate the idea that discrimination may be perceived as more appropriate along certain dimensions of identity than others, consider sports or music fandom versus ethnicity or gender. Norms may render it appropriate to discriminate against others who support a different football team or listen to a different type of music, but not appropriate to discriminate against others who are different in terms of ethnicity or gender.
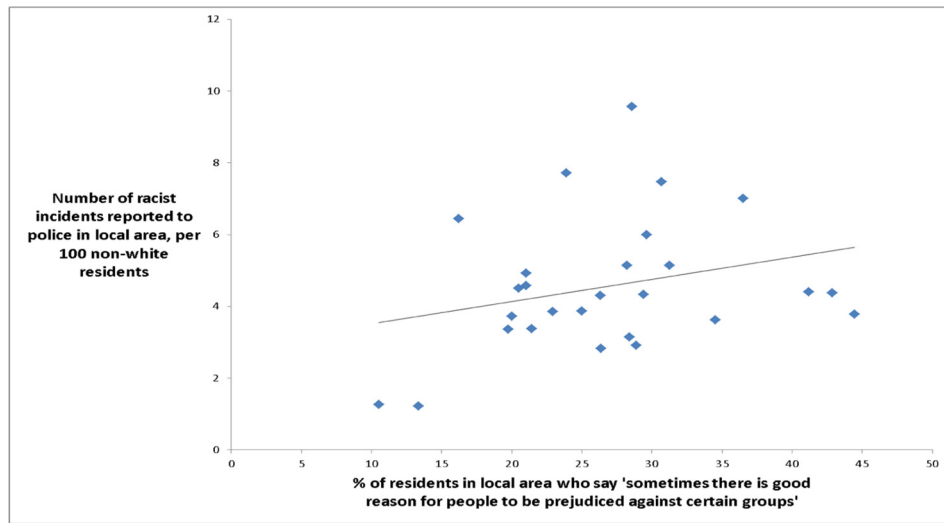
Fig. 1. Variations in attitudes towards racial prejudice and race crimes across Scottish local authority areas. Note: Fig. 1 plots, at the level of local authority area, the relationship between attitudes to prejudice, as reported in the Scottish Social Attitudes Survey 2002, and the frequency of racist incidents reported to the police in the financial year 2003–4. Each data-point represents one local authority area in Scotland.

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. (2017), who 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.'s, our experiment 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 identities of the decision-makers alone, we focus on the social identities of *both* the decision-makers *and* other individuals affected by the decision-makers' 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.[8] 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 social norms, but also that these norms vary across particular dimensions of an individual's identity.

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

2. Theoretical framework

Our simple model of social norm-compliance and identity closely follows Krupka and Weber (2013) and Chang et al. (2017), who based theirs on Akerlof and Kranton (2000, 2005). An individual i's utility U_i depends on the payoff-determining actions of him- or herself and others, $a = (a_i, a_{-i})$, the social identities of him- or herself and others, $I = (I_i, I_{-i})$, and the dimension of identity that is salient in the decision situation, d:

$$U_i(a, I, d) = V_i(a) + S_i(a|I) + \gamma_i N(a_i|I, d) \tag{1}$$

We assume that the decision-maker's utility can be broken into three components. The first component, $V_i(a)$, describes individual i's utility over material payoffs, which in turn depends 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

---

[8] 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.
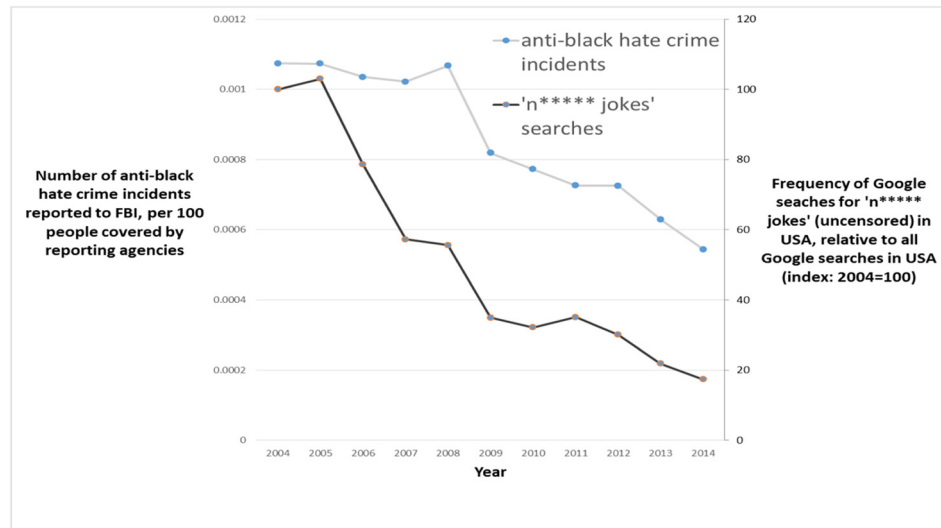
Fig. 2. Google searches for racist jokes about black people and anti-black hate crimes in the US, 2004–14. Note: The light grey line plots the number of anti-black hate crime incidents reported to the FBI each year, adjusted for the population size covered by reporting agencies at the time. The dark 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.

