The normative permissiveness of political partyism

Tom Lane [a], Luis Miller [b,*], Isabel Rodriguez [b,c]

[a] Newcastle University Business School, 5 Barrack Road, Newcastle upon Tyne NE1 4SE, United Kingdom
[b] Spanish National Research Council (IPP-CSIC), C/ Albasanz 26, 28037 Madrid, Spain
[c] International Doctorate School of the Spanish National Distance Education University (EIDUNED)

A R T I C L E   I N F O

A B S T R A C T

Political party identity has become one of the strongest social divides within many Western societies. This paper employs experiments to measure discrimination along different dimensions of social identity, and replicates previous findings showing the strongest discrimination against out-groups occurs in the political party domain. Moreover, we explore a possible explanation for this phenomenon based on social norms. We measure the social appropriateness of discrimination along each identity dimension. The ranking of dimensions by discrimination against out-groups reflects the extent to which such behaviour is normatively permissible, with the weakest anti-discrimination norms on the political party dimension. Results are qualitatively similar in two European countries. We argue that, while strong norms sanctioning discrimination on other dimensions have developed historically, no such process has taken place concerning party affiliation, bringing partisan identity to the fore and helping polarisation flourish.

1. Introduction

Why has political party identity become one of the most important social divides in the Twenty-First Century? A fast-growing literature in economics and political science has established the pre-eminence of partyism over other relevant social divides, like race, ethnicity, religion, or national identity. Partyism – also labelled party affective polarization (Iyengar et al., 2019) – refers to any form of prejudice or hostility towards voters or supporters of opposing political parties (Sunstein, 2016). Observational (mostly survey) and experimental studies on partyism in the United States and many European Countries abound (Iyengar et al., 2019; Iyengar and Westwood, 2015; Westwood et al., 2018). Such research shows people have stronger preferences to engage in economic discrimination against party political out-groups than against out-groups along any other dimension of social identity.

Yet despite this proliferation of empirical studies, the underlying theoretical mechanisms explaining the prominence of partyism in contemporary Western societies remain unclear. Most studies resort to the human tendency for tribalism, an evolved predisposition to group conflict that makes us favour and be loyal to *our* groups and hostile to *other* groups with whom we compete (Clark et al., 2019). Unfortunately, while tribalism can explain partyism, on its own it would seem to equally predict such strong in-group favouritism and out-group hostility based on other social divides. To explain the dominance of party affective polarization, we need something else beyond our natural predisposition toward being tribal. Here, we argue that one of the missing pieces in the puzzle of the currently exacerbated level of party affective polarization is the role of social norms.

Social norms are the unwritten 'grammar of society' (Bicchieri, 2005), the informal rules which dictate which behaviours should or

---

should not be taken.[1] In recent years, there has been an explosion of interest in economics in social norms. A vast body of empirical evidence suggests that economic behaviour – across a wide range of domains – is guided and constrained by social norms (Allcott, 2011; Fehr and Schurtenberger, 2018; Gachter et al., 2013; Kimbrough and Vostroknutov, 2016; Köbis et al., 2022; Krupka and Weber, 2013; Krupka et al., 2017; Szekely et al., 2021). More specifically related to our paper's research question, there is evidence that the strength of economic *discrimination* is driven by the extent to which norms fail to prevent such behaviour (Barr et al., 2018; Coffman et al., 2021; Restrepo-Plaza and Fatas, 2022).

We argue that, historically, norms have developed which characterize as morally reprehensible discrimination against out-groups along many dimensions of social identity, especially race and gender (Gaertner and Dovidio, 2000). Anecdotal evidence for this comes from the social sanctions imposed upon public figures caught expressing racist or sexist remarks which might have gone unpunished in previous decades. These norms act as a constraint on discriminatory behaviour. Our conjecture, however, is that Western societies have not evolved corresponding social pressures or sanctions that mute disapproval of partisan opponents. In today's world, it is not uncommon to find left-wing opponents of the British government denouncing its supporters on Twitter as 'scum' or boasting that they have 'never kissed a Tory' (Cohen, 2016). While such actions would seem likely to trigger backlash if taken against out-groups on other domains, we argue that lax norms along the political party identity domain fail to restrict expressions of animus and discriminatory behaviour toward opponents. Based on the previous literature, we cannot claim that there are norms condoning discrimination against partisan out-groups. We rather focus on the weakness of norms regulating partyism as one plausible explanation of its development.

In this article, we provide empirical evidence to support our argument. This comes from a set of laboratory experiments designed to study the role of social norms in partisan and other types of social intergroup interactions. We focus specifically on interpersonal discrimination, which can be regarded as one manifestation of intergroup hostility and, in the political party domain, one of the "dark consequences" of polarization (Finkel et al., 2020). Previous studies have reported discrimination based on partisan and other types of political affiliation in the labour market, in college admissions, and in everyday economic interactions (Dimant, 2023; Iyengar et al., 2019; Michelitch, 2015). In our experiments, we employed a standard norm-elicitation task to establish the social appropriateness of economic discrimination along different dimensions of social identity, as well as an incentivized task to measure such discrimination on each dimension.

We hypothesize that, like in previous research, subjects will discriminate less on the basis of other social identities, such as religion, than on the basis of political party support. Moreover, our conjecture is that participants will perceive it to be correspondingly less socially inappropriate to discriminate in the partisan identity context. In our experiments run in both the UK and Spain, we divided participants into groups on the basis either of their support for political parties or their religious affiliation.[2] The dimension of identity upon which we divided participants represents our treatment manipulation. Having formed these groups, we implemented a standard distribution game in which participants allocated money between in-group and out-group individuals, therein facilitating the measurement of discrimination in favour of the in-group over out-groups. In each treatment, for every possible level of discrimination a participant could enact in the allocation task, we have a corresponding estimate from the norm-elicitation task of the perceived social appropriateness of this behaviour.

The results are in line with our hypotheses. We observe significantly stronger levels of discrimination based on partisan identity than on religious identity. We also find significant differences in norms between treatments: discrimination in favour of one's in-group is perceived to be significantly less inappropriate when the groups are based on political parties than based on religion. The relatively permissive norms regarding partisan discrimination emerge as a compelling explanation for the relative prominence of partyism over other social divisions that has been identified in the existing literature.