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). Importantly, this component of utility does not depend on the identities of the decision-maker or the others.

In contrast, we assume that the second and third components of utility depend on social identities. These capture the decision-maker's willingness to treat others differently depending on how those others' identities compare to his or her own identity. There are several psychological mechanisms that form the basis for these components. Social identity theory (Tajfel and Turner, 1979), for example, posits that discrimination helps individuals satisfy their *need for positive self-esteem* since it confers a relatively high status on the in-group at the expense of the out-group. Subjective uncertainty reduction theory (Hogg, 2000) takes a different approach: individuals strive to reduce *uncertainty about their attitudes, beliefs, and perceptions*. Self-categorization and identification with groups that provide normative prescriptions for behaviour can reduce this uncertainty and lead to differential treatment of in-groups and out-groups.

We view these psychological mechanisms as distal motivations for the 'taste for discrimination' that has been discussed in the economics literature. In our model we operationalise these mechanisms using two distinct components of utility. The component $S_i(a|I)$ captures what has traditionally been thought of as the taste for discrimination, i.e. utility derived by individual *i* from *i*'s and others' material payoffs that is conditional on how the social identities of individual *i* and the others relate, one to another. In our model, this component of utility can be thought of as primal – as the direct utility that is or would be derived from favouring the in-group in the absence of any self-moderation. As in the utility functions proposed by McLeish and Oxoby (2007), Chen and Li (2009), and Chen and Chen (2011), this component of utility is not conditional on which specific dimension of identity is salient in the decision-making environment. Rather, we simply assume that *i* places a higher weight on the material payoffs of those players who are in-group, i.e., have the same social identity as him- or herself, as compared to the payoffs of players who are out-group, i.e., have a different identity.[9] This can accommodate simple forms of favouritism towards the in-group, such as in-group altruism, as well as more complex forms of identity-contingent other-regarding

preferences, as in the models by Chen and Li (2009) and Chen and Chen (2011).

The third component of utility in our model, $\gamma_i N(a_i|I,d)$, captures the decision-maker's preference to self-moderate his or her primal inclination to favour the in-group with reference to what is or is not socially appropriate. Specifically, we assume that the individual derives utility from complying with normative prescriptions, captured by the function $N(.)$, which defines the social appropriateness of each action a_i available to individual *i*. These normative prescriptions depend on social identities. They may, for example, prescribe different behaviours towards in-group and out-group others. In addition, these prescriptions depend on the *dimension* of social identity, *d*, that is salient given the decision-making context. So, the same action may be viewed as more or less socially appropriate depending not only on how the identities of the decision-maker and others compare, but also on what dimension of identity it is appropriate or meaningful to compare given the context. In some contexts, this third term mitigates the second. For example, an individual might have a primal desire to direct mildly insulting comments at members of an ethnic group other than his or her own, but refrains from doing so because it is socially inappropriate. In other contexts, this third term may build on the second. For example, an individual might have a primal desire to direct mildly insulting comments at the supporters of a soccer team other than the one he or she supports and is further motivated to do so because, especially on match days, such behaviour is socially appropriate. 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), where they have to divide an amount of money between two other participants. 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*.

Moreover, in all treatments subjects are asked to divide the money between a participant who belongs to the same identity group as themselves and a participant who belongs to a different identity group. Thus, the second component $S_i(a|I)$ of the utility function is also kept constant across treatments.

Our treatments vary the dimension of identity *d* 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 identity groups are formed on

---

[9] During our analysis in Section 4, we investigate whether this assumption should be relaxed, i.e., whether utility from in-group favouritism depends on the dimension of identity that is salient in the decision situation.

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, $N(a_i|I,d)$, that regulate the third 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. Our experiment empirically explores the effect of varying the salient dimension of identity on the normative prescriptions relating to discriminatory behaviour and the role of these normative prescriptions in predicting such behaviour.

Note that our model does not specify ex-ante the underlying determinants of the perceptions of appropriate behaviour captured by the function $N(.)$, or how these will vary across treatments. Instead, we follow Krupka and Weber (2013) and Chang et al. (2017) and employ a norm-elicitation technique to quantify, in an incentive-compatible way, the function $N(.)$ in each treatment.[10] This allows us to assess empirically 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 predicted by differences in the perception of its appropriateness.

3. Experimental design

3.1. 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 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 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 maximize our chances of discerning treatment differences, we wanted a task that reliably

produces discriminatory behaviour and, 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 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.

3.2. 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 an allocator could take, and were asked to evaluate the social appropriateness of each by selecting one option on a four-point scale: 'Very socially inappropriate', 'Somewhat socially inappropriate', 'Somewhat socially appropriate' or 'Very socially appropriate.'

To ensure that the relevant perceptions of appropriateness are measured, the evaluators should be, to the greatest extent possible, in the mind-set of the person making the decision they are evaluating. In our experiment, 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.

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. These 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 recognised rules of behaviour, rather than personal opinions about behaviours (e.g. Elster, 1989; Ostrom, 2000).

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

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

---

[10] This is one of the main advantages of the social norms approach proposed by Krupka and Weber (2013) since the researcher does not have to rely on introspection or casual empiricism to specify ex-ante the underlying normative structure of the decision situation he or she is interested in studying, but can rather let this be revealed directly by the data.

[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.

[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.

[14] See also Erkut et al. (2015) and D'Adda et al. (2016) for further evidence on the validity of the Krupka and Weber elicitation task in within-subject experimental designs.

3.3. 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.

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 conjectured that the normative prescriptions regulating the third component of utility in Eq. (1) would differ across the two treatments. Specifically, we conjectured that favouring the in-group at the expense of the out-group would be viewed as *less* appropriate in the *Nationality* compared to the *Artificial* treatment. This specific conjecture was derived from the following assumptions and observations. First, we considered the *Nationality* treatment. In Britain, a liberal society with a long history of in-migration, it seemed reasonable to assume: first, the existence of a norm proscribing discrimination against people from nations other than one's own; and second, that both the British and the Chinese participants in our experiment recognised such a norm and its relevance under the *Nationality* treatment. Under these assumptions, the third component of utility in Eq. (1) is discrimination prohibiting.

Second, we considered the *Artificial* treatment. In this case, there was no directly relevant social norm to which participants could refer.

However, we conjectured that participants could have referred to one or more apparently partially relevant social prescriptions or norms.[17] So for some, the specifics of the *Artificial* treatment could have brought to mind dimensions of identity such as nationality or ethnicity. However, it seemed reasonable to assume that for others it would have been more likely to invoke dimensions of identity such as sports fandom and team game-playing, across which discrimination is condoned. Thus, we conjectured that, under *Artificial*, to the extent that any normative prescription or prescriptions were regulating the third component of utility in Eq. (1), on average, they would be *less* discrimination prohibiting than those at work under *Nationality*.

Finally, we noted that this conjecture is consistent with the fact that previous 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.[18] Indeed, 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.

3.4. 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.[19]

4. Results

4.1. Treatment differences – social norms

We look first at the social appropriateness of discrimination in each treatment, as measured by the norm-elicitation task. Fig. 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,

---

[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.

[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] Hertel and Kerr (2001) present evidence that supports the notion that participants in a minimal group paradigm may refer to multiple, partially relevant normative prescriptions, by showing that discrimination can be manipulated by priming, alternatively, principles of loyalty or equality.

[18] Interestingly, however, a number of previous studies have failed to find strong discrimination in minimal group experiments conducted in China (see Cadsby et al., 2016, Section 3.1). This raises the possibility that discriminatory behaviour by Chinese participants may be difficult to observe. Our experimental evidence, discussed below, does not support this conjecture.

[19] 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.
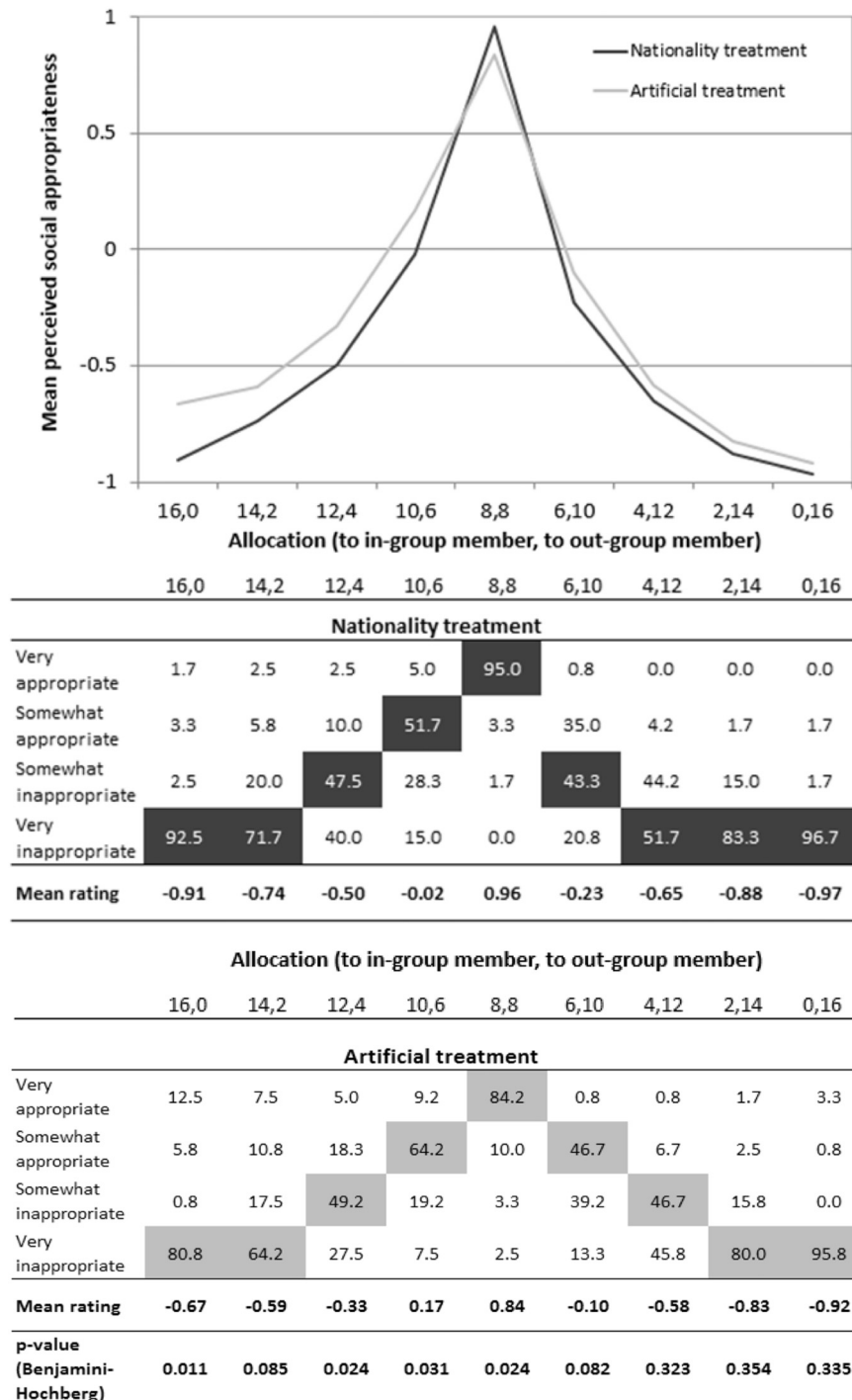
	16,0	14,2	12,4	10,6	8,8	6,10	4,12	2,14	0,16
Nationality treatment									
Very appropriate	1.7	2.5	2.5	5.0	95.0	0.8	0.0	0.0	0.0
Somewhat appropriate	3.3	5.8	10.0	51.7	3.3	35.0	4.2	1.7	1.7
Somewhat inappropriate	2.5	20.0	47.5	28.3	1.7	43.3	44.2	15.0	1.7
Very inappropriate	92.5	71.7	40.0	15.0	0.0	20.8	51.7	83.3	96.7
Mean rating	-0.91	-0.74	-0.50	-0.02	0.96	-0.23	-0.65	-0.88	-0.97