Our findings are drawn from experiments following a design similar to that introduced in a previous paper, Barr et al. (2018). That study measured discrimination between groups – and the social norms pertaining to such behaviour – based on either nationality or minimal-group identity. Like the current paper, Barr et al. found strongest discrimination on the identity dimension for which the social norms against discrimination were weakest. The purpose of their study was to identify social appropriateness as a potential moderator of discrimination in general, as well as to explain a surprising result in the previous literature: that discrimination in lab experiments tends to be stronger between minimal groups than between groups based on various types of natural identity (Lane, 2016).

The main purpose of our study differs from Barr et al. in that we are focused primarily on explaining the strength of discrimination along one particular dimension: political party identity. That partisan discrimination and its normative permissiveness outrank the equivalent phenomena on the religious identity dimension – which represents a strong natural characteristic – provides evidence for our argument that political party support is special among group identity dimensions. Note that our experiment does *not* show that social norms *favour* discrimination against partisan opponents; the norm-elicitation task reveals that, along all dimensions of identity, the most socially appropriate behaviour is to provide equal treatment to in-groups and out-groups. However, there is lesser agreement about this on the partisan dimension, and greater acceptance towards those who deviate from equality. Our results suggest that those who discriminate against party political opponents may escape with only mild social sanctions, resulting in discriminatory actions taking on a more attractive cost-benefit profile than they would in other social arenas. Our findings are qualitatively similar in the two

---

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

[2] We conducted a third treatment where participants were divided on the basis of a trivial characteristic (colour of the ball the participants randomly picked during the experiment) representing a 'minimal group' identity (Tajfel, 1970). Since the recruitment procedure for this treatment was not perfectly comparable to the other treatments, results comparing this treatment with the others are reported mostly in the Online Appendix and only briefly in the main text.

countries where the experiments were run, lending credence to their generalisability across different party political contexts.

Of course, social norms do not account for party polarization as a whole. Comparative studies have identified other variables that correlate with country-level polarisation. For instance, economic inequality (e.g., as in the US), high levels of unemployment (e.g., as in Southern and Eastern Europe), and majoritarian political systems (e.g., as in the UK and the US) correlate with affective party polarization at the macro level (Gidron et al., 2020). Our norm-based explanation adds to a growing cross-disciplinary literature on the determinants of partisan polarisation.

2. Methods

2.1. Research Design

Our objective is to explore whether differences in the strength of discrimination along different identity dimensions can be explained by equivalent differences in the strength of social norms prohibiting such discrimination. Our design mostly follows Barr et al., yet departs not only by our treatment manipulation employing three rather than two dimensions of social identity, but also in that, along the identity dimension made salient in a given treatment, subjects are divided into four (rather than two) identity groups. For instance, in the treatment of our UK experiment where we make the partisan identity dimension salient, subjects are divided into groups of supporters of the Conservative, Labour, Liberal Democrat and Green parties. In each treatment, for members of each of the four groups, we estimate three measurements of discrimination in favour of their in-group, separately against each of the three out-groups they face on this identity dimension; correspondingly, we also separately elicit the social appropriateness they perceive of discrimination in favour of their in-group over each out-group. The separation of subjects into four groups more naturally reflected the real-world structure of the social identity dimensions employed in our experiment. Note, however, that we are not primarily interested in drawing comparisons between the behaviour of different groups within an identity dimension, but rather in making aggregate-level comparisons across dimensions, to which end the division of subjects into groups was an essential design feature.

Our UK experiment was run at the University of Nottingham from December 2022 to January 2023, and our Spain experiment at the University Jaume I in Castellón in April 2023. Both were pre-registered.[3] Below, we outline the design of the first experiment in the UK, and in the process highlight any deviations from this design in the implementation of the second experiment in Spain. Note that our study is not an attempt to conduct a controlled cross-cultural comparison between the UK and Spain. In particular, the religious and partisan identity dimensions are differently composed in each country with their own groups and local characteristics. Rather, we are interested in the results across treatment manipulation in each country in their own right, as well as in identifying whether they allow us to draw qualitatively similar conclusions.

2.2. Discrimination measurement task

To measure discrimination, we used a simple third-party allocator game. See Barr et al. (2018) for a discussion of the advantages of this task in our research context. In the game, a decision-maker is endowed with £16 (€16 in Spain) and required to divide it between two passive players, one belonging to the decision-maker's in-group and the other belonging to an out-group.[4] The money can be split however the decision-maker prefers, as long as the amount given to each player is a multiple of two. The interaction is anonymous, and the decision-maker receives no information about either of the passive players except for their group identity.

In order to maximize sample sizes, we employed two elements of randomization. First, while only one third of subjects would be allocators and the other two thirds would be passive players, all subjects were required to make decisions in the role of allocator. It was made clear that, after the end of the experiment, it would be randomly determined which subjects had been assigned to the allocator role, and the decisions of those not assigned to it would be discarded. Secondly, while subjects were told that – if assigned to the role of allocator – they would definitely be matched with one in-group and one out-group player, they were not informed the specific group identity of the out-group player. This would be randomly determined after the allocator role assignments. In the meantime, subjects were required to commit to three allocation decisions, one for each of the possible groups the out-group player might belong to; when it was later determined which out-group an allocator had been matched with, the allocation they had committed to make if matched with a member of this out-group was automatically implemented, while their other two allocation decisions were discarded. Since this randomization approach was relatively complex, care was taken to write the instructions with sufficient detail and clarity to avoid confusion (our full instructions can be found in section L of the Online Appendix, and subjects were required to answer understanding test questions before they could proceed to their allocation decisions).[5] The order of the three allocation decisions was also

---

[3] The pre-registration can be found at https://www.socialscienceregistry.org/trials/10172 and https://www.socialscienceregistry.org/trials/11275. Ultimately, our execution of the experiments deviated slightly from the initial plan in that we were practically unable to attain the desired sample sizes and the number of sessions, but otherwise did not depart from it.

[4] Although the decision-maker could not keep any money for their self, they knew they would receive a payment from the task of either £6, £8 or £10 (or equivalent in €), randomly selected with equal probability.

[5] Subjects faced five control questions enquiring about the structure of the experiments and the incentives scheme after reading the instructions. Subjects answered 94% of questions correctly at their first attempt in the Nottingham experiments and 89% in the Castellón experiments. They had the opportunity to ask for clarifications and were obliged to enter the correct answer on their second attempt before they could proceed with the experiment.

randomized.

2.3. Norm-elicitation task

The social appropriateness of discrimination in the allocator game was measured using the well-established norm-elicitation method first introduced by Krupka and Weber (2013). Subjects were described the allocator game and required to evaluate the social appropriateness of each of the different allocations available to the decision-maker, by selecting on a four-point scale one of the following options: 'Very socially inappropriate', 'Somewhat socially inappropriate', 'Somewhat socially appropriate' or 'Very socially appropriate'.

When making evaluations, subjects were informed of the group identity of the decision-maker whose behaviour they were assessing, as well as that of both of the passive players the decision-maker was allocating between. Subjects only evaluated allocations made by decision-makers of their own group. They made three sets of evaluations, one for each of the groups the passive out-group player could belong to. Thus, the task separately reveals each subject's perception of the social appropriateness of any possible level of discrimination by members of their own group towards members of each of the other three groups along the relevant identity dimension in their treatment.

Subjects making the evaluations were the same as those who participated in the allocator games described in the task. Although there may be a concern in principle that participation in either part of the experiment might influence responses to the other, Barr et al. found an absence of evidence for this. Nevertheless, in the UK experiment we randomized across sessions whether subjects first partook in the allocator game or norm-elicitation task, so that we can also check for order effects in the current study.[6] The order in which subjects completed their three sets of evaluations was also randomized (within sessions), but subject to the constraint that the order of the three out-groups was the same for any given subject in the norm-elicitation task as it was in the allocator game.

Evaluations were incentivized. Each subject knew that, at the end of the experiment, one of the actions from one of the three sets of evaluations they had made, would be randomly selected. The subject's evaluation of this action would be compared with that of another randomly selected subject from the same treatment. The subject would receive a bonus of £8 if and only if their evaluation matched that of the person they were compared against. The norm-elicitation task, therefore, takes the form of a coordination game, where subjects are incentivized to provide the same evaluations as others. Importantly, subjects were told that the person to whom their evaluation would be compared would belong to their own identity group.[7]

The incentives are thus designed to guide subjects to reveal perceptions of appropriateness as commonly agreed by members of their own identity group (i.e. group-specific social norms), rather than personal opinions about morality. See Bicchieri (2005) for a discussion of the important distinction between personal opinions and social norms, the latter of which can be regarded as second-order beliefs about the appropriateness of behaviour. The way 'socially appropriate' behaviour is defined to subjects before they undertake a Krupka-Weber task is aimed at conveying to them this concept – in our experiment, the instructions told them to think of it as 'behaviour that you think most participants [of your identity group] in this experiment would agree is the "correct" thing to do.'

A possible concern is that the coordination incentives could lead subjects to report third – or higher–order beliefs. However, if subjects use salient focal points to coordinate, in the manner suggested by Schelling (1980), it is likely they will indeed report second-order beliefs since this is what they are asked to do, and answering the question truthfully seems by far the most salient available strategy. In principle, the Krupka-Weber method could produce responses wholly unrelated to social norms if subjects could find an alternative strategy to coordinate, but existing empirical research suggests this does not happen, even when other plausible focal points are made available (Fallucchi and Nosenzo, 2022). There is also evidence that the Krupka-Weber method provides norm estimates consistent with other methods where alternative coordination strategies are excluded by design (Bicchieri et al., 2022; Lane et al., 2023).[8]

2.4. Treatments

We conduct a between-subject design with three treatments, which differed according to the dimension of identity that the groups in the experiment were formed on the basis of. This dimension of identity was made salient at the beginning of the experiment by announcing the four groups, along this dimension, that the subjects in the treatment were divided between.

In two of the treatments, the groups were determined prior to the experiment based on real-world identity characteristics. Of course, the labels of the groupings differ between the two experiments, on account of the differences between the two countries in which they were run. In the UK, in the *Religion* treatment, we invited subjects who were either *Christian, Hindu, Muslim,* or *non-religious*. In the *Partisan*

---

[6] Subjects were not informed about the nature of the second activity until they had completed the first one. However, those who did the norm-elicitation task first were, just like those who first played the allocator game, made aware at the start of the experiment about the four different identity groups that subjects in their session were divided between. Thus, the dimension of identity made relevant in their treatment had already been made salient to all subjects before they began whichever was their first task.

[7] By incentivizing subjects to match the evaluation of *one* other subject, we depart slightly from the original design of Krupka and Weber (2013), in which they were incentivized to match the modal evaluation. However, the variant of the incentive scheme we use has also been frequently used, from early on, in the norm-elicitation literature (e.g. by Burks and Krupka, 2012; Barr et al., 2018). Conceptually, the two approaches should be strategically equivalent.

[8] For a methodological discussion of the Krupka-Weber method, see Nosenzo and Gorges (2020).

treatment, the invited subjects were supporters of one of the *Conservative, Green, Labour* or *Liberal Democrat* parties.[9] In the Spanish experiment, in the *Religion* treatment, we invited subjects who were either *Catholic, Muslim, agnostic*, or *atheist*. In the *Partisan* treatment, the invited subjects were supporters of one of the *PSOE* (Partido Socialista Obrero Español – Spanish Socialist Workers' Party), *PP* (Partido Popular – People's Party), Vox (Voice), and *Unidas Podemos* (United We Can) parties.[10] Meanwhile, in the *Artificial* treatment, group identity was created at the beginning of the experiment itself, using a similar method to Barr et al., which itself followed in the long tradition – stemming from Tajfel (1970) – of inducing 'minimal group' identity within an experiment. Upon entering the lab, subjects were instructed to blindly draw a ball from a bag. The colour of the ball – *blue, pink, red* or *yellow* – determined the subject's group.

We knew subjects' religious or partisan affiliation from a pre-survey we conducted in the weeks leading up to the lab experiment (see Online Appendix, section A for results and section L for the questionnaire). In the UK, partisan identity was derived from subjects' response to the question: 'If the next general election were held tomorrow, which party would you vote for?'[11] Given difficulties in Nottingham to attain a pool of eligible subjects of sufficient size, in Castellón we used both the voting question and the question 'name the party that you feel closer to than any other' to identify partisan identity. The pre-survey was distributed to subjects at the Universities of Nottingham and Jaume I registered on ORSEE for participation in economic experiments. From the responses, we identified the four most popular political parties and religious identities to employ as our groups in these treatments (see section A in the Online Appendix). We assigned to the experiment only those who belonged to one of the four chosen groups for both identity dimensions; these subjects were then randomly assigned either to the *Religion* or *Partisan* treatment. This ensures that subjects in these two treatments are drawn from the same wider sample, and therefore any differences in the outcomes of the experiment should be driven by the treatment manipulation (i.e. dimension of identity) rather than demographic differences between the subjects assigned to each treatment. We would ideally have recruited subjects for the *Artificial* treatment also from the same pool of individuals identified by the pre-survey as eligible for participation in both of the other treatments. Unfortunately, however, this pool was not sufficiently large to fill three treatments.[12] The most important results of our study should be considered the comparisons between the *Religion* and *Partisan* treatments, with the *Artificial-Religion* and *Artificial-Partisan* comparisons regarded as suggestive but not perfectly controlled.