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

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

Fig. 3. Perceived social appropriateness of actions in allocator game. Notes: Fig. 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.

1995), we sort our p-values in ascending rank and multiply each by the number of separate tests being performed (in our case nine, one for each possible allocation) before dividing each by its rank – thus greater adjustments are made to smaller p-values.[20]

---

[20] All p-values reported in this paper are two-sided, 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.

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.[21]

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, Fig. 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.[22]

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 better in three out of the 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; Fig. 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.[23]

4.2. Treatment differences – discrimination

Fig. 4 presents 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 participant 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.[24]

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.[25,26]

4.3. Econometric analysis of individual perceptions of social appropriateness and 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 three components, defined respectively on material payoffs, identity-contingent preferences over material payoffs, and normative prescriptions. We assume that the first component depends on the squared difference between the material payoffs of the two passive players implied by allocation x.

The second component depends on the material payoff of the passive player who shares the same identity as the allocator. The third component depends on the social appropriateness of the allocation. For allocator i,

$$U_i(a_x) = v\left[\pi_j(a_x) - \pi_k(a_x)\right]^2 + s\left[\mathbb{1}_{I_i=I_j}\pi_j(a_x) + \mathbb{1}_{I_i=I_k}\pi_k(a_x)\right] + \gamma N_i(a_x) \quad (2)$$

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; $\mathbb{1}_{I_i=I_j}$ and $\mathbb{1}_{I_i=I_k}$ are indicator functions that take value 1 if the allocator i and player j (or k) belong to the same group and 0 otherwise; and $N_i(a_x)$ is the social appropriateness

---

[21] 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.

[22] In Supplementary Online Materials D we show that these average treatment differences are driven by systematic cross-treatment variations in subjects' response patterns to the norm-elicitation task. In particular, in the *Artificial* treatment more subjects assigned the highest appropriateness rating to the (16,0) allocation and then monotonically decreasing ratings of appropriateness as more money was given to the out-group member. This is consistent with a social norm of in-group favouritism.

[23] Fig. 3 also suggests and further analysis confirms that the perceived appropriateness of discrimination varies more in the *Artificial* compared to the *Nationality* treatment. The variance in appropriateness is significantly (5% level) greater in the *Artificial* compared to the *Nationality* treatment for 5 out of the 9 possible allocations and, for the remaining 4, we cannot reject the null of equal variance (*p*-values adjusted for multiple tests using the Benjamini-Hochberg False Discovery Rate method). This is consistent with our argument that, in the *Artificial* treatment, the participants could have referred to a variety of apparently partially relevant social prescriptions or norms. An interesting avenue for further research would be to explore the impact that this normative uncertainty and heterogeneity may have on behaviour.