2.5. Procedure and sample

After participating in both the allocator game and norm-elicitation task, each subject was randomly paid their earnings from only one of them, as determined randomly by a coin toss at the end of the session (this rule was made clear to subjects from the outset). All subjects additionally received a show up fee of £4 (€3 in Spain).

The experiment was conducted using Z-Tree (Fischbacher, 2007). In the UK, we ran four sessions for the *Artificial* and *Partisan* treatments, and three sessions for the *Religion* treatment; since we randomized across sessions whether the allocator game or norm-elicitation task was run first, we are able to check for order effects. Consistent with Barr et al., we do not find such effects (see section J in the Online Appendix) and will therefore pool our analysis in the next section across ordering conditions. In Spain, we conducted three sessions (one per treatment) and the norm elicitation task was always implemented after the decision task.

Sessions ranged in size from 13 to 29 subjects in Nottingham and from 57 to 60 in Castellon.[13] Table 1 presents the total number of subjects in each treatment. Relative sizes of each identity group can be found in section C of the Online Appendix.[14] Across the two experiments, subjects were 57% female, 23 years old on average, and 11% from rural areas. According to the self-reported ideology variable, 45% are left-wing, 25% centre and 30% right-wing. All subjects in the UK, and the vast majority in Spain, were students. Individual characteristics are similarly distributed between experimental conditions.

---

[9] The *Conservatives* are the governing right-wing party in the UK. Among our participants, Conservatives score 6.6 on the 1-10 left-right self-placement scale. *Labour* is the main, left-wing opposition, scoring 3 on the self-placement scale. The *Liberal Democrats* and *Greens* are smaller parties (but relatively popular among student populations) occupying, respectively, centrist internationalist and environmental leftist positions, scoring 4.4 and 3.9 on the scale, respectively.

[10] PSOE (centre-left, with supporting among our participants scoring on average 3 on the left-right scale) and PP (centre-right, 7.1) are the two majoritarian Spanish parties. *Vox* (8.2) is a relatively new far-right party, currently third at the national level. *United We Can* (1.5) is a left-wing coalition.

[11] Subjects were allowed to not select any party to this question if they did not consider any party worth voting for. Thus, we did not force partisan identity on subjects. In all treatments, subjects were reminded in the lab experiment of their group identity – this seemed particularly necessary in the *Partisan* treatment, in case some subjects had forgotten their answer to the voting intention question. Using vote intention as partisan identity is not without shortcomings. We acknowledge this weakness in the design. Fortunately, in Castellón we used party identification as the main variable to define the groupings, and we found the same treatment effects in both locations. Furthermore, party identification and vote intention are almost perfectly correlated. 92% of the participants feel closest to the party they intend to vote for.

[12] See sections A and B of the Online Appendix for further information on our recruitment approach and related methodological discussion.

[13] The maximum capacity of the Nottingham lab is 32; in Castellon it is 66.

[14] In Nottingham, some procedural errors were made in conducting the experiments. In the first *Artificial* session, a typo at one point on the instructions referred to "religion", where it should have said "ball colour grouping". At this point, the experimenter immediately verbally corrected the instructions to clarify what they should have said. In one *Religion* session in which the allocator game was conducted first, the full instructions from the previous *Religion* session (in which the norm-elicitation task was run first) were inadvertently left on subjects' desks and only spotted and removed after a few subjects had already taken their seats – therefore, these subjects may have briefly been exposed to the information about both tasks at the beginning of the experiment. Because these errors could in principle bias our results, we perform a robustness analysis in section I of the Online Appendix to check whether our main results are driven specifically by the flawed sessions. The analysis suggests this is not the case.

Table 1
Descriptive characteristics of participants and treatment assignment, by experiment.

	Total	Nottingham, UK	Castellón, Spain
	$N = 418$	$N = 241$	$N = 177$
Age	22.8 (7.6)	21 (2.9)	25.2 (10.7)
Gender			
Male	172 (41.1%)	108 (44.8%)	64 (36.2%)
Female	240 (57.4%)	129 (53.5%)	111 (62.7%)
Other	6 (1.4%)	4 (1.7%)	2 (1.1%)
Background			
Rural	47 (11.2%)	31 (12.9%)	16 (9.0%)
Urban	207 (49.5%)	123 (51.0%)	84 (47.5%)
Mixed	164 (39.2%)	87 (36.1%)	77 (43.5%)
Ideology			
Left	176 (42.1%)	102 (42.3%)	74 (41.8%)
Centre	97 (23.2%)	61 (25.3%)	36 (20.3%)
Right	115 (27.5%)	48 (19.9%)	67 (37.9%)
Missing	30 (7.2%)	30 (12.4%)	0 (0.0%)
Treatment			
Artificial	146 (34.9%)	86 (35.7%)	60 (33.9%)
Partisan	145 (34.7%)	85 (35.3%)	60 (33.9%)
Religion	127 (30.4%)	70 (29.0%)	57 (32.2%)

Notes: Distribution of gender, background and ideology of participants. Age is presented as a continuous variable, including the mean and standard deviation. Ideology is measured using a standard 0–10 left-right survey question. *Left* is defined as 0–4, *Centre* as 5, and *Right* as 6–10. Treatment assignment is also included, by location of the experiments.

2.6. Predictions

Following the literature on the pre-eminence of partyism (Westwood et al., 2018), Lane's (2016) meta-analysis on discrimination based on natural and artificial identities, and the previous behavioural study of Barr et al. (2018), we derive two core predictions:

Prediction 1. the social appropriateness of in-group favouritism will be ranked across treatments as following:

Partisan > Artificial > Religion

Prediction 2. participants will behave according to social norms when presented with the allocation task.

3. Results

In the following, we focus almost exclusively on the results of the *Partisan* and *Religion* treatments. Since the recruitment procedure for the *Artificial* treatment was not perfectly comparable to the other treatments, we only report detailed analyses comparing this treatment with the others in the Online Appendix, section E. However, Figs. 1 and 2 below include the descriptive results of the *Artificial* treatment for the sake of completeness. Results are pooled across the two countries for ease of exposition because they largely do not differ. However, we will point out any notable differences between countries that do arise. The complete analysis for each country is separately reported in full in the Online Appendix, section F.

We start by examining the results from the norm-elicitation task, in which individuals assess the social appropriateness of the range of allocation decisions. There is a clear consensus in recognizing the equal-split choice – allocating 8 monetary units to each participant – as the most appropriate behaviour: pooling across the *Religion* and *Partisan* treatments, 80% of ratings for this decision were "very socially appropriate". Yet, variations across treatments are observed. To examine these, following Krupka & Weber (2013) and Barr et al. (2018), we transform evaluations into numerical values, assigning evenly-spaced values of −1 for the "very socially inappropriate" rating, −0.33 in the case of "somewhat socially inappropriate", 0.33 for "somewhat socially appropriate" and 1 for "very socially appropriate". We then present, in Fig. 1, mean ratings for each allocation in each treatment, pooling all assessments made by subjects of all groups regarding the given allocation towards each of their out-groups. The *Artificial* treatment is depicted with a dashed line to highlight the difference in recruitment procedure from the other treatments. The complete distributions these means are based on are presented in Table 2, along with the p-values on Fisher-Pitman permutation tests comparing them between treatments.[15]

---

[15] An alternative approach to determine the statistical significance of these treatment differences in social appropriateness would be performing regression analysis. We have conducted this analysis and verified the results are qualitatively the same as the permutation tests', except for the two extreme allocations favouring the out-group (where the reported significance of the differences between treatments is at <10% in the regressions). We opt for permutation tests as they are non-parametric, allowing inference in a more flexible statistical framework which arguably better fits experimental data. As we perform a total of 27 tests, 9 by treatment comparison – including those involving the *Artificial* treatment (see Online Appendix, Section E) – the reported p-values are corrected with the Benjamini-Hochberg False Discovery Rate method. This method accounts for the increasing probability of reporting a false result when performing multiple tests. The Benjamini-Hochberg method consists of sorting the p-values in ascending order, multiplying by the total number of tests performed and dividing by the rank number.
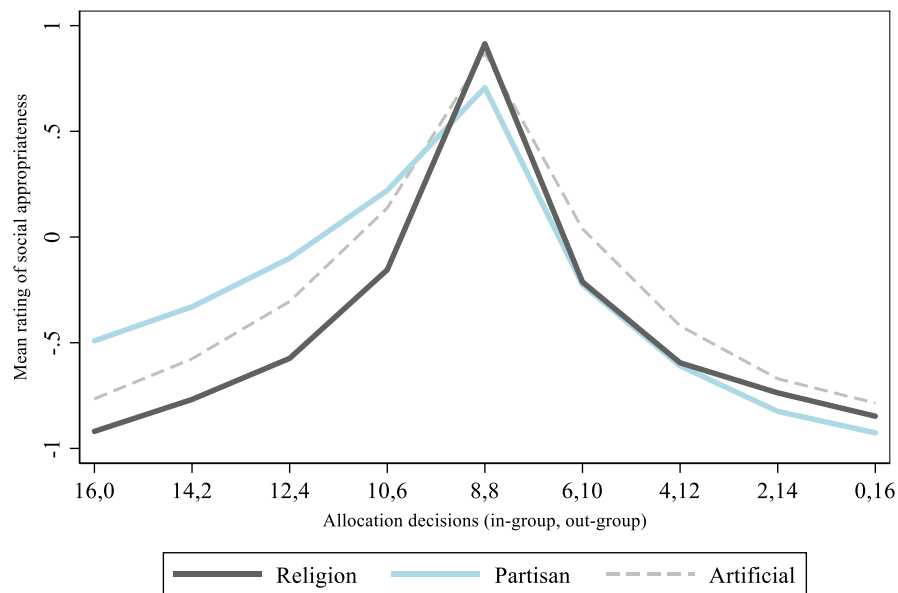
Fig. 1. Mean evaluation of social appropriateness of allocation decisions, by treatment.
Note: Social appropriateness ratings range from −1 (if complete consensus on social inappropriateness) to 1 (in the case of absolute consensus on social appropriateness). Ratings comprise 381 observations from 127 subjects in the *Religion* treatment, 435 observations from 145 subjects in the *Partisan* treatment, and 438 observations from 146 subjects in the *Artificial* treatment.

Dividing the money equally is regarded significantly less appropriate in the *Partisan* treatment than in the *Religion* treatment. 70.4% of responses in the *Partisan* treatment rated this choice as "very socially appropriate", in contrast to 90.8% and 87.7% in the *Religion* and *Artificial* treatments, respectively. In all the treatments, social appropriateness decays rapidly as choices deviate from the equal-split standard toward more discriminatory choices. However, for in-group favouring allocations, that decline is faster in the *Religion* treatment; all choices that favour the in-group participants are significantly less socially inappropriate in the *Partisan* treatment than in the *Religion* treatment. Even the smallest deviation possible from the equal choice is extensively more censured in the *Religion* treatment than in the *Partisan* case: 54.9% of responses deem the (10,6) allocation – giving 10 monetary units to the in-group member and 6 to the out-group member – to be either "somewhat socially inappropriate" or "very socially inappropriate", in contrast to 29.3% in the *Partisan* treatment. In contrast, actions favouring the out-group are rated as more inappropriate in the *Partisan* treatment than the equivalent actions are in the *Religion* treatment, significantly so for the most extreme allocations – giving 16 or 14 monetary units to the out-group member and 0 or 2 to the in-group member.