[24] An alternative explanation could be that it is participants' behaviour in the allocator game that shapes the appropriateness ratings they supply in the norm-elicitation task. For instance, participants may reflect on the way they behaved – or, for those playing the allocator game second, the way they think they would behave – and assume others would consider this an appropriate way to act. Although we cannot eliminate this possibility, we can investigate whether, if we hold behaviour constant by focusing on those who did not discriminate in the allocator game (76% of the sample), we can still observe a difference in the social appropriateness of discrimination across treatments. This analysis reveals that discrimination is perceived to be more inappropriate in the *Nationality* treatment than the *Artificial* treatment even when behaviour is constant (see Supplementary Online Materials E). The pattern of evaluations is similar to that reported in Figure 3. The differences for all allocations are in the expected direction, although after applying the Benjamini-Hochberg adjustment the treatment difference is significant at the 10% level only for the (16,0) allocation, possibly reflecting the smaller sample size.

[25] Table 1 also reveals that Chinese participants discriminated more than British participants. We also found that, under both *Artificial* and *Nationality* treatments, Chinese participants perceived discrimination to be less socially inappropriate than British participants. These cross-national differences are explored in more detail in Supplementary Online Materials F. One possible reason for the stronger discrimination and more favourable perception of it by the Chinese in our experiment 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).

[26] We ran further models on the British and Chinese sub-samples 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.
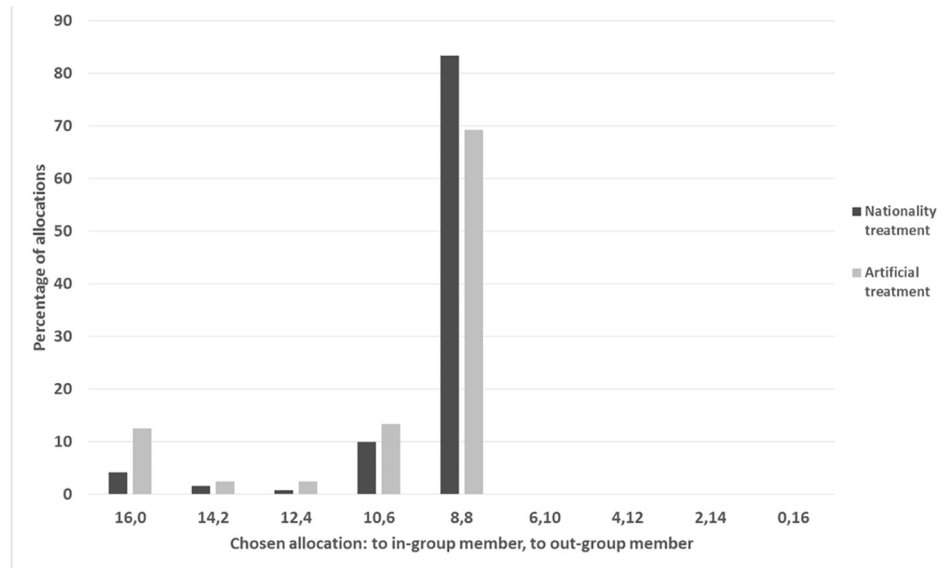
Fig. 4. Discrimination in the allocator game. Notes: Fig. 4 shows the percentage of participants in each treatment who choose each allocation. Allocations are denoted by the amount given to the in-group member on the left, and the amount given to the out-group member on the right – e.g. (16,0) denotes allocating £16 to the in-group member and £0 to the out-group member.

that allocator i ascribes to allocation x, as measured in the norm-elicitation task.[27]

The parameter v captures the weight that the allocator places on the material payoff component of the utility function, regardless of the identities of the passive players: allocations that implement unequal pay-offs 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 parameters s and γ capture the weight that the allocator places on the components of utility that are contingent on the identities of the passive players. The parameter s captures simple in-group altruism: the allocator places an extra weight on the material payoff of the passive player who belongs to the same group as him- or herself (and zero weight on the payoff of the out-group). Finally, the parameter γ captures the weight that allocators place on (identity-related) normative prescriptions.

Following Gächter et al. (2013) and Krupka and Weber (2013), we use fixed-effects conditional logit regressions to estimate the weights v, s and γ on the three components of the utility function shown in Eq. (2). 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:

$$Pr(a = a_x) = \frac{exp\{U(a_x)\}}{\sum_{l=1,\dots,9} exp\{U(a_l)\}}, x = 1, \dots, 9 \qquad (3)$$

Our objective, here, is to show that a norm-augmented model is better able to capture the data patterns observed in the experiment than a model which contains only the first two components of utility captured in Eq. (2) above. Thus, in Table 2 we report the output of two fixed-effects conditional logit models, each estimated using all of the allo-cation decisions and, in Model (B), all of the social appropriateness

[27] 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.

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)
R²	0.032		0.082	
N	240		234	

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.
*** $p < 0.01$.
** $p < 0.05$.
* $p < 0.1$.