In the case of *Religion*, the inappropriateness of discriminatory choices is largely symmetric, meaning participants evaluate favouring the in-group participant and favouring the out-group participant alike. Hence, in the *Religion* framework, discriminating in favour of a participant of your group is regarded as equally socially inappropriate, or even more so, than favouring other groups. For the *Partisan* framework, the opposite is true. Evaluations in this treatment are not symmetric: the social appropriateness of discriminatory choices favouring the out-group declines faster than those favouring the in-group. In the extreme splits, allocating all the money to the in-group participant is regarded as inappropriate – either somewhat or very – by 75% of responses, in contrast to 97% when allocating all the money to the out-group. Between 10 and 20% of responses rate any discriminatory behaviour favouring the party political group of the allocator as "very appropriate", in contrast to almost none providing the same rating to behaviour favouring the out-group. All in all, the data reveal a clear pattern in the *Partisan* treatment of strong rejection of out-group advantage and weaker rejection of in-group favouritism. Does this pattern of normative beliefs guide actual behaviour?

We now focus on decisions made in the allocation tasks. As with the evaluations, we report detailed analyses corresponding to the *Religion* and *Partisan* treatments and additionally include descriptive results of the *Artificial* treatment in white bars in Fig. 2 for completeness. This figure reveals two broad tendencies quite clearly. First, in correspondence to the norm elicitation results, the modal distribution is the equal split, which accounts for 60% of allocations across the *Religion* and *Partisan* treatments. Second, those that deviate do so mostly to favour the in-group member, with only 2.8% of allocations favouring the out-group member. On average, participants in these treatments allocate 3.03 monetary units more to the partner with whom they share an identity. Even so, as was the case for the norm-elicitation task, there are significant differences across treatments. The average in-group premium is 1.85 monetary units in the *Religion* treatment, with 74.3% of allocations producing an equal (8,8) split, compared to 4.06 monetary units in the *Partisan* treatment, where only 47.8% of allocations were the equal split.
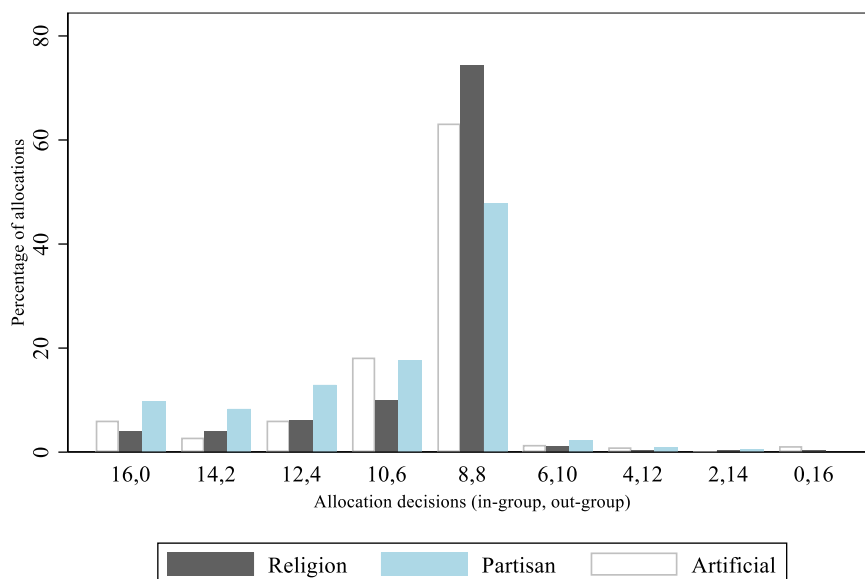
Fig. 2. Allocation decisions.

Note: Percentages of allocation decisions, by treatment. Each participant made three allocation decisions, one for each different possible affiliation of the out-group participant. The figure includes 381 allocation decisions from 127 subjects in the *Religion* treatment, 433 allocation decisions from 145 subjects in the *Partisan* treatment, and 429 allocation decisions from 143 subjects in the *Artificial* treatment. In the *Partisan* treatment, there are 2 missing observations from 2 subjects who could not complete one of their allocation decisions.

Table 3 presents a random effects model[16] of the allocation decisions. The dependent variable is the in-group premium, measured by subtracting the amount allocated to the out-group from the amount allocated to the in-group. Columns (1) and (2) show aggregated results for the full sample while columns (3) and (4) include a binary variable and an interaction term showing differences between the Nottingham and Castellón experiments. Outcomes confirm the observed difference of more than two additional pounds in the in-group premium between the *Religion* and *Partisan* treatments. That difference is statistically significant at a 99% confidence level. This result is robust to a set of control variables that account for individual characteristics. Alternative specifications including other individual characteristics such as ideology, party affiliation and religion do not alter the results, with no statistically significant effects of the additional control variables either.

Thus, the ranking of discrimination across the treatments matches that of the perceived social appropriateness of discrimination discussed above. As shown in Online Appendix E, this remains true when adding into the analysis the *Artificial* treatment, which lies in between the other two both in the amount of discrimination and the social inappropriateness of in-group-favouring allocations. This is in line with our predictions, although not all treatment differences involving the *Artificial* condition are significant.

The crucial contribution of having the experiments in both countries is that although there is a strong and significant difference between countries in the overall levels of discrimination (see the coefficient for *Spain* in columns (3) and (4) of Table 3), there are not significant differences in the treatment effect (see the interaction *Spain*Partisan*). Furthermore, in both experiments the ranking of treatments by discrimination reflects the ranking according to the social appropriateness of discrimination. This confers credibility to the argument that our findings on partisan identity are robust to changes in the social context. Regardless of a society's baseline tendency for discrimination, the political party domain induces higher levels of discrimination than religion, which reflects the relatively weak norms it induces against such discrimination.

Finally, an interesting result emerges that fosters our confidence on the independence of the normative and behavioural measures. Although in the *Religion* treatment in Castellón there is high agreement on the social inappropriateness of in-group favouritism, participants in this treatment do discriminate somewhat frequently. This means that discriminatory behaviour is not fully determined

---

[16] Given that each participant had to make three choices, one for each possible out-group, a specification including random effects allows us to control for individual observable and unobservable characteristics beyond the included control variables.

Table 2
Distribution of social appropriateness ratings, by treatment.