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

Dependent variable = 1 if action is chosen; 0 otherwise		
Model	(A)	(B)
v (weight on payoff inequality)	−0.022***	−0.005**
	(0.002)	(0.002)
s (weight on in-group altruism)	0.194***	0.206***
	(0.032)	(0.050)
γ (weight on normative prescriptions)		1.414***
		(0.104)
Pseudo R²	0.277	0.520
Bayesian Information Criterion	778.12	529.04
Number of Observations	2160	2160

Standard errors are in parentheses.
*** $p < 0.01$.
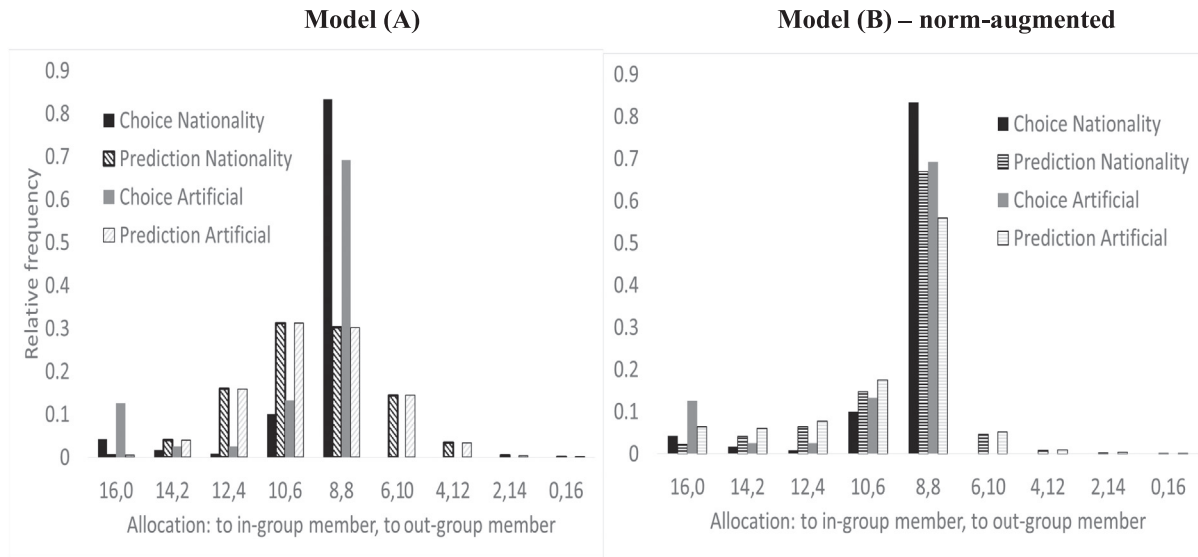** $p < 0.05$.
* $p < 0.1$.

Fig. 5. Actual choice rates in the allocator game and choice rates predicted by conditional logits. Notes: Fig. 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 and in-group altruism, 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.

evaluations generated under either the *Nationality* or the *Artificial* treatment. In the first model we impose the restriction $\gamma = 0$ to the utility function in Eq. (2) and, thus, estimate a choice model where the decision-maker is purely concerned with payoff inequality and simple in-group altruism. In the second model this restriction is removed and utility is allowed to depend on payoff inequality, simple in-group altruism, and the individual's normative evaluation of the action under consideration.

The significant negative estimates of *v* in both models indicate that actions which yield larger payoff inequalities are less likely to be chosen. The estimate of *s* is positive and significant in both models, indicating that allocations which favour the in-group are more likely to be chosen. Finally, the significant positive estimate of γ in model (B) 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* and *s* parameters indicates that the normative component of the utility function can explain variation in choice behaviour that cannot be entirely captured by (identity-blind) inequality considerations combined with simple in-group altruism. This also explains why the Bayesian Information Criterion is significantly lower for model (B) than (A) ($p < 0.001$ 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 Fig. 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 Fig. 4). The left-hand panel of Fig. 5 presents the choice rates predicted by model (A). The right-hand panel presents the choice rates predicted by model (B). 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.

Model (A), in which participants care only about inequality and in-group altruism, captures some important aspects of the choice data. In particular, the model predicts that deviations from equality are asymmetric across the choice space. That is, the probability of choosing an unequal and in-group-favouring allocation is predicted to be higher than that of choosing an allocation which creates the same payoff inequality but favours the out-group. This is what we observe in the

actual choice data, as no-one chooses out-group-favouring allocations, while 24% of participants choose an in-group-favouring allocation.

However, Model (A) fails to capture a second key feature of the choice data, the difference in allocations across treatments. In contrast, the norm-augmented model (B) predicts a lower probability of choosing the equal split allocation and higher probabilities of choosing in-group-favouring allocations in the *Artificial* compared to the *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, they are markedly lower than those predicted by Model (A).

Before concluding, we need to investigate whether identity-contingent preferences over material payoffs (captured by the term $s [\mathbb{1}_{I_i=I_j}\pi_j(a_x) + \mathbb{1}_{I_i=I_k}\pi_k(a_x)]$ in Eq. (2)) depend on the dimension of social identity that is salient in the decision-making situation. We would expect to observe such a dependence if, for example, in-group altruism varies depending on how strongly individuals identify with the group and this, in turn, depends on the dimension of identity along which groups are defined.