	16–0	14–2	12–4	10–6	8–8	6–10	12–4	14–2	0–16
	Favouring the in-group					Favouring the out-group			
Partisan treatment									
Very appropriate	17.8	13.2	13.4	19.2	70.4	2.5	0.5	0.2	0.2
Somewhat appropriate	6.9	17.3	31.9	51.5	18.5	31.6	7.4	3.2	2.5
Somewhat inappropriate	9.5	26.6	31.4	22.6	7.6	44.8	42.0	18.7	4.9
Very inappropriate	65.8	43.0	23.3	6.7	3.5	21.0	50.1	77.8	92.4
Mean rating	**−0.49**	**−0.33**	**−0.10**	**0.22**	**0.71**	**−0.23**	**−0.61**	**−0.83**	**−0.93**
Religion treatment									
Very appropriate	2.1	2.6	2.6	4.7	90.8	5.3	3.9	2.9	3.4
Somewhat appropriate	0.8	1.8	9.5	40.4	6.8	33.3	8.4	6.0	2.9
Somewhat inappropriate	4.2	23.1	37.0	31.8	1.1	35.7	32.0	18.6	6.8
Very inappropriate	92.9	72.4	50.9	23.1	1.3	25.7	55.6	72.4	86.9
Mean rating	**−0.92**	**−0.77**	**−0.57**	**−0.16**	**0.91**	**−0.21**	**−0.59**	**−0.74**	**−0.85**
p-value	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.721**	**0.687**	**0.003**	**0.001**

Notes: Percentage of responses corresponding to each of the available social appropriateness ratings for each of the nine possible decisions in the allocator game. Decisions range from allocating all the money to the in-group participant (16,0) to allocating all the money to the out-group participant (0,16). The modal evaluation for each outcome is highlighted. Mean ratings are computed assigning values of 1, 0.33, −0.33 and −1 to assessments of "very socially appropriate", "somewhat socially appropriate", "somewhat socially inappropriate" and "very socially inappropriate", respectively. Therefore, mean ratings range from −1 (if complete consensus on social inappropriateness) to 1 (in the case of absolute consensus on social appropriateness). Ratings comprise 381 observations from 127 subjects in the *Religion* treatment and 435 observations from 145 subjects in the *Partisan* treatment. P-values obtained from Fisher-Pitman permutation tests for two independent samples, corrected using the Benjamini-Hochberg False Discovery Rate method.

by norms reported in the lab, but also reflects personal preferences which in some cases may be to violate them.[17] Furthermore, these results can be interpreted to mean that different individuals are more or less sensitive to the norm.[18]

4. Discussion

Political sectarianism, extremism, and polarization are among the social phenomena about which citizens in Western democracies are more concerned (Finkel et al., 2020). These issues feature highly in the news and all forms of media, traditional or new. They are also among the most frequent social problems identified by citizens. Political polarization is becoming so severe that it is even bringing democracy itself and societal stability under threat as it undermines support for democratic norms (Kingzette et al., 2021). Descriptive studies in the social sciences defining these topics abound. However, there is a lack of studies attempting to address the causal processes underlying such social phenomena. In this paper, we explore one mechanism explaining the pre-eminence of political polarization over other social divides: the role of social norms. We follow a long tradition in social psychology and economics that goes back to the works of Tajfel (1970) and conduct experiments in which we use group identities and elicit discriminatory choices. More specifically, we base our study on the previous investigation of Barr et al. (2018) that introduced an experimental design to study norms and discriminatory behaviour in different social contexts. We focus specifically on partisanship, a form of political identity.

Our main results are two. First, we replicate the finding that discrimination is largest when norms against it are weaker. Second, regardless of the experimental setting, most discrimination is observed in the partisan domain. We conducted the experiment in two locations that differed in their baseline tendency for discrimination, but always found by far the strongest discrimination in the *Partisan* treatment, which also mirrored the fact that in both countries norms against discrimination were weakest in this domain.

Why is the party political domain different? We argue that this domain may lack the sort of norms that prevent discrimination, prejudice, and hostility between social groups in society. Interactions across religion or nationality, also race, gender, and other social divides are constrained by social norms, but there seem not to be corresponding pressures or sanctions that mute disapproval of party political opponents. Partisans, therefore, are not strongly restricted from expressing animus and engaging in discriminatory behaviour toward their opponents. We acknowledge that partisanship is only one form of political identity; future research could explore whether a similar relationship between norms and political discrimination emerges from investigating other forms.

---

[17] For additional credibility in the relationship between norms and discriminatory behaviour, we perform a regression analysis demonstrating how, in each domain, the differences in social appropriateness of allocations are reflected in differences in their likelihood of being selected. See Online Appendix, section K.

[18] In section G of the Online Appendix we take advantage of the four identity categories in the *Partisan* treatment to study whether results are driven by some subgroup behaviour. Restrepo-Plaza and Fatas (2023) show that the ideological distance between groups is important in a polarized context. Our design incorporates this idea, by eliciting decisions and norms for every possible in-group-out-group pairing. Hence, if different out-groups yield different degrees of discrimination based on social distance, they are all jointly considered in every treatment. See also section D of the Online Appendix for a detailed description of the average allocation decisions across treatments and sub-groups, including cross-party and religion allocations.

Table 3

Regression analysis of allocation decisions.

| | Dependent variable | | | |
| | Amount allocated to in-group – amount allocated to out-group participant | | | |
	(1)	(2)	(3)	(4)
Treatment:				
Partisan	2.214***	2.388***	2.191***	2.247***
	(0.486)	(0.496)	(0.516)	(0.636)
Spain			2.165***	2.155***
			(0.624)	(0.733)
Spain*Partisan			0.235	0.319
			(0.986)	(0.97)
Controls	✘	✔️	✘	✔️
Constant	1.848***	2.161	0.876***	2.357
	(0.31)	(1.534)	(0.294)	(1.489)
Observations	814	814	814	814

Note: Results for random effects models. The reference category is the *Religion* treatment. Control variables include: gender (binary variable, 1=female), age, year of university degree the participant is in, rural if they are originally from a rural area, and household income. Standard errors robust to heteroscedasticity are in parentheses. Stars indicate significance level: *** $p<.01$, ** $p<.05$, * $p<.1$. Observations[*] include 381 allocation decisions from 127 subjects in the Religion treatment and 433 allocation decisions from 145 subjects in the Partisan treatment.

[*] In the Nottingham experiments, we have incomplete allocation decisions from five subjects as a result of glitches in the software which prevented them from entering responses. This results in 11 missed allocation observations in the final dataset: 9 in the *Artificial* treatment (not included here) from three subjects who could not introduce any allocation decision and 2 additional missing observations in the *Partisan* treatment from 2 subjects who could not complete one of their allocation decisions. We also had a person who came twice to the experiment; all the data from their second participation has been excluded from all analyses.