To explore this, we consider the following alternative utility specification:

$$U_i(a_x) = v\left[\pi_j(a_x) - \pi_k(a_x)\right]^2 + s\left[\mathbb{1}_{I_i=I_j}\pi_j(a_x) + \mathbb{1}_{I_i=I_k}\pi_k(a_x)\right]$$
$$+ \sigma\mathbb{1}_{d=nationality}\left[\mathbb{1}_{I_i=I_j}\pi_j(a_x) + \mathbb{1}_{I_i=I_k}\pi_k(a_x)\right] + \gamma N_i(a_x) \quad (4)$$

which is identical to Eq. (2) except for the inclusion of the third term, that captures the incremental utility from in-group altruism when identities are defined over *national* groups ($\mathbb{1}_{d=nationality}$ is an indicator variable that takes value 1 when groups are based on national identities). Thus, in Eq. (4) *s* captures in-group altruism towards minimal groups, while $s + \sigma$ captures in-group altruism towards national groups. In Table 3 we use a fixed-effects conditional logit regression to estimate the weights *v*, γ, *s* and σ.

Here, as in Model (B) in Table 2, *v* is highly significant and negative and both γ and *s* are highly significant and positive, while the newly added σ is statistically insignificant.[28] Thus, we cannot reject the null

---

[28] Note also that, the Bayesian Information Criterion is not significantly different between the model in Table 3 and model (B) in Table 2 ($p = 0.195$ on a likelihood-ratio test).

Table 3
Alternative conditional logit regression.

Dependent variable = 1 if action is chosen; 0 otherwise	
v (weight on payoff inequality)	−0.006***
	(0.002)
s (weight on in-group altruism for artificial groups)	0.248***
	(0.061)
σ (extra-weight on in-group altruism for national groups)	−0.088
	(0.068)
γ (weight on normative prescriptions)	1.398***
	(0.104)
Pseudo R^2	0.522
Bayesian Information Criterion	535.03
Number of Observations	2160

Standard errors are in parentheses.
*** $p < 0.01$.
** $p < 0.05$.
* $p < 0.1$.

hypothesis that direct utility derived from in-group altruism is independent of the dimension of social identity that is salient in the decision-making situation, and retain Eq. (2) as our preferred specification of utility.

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 less inappropriate 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 less inappropriate, and that, at the individual level, perceived inappropriateness predicts actual behaviour.

Our findings are supportive of a theoretical framework within which taste-based discrimination is partly driven by normative considerations about the appropriateness of discriminatory behaviour. We offer direct evidence that, across choice contexts that are otherwise identical, differences in the way identity groups are defined translate into differences in the perceived normative prescriptions, and corresponding differences in behaviour towards in-group and out-group members. These findings are in line with models of social identity that have emphasised the role of social norms, such as Akerlof and Kranton (2000, 2005).

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.

We remain agnostic as to exactly why these norms opposing discrimination are more strongly triggered by national identity than by minimal group identity. National groups are different from minimal groups in various ways, so there are multiple possible explanations. We conjecture that a norm proscribing discrimination on the basis of nationality is likely to have developed over time in a liberal society, with a long history of in-migration, such as Britain, perhaps enhanced by sensitivities about the negative historical effects of racism and xenophobia. For these reasons, grouping participants by nationality seems likely to invoke stronger norms against discrimination than if we grouped them by other types of natural identity, such as university affiliation (as implemented, for instance, in Ockenfels and Werner, 2014). In the case of minimal groups, while their members have no directly relevant norms to refer to, we argue that they are relatively likely to invoke types of identity such as sports fandom and team game-playing, across

which discrimination is considered harmless. It is also possible that some participants under minimal group identity perceive that the experimenter intends for them to discriminate, and that this experimental demand leads them to perceive a social norm in favour of discrimination. These are matters for interesting future research.

While further investigation is needed, our findings are consistent with and strongly suggest that shared norms opposing discrimination do, as one would expect, help moderate discrimination. One likely 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 with the recent co-emergence of a 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.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jpubeco.2018.06.004.

References

Abbink, K., Harris, D., 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. J. Econ. Psychol. 28 (3), 324–337.

Akerlof, G.A., Kranton, R.E., 2000. Economics and identity. Q. J. Econ. 715–753.

Akerlof, G.A., Kranton, R.E., 2005. Identity and the economics of organizations. J. Econ. Perspect. 9–32.

Allcott, H., 2011. Social norms and energy conservation. J. Public Econ. 95 (9–10), 1082–1095.

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

Ashburn-Nardo, L., Voils, C.I., Monteith, M.J., 2001. Implicit associations as the seeds of intergroup bias: how easily do they take root? J. Pers. Soc. Psychol. 81 (5), 789–799.

Banerjee, R., 2016. On the interpretation of bribery in a laboratory corruption game: moral frames and social norms. Exp. Econ. 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. J. R. Stat. Soc. Ser. B Methodol. 289–300.

Bernhard, H., Fehr, E., Fischbacher, U., 2006. Group affiliation and altruistic norm enforcement. Am. Econ. Rev. 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. J. Appl. Psychol. 79 (6), 993.