The deeper question of *why* norms have not developed to counter partisan discrimination is beyond the scope of this paper, but we can offer the reader speculation. In a democratic system, people are free to select in and out of partisan groups. A person's political party leanings are a reflection of their philosophical values and opinions about society. It may be regarded as less unacceptable to dislike a person because of their worldview than on the basis of a characteristic such as race or sex which is unchosen and fixed from birth.[19] This is an important matter for future research.

We believe our design sets the stage for international comparisons of the role of social norms on political discrimination and polarization. We have shown that our treatment effects remain qualitatively the same in two societies that differ in their baseline tendencies for discrimination and in their political ecosystems. A worthwhile endeavour would be to extend the analysis to a wider set of more different societies, including non-Western countries and those with less democratic political systems. The relative simplicity of our experimental design would allow it to be easily implemented in large-sample online studies. Finally, understanding the role of norms in political polarization may help devise interventions to return political conflict to the realm of ideas, reducing interpersonal hostilities. Once knowledge is established about existing norms, how they vary across contexts and the mechanisms behind their development and enforcement, one could envision and formulate social interventions that help break the vicious cycle of political group hostility (Dimant, 2023).

Acknowledgments

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.euroecorev.2023.104661.

---

[19] It is interesting to note that religion, which we also employed in our experiment, is at its root also a matter of belief, with people able to switch between affiliations. In practice, however, religion has evolved into a characteristic that is explicitly inherited from one's parents, and conversion is a rare event. As such, religion may be viewed less as a choice-based identity group like partisanship, and more akin to an immutable one like ethnicity.

References

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

Barr, A., Lane, T., Nosenzo, D., 2018. On the social inappropriateness of discrimination. J. Public Econ. 164, 153–164.

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

Bicchieri, C., Dimant, E., Gächter, S., Nosenzo, D., 2022. Social proximity and the erosion of norm compliance. Game Econ. Behav. 132, 59–72.

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. Manage. Sci. 58 (1), 203–217.

Clark, C.J., Liu, B.S., Winegard, B.M., Ditto, P.H., 2019. Tribalism is human nature. Curr. Dir. Psychol. Sci. 28 (6), 587–592.

Coffman, K.B., Exley, C.L., Niederle, M., 2021. The role of beliefs in driving gender discrimination. Manage. Sci. 67 (6), 3551–3569.

Cohen, N., 2016. Did better broadband make Americans more partisan? The Guardian. https://www.theguardian.com/commentisfree/2016/aug/27/donald-trump-fox-news-charlie-sykes-modern-debate.

Dimant, E., 2023. Hate trumps love: the impact of political polarization on social preferences. Manage. Sci.

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

Fallucchi, F., Nosenzo, D., 2022. The coordinating power of social norms. Exp. Econ. 25 (1), 1–25.

Fehr, E., Schurtenberger, I., 2018. Normative foundations of human cooperation. Nat. Human Behav. 2 (7), 458–468.

Finkel, E.J., et al., 2020. Political sectarianism in America. Science 370 (6516), 533–536.

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.

Gaertner, S.L., Dovidio, J.F., 2000. Reducing intergroup bias: The Common Ingroup Identity Model. The Psychology Press.

Gidron, N., Adams, J., Horne, W., 2020. American Affective Polarization in Comparative Perspective. Cambridge University Press.

Iyengar, S., Westwood, S.J., 2015. Fear and loathing across party lines: new evidence on group polarization. Am Polit Sci Rev 59 (3), 690–707.

Iyengar, S., Lelkes, Y., Levendusky, M., Malhotra, N., Westwood, S.J., 2019. The origins and consequences of affective polarization in the United States. Annu Rev of Polit Sci 22 (1), 129–146.

Kimbrough, E.O., Vostroknutov, A., 2016. Norms make preferences social. J. Eur. Econ. Assoc. 14 (3), 608–638.

Kingzette, J., Druckman, J., Klar, S., Krupnikov, Y., Levendusky, M., Ryan, J.B., 2021. How affective polarization undermines support for democratic norms. Public Opin. Quart. 85 (2), 663–677.

Köbis, N.C., Troost, M., Brandt, C.O., Soraperra, I., 2022. Social norms of corruption in the field: social nudges on posters can help to reduce bribery. Behav. Public Policy 6 (4), 597–624.

Krupka, E.L., Leider, S., Jiang, M., 2017. A meeting of the minds: informal agreements and social norms. Manage. Sci. 63 (6), 1708–1729.

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.

Lane, T., 2016. Discrimination in the laboratory: a meta-analysis of economics experiments. Eur. Econ. Rev. 90, 375–402.

Lane, T., Nosenzo, D., Sonderegger, S., 2023. Law and Norms: empirical Evidence. Am. Econ. Rev.

Michelitch, K., 2015. Does electoral competition exacerbate interethnic or interpartisan economic discrimination? Evidence from a field experiment in market price bargaining. Am. Pol. Sci. Rev. 109 (1), 43–61.

Nosenzo, D., Görges, L., 2020. Measuring social norms in economics: why it is important and how it is done. Analyse & Kritik 42 (2), 285–312.

Ostrom, E., 2000. Collective action and the evolution of social norms. J. Econ. Perspect. 14 (3), 137–158.

Restrepo-Plaza, L., Fatas, E., 2022. When ingroup favoritism is not the social norm a lab-in-the-field experiment with victims and non-victims of conflict in Colombia. J. Econ. Behav. Organ. 194, 363–383.

Restrepo-Plaza, L., Fatas, E., 2023. Building inclusive institutions in polarized scenarios. Const. Polit. Econ. 34, 88–110.

Schelling, T.C., 1980. The Strategy of Conflict. Harvard university press.

Sunstein, C.R., 2016. Partyism. Univ. Chicago Leg. For. 2015 (2), 1–26.

Szekely, A., Lipari, F., Antonioni, A., Paolucci, M., Sánchez, A., Tummolini, L., Andrighetto, G., 2021. Evidence from a long-term experiment that collective risks change social norms and promote cooperation. Nature Commun 12 (1), 5452.

Tajfel, H., 1970. Experiments in Intergroup Discrimination. Sci Am 223 (5), 96–103.

Westwood, S.J., Iyengar, S., Walgrave, S., Leonisio, R., Miller, L., Strijbis, O., 2018. The Tie That Divides: cross-National Evidence of the Primacy of Partyism. Eur. J. Polit. Res. 57 (2), 333–354.