Bolton, G.E., Ockenfels, A., 2000. ERC: a theory of equity, reciprocity, and competition. Am. Econ. Rev. 166–193.

Buonanno, P., Montolio, D., Vanin, P., 2009. Does social capital reduce crime? J. Law Econ. 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. Manag. Sci. 58 (1), 203–217.

Cadsby, C.B., Du, N., Song, F., 2016. In-group favoritism and moral decision-making. J. Econ. Behav. Organ. 128, 59–71.

Chang, D., Chen, R., Krupka, E., 2017. Rhetoric Matters: A Social Identity Explanation for the Anomaly of Framing (Unpublished manuscript).

Chen, R., Chen, Y., 2011. The potential of social identity for equilibrium selection. Am. Econ. Rev. 101 (6), 2562–2589.

Chen, Y., Li, S.X., 2009. Group identity and social preferences. Am. Econ. Rev. 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 Econ. Behav. 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. J. Pers. Soc. Psychol. 82 (3), 359.

D'Adda, G., Drouvelis, M., Nosenzo, D., 2016. Norm elicitation in within-subject designs: testing for order effects. J. Behav. Exp. Econ. 62, 1–7.

Elster, J., 1989. Social norms and economic theory. J. Econ. Perspect. 3 (4), 99–117.

Erkut, H., Nosenzo, D., Sefton, M., 2015. Identifying social norms using coordination games: spectators vs. stakeholders. Econ. Lett. 130, 28–31.

Fehr, E., Schmidt, K.M., 1999. A theory of fairness, competition, and cooperation. Q. J. Econ. 817–868.

Fischbacher, U., 2007. z-Tree: Zurich toolbox for ready-made economic experiments. Exp. Econ. 10 (2), 171–178.

Gächter, S., Nosenzo, D., Sefton, M., 2013. Peer effects in pro-social behavior: social norms or social preferences? J. Eur. Econ. Assoc. 11 (3), 548–573.

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.

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. Am. Econ. Rev. 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 date: 8 June 2016.

Greiner, B., 2015. Subject pool recruitment procedures: organizing experiments with ORSEE. J. Econ. Sci. Assoc. 1–12.

Guillen, P., Ji, D., 2011. Trust, discrimination and acculturation: experimental evidence on Asian international and Australian domestic university students. J. Socio-Econ. 40 (5), 594–608.

Hargreaves Heap, S.P., Zizzo, D.J., 2009. The value of groups. Am. Econ. Rev. 295–323.

Harris, D., Herrmann, B., Kontoleon, A., Newton, J., 2014. Is it a norm to favour your own group? Exp. Econ. 1–31.

Hertel, G., Kerr, N.L., 2001. Priming in group favoritism: the impact of normative scripts in the minimal group paradigm. J. Exp. Soc. Psychol. 37, 316–324.

Hogg, M.A., 2000. Subjective uncertainty reduction through self-categorization: a motivational theory of social identity processes. Eur. Rev. Soc. Psychol. 11 (1), 223–255.

Huang, P.H., Wu, H.M., 1994. More order without more law: a theory of social norms and organizational cultures. J. Law Econ. Org. 10, 390.

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

Krupka, E.L., Weber, R.A., 2013. Identifying social norms using coordination games: why does dictator game sharing vary? J. Eur. Econ. Assoc. 11 (3), 495–524.

Krupka, E.L., Leider, S., Jiang, M., 2016. A meeting of the minds: informal agreements and social norms. Manag. Sci. 63 (6), 1708–1729.

Kübler, D., 2001. On the regulation of social norms. J. Law Econ. Org. 17 (2), 449–476.

Lane, T., 2016. Discrimination in the laboratory: a meta-analysis of economics experiments. Eur. Econ. Rev. 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. J. Law Econ. Org. 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.

McLeish, K.N., Oxoby, R.J., 2007. Identity, Cooperation, and Punishment.

Moir, R., 1998. A Monte Carlo analysis of the Fisher randomization technique: reviving randomization for experimental economists. Exp. Econ. 1 (1), 87–100.

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 Dev. 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.

Ockenfels, A., Werner, P., 2014. Beliefs and ingroup favouritism. J. Econ. Behav. Organ. 108, 453–462.

Ostrom, E., 2000. Collective action and the evolution of social norms. J. Econ. Perspect. 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., Turner, J.C., 1979. An integrative theory of intergroup conflict. In: Austin, W.G., Worchel, S. (Eds.), The Social Psychology of Intergroup Relations. Brooks/Cole, Monterey, CA, pp. 33–47.

Tajfel, H., Billig, M.G., Bundy, R.P., Flament, C., 1971. Social categorization and intergroup behavior. Eur. J. Soc. Psychol. 1 (2), 149–178.

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

United States Department of Justice, Federal Bureau of Investigation, 2015. Hate Crime Statistics, 2004–2014. Retrieved 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. J. Econ. Psychol. 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. J. Exp. Soc. Psychol. 43 (6), 867–876.