University of Nottingham
UK | CHINA | MALAYSIA

NICEP
Nottingham Interdisciplinary Centre for Economic and Political Research

NICEP Working Paper: 2020-06

# The Brexit referendum and the rise in hate crime; conforming to the new norm

**Facundo Albornoz, Jake Bradley and Silvia Sonderegger**

**Nottingham Interdisciplinary Centre for Economic and Political Research**
https://nicep.nottingham.ac.uk
School of Economics and School of Politics and International Relations,
University Park, Nottingham, NG7 2RD
ISSN 2397-9771

# The Brexit referendum and the rise in hate crime; conforming to the new norm[a]

Facundo Albornoz[b] Jake Bradley[c] Silvia Sonderegger[d]

November 9, 2020

**Abstract**

We show that the sharp increase in hate crime following the Brexit referendum was more pronounced in more pro-remain areas. This is consistent with a model where behavior is dictated by the desire to conform to imperfectly observed social norms in addition to following individual preferences, and where the referendum revealed that society's real preferences over immigration were less positive than previously thought. For identification, we exploit the feature that the referendum revealed new information overnight in a context where other determinants of attitudes remained constant. The data can be replicated with a sensible parameterization of the model.

**Keywords:** Hate crime, Brexit, attitudes towards immigrants, social norms, value of information.

[a]We are grateful to Fabrizio Adriani, Gilat Levy, Daniel Seidmann, Francesco Squintani, and especially Antonio Cabrales, Giacomo Corneo, Guillermo Cruces and Carlo Schwarz, for valuable comments and suggestions, as well as seminar participants at the Freie Universität Berlin, King's College London, University of Birmingham, University of Bristol, University of East Anglia, University of Loughborough, University of Nottingham, University of Pennsylvania, Royal Economic Society 2019 and NICEP 2018. Justus Meyer and Nada Abdelghany provided outstanding research assistance.

[b]University of Nottingham and CONICET; email: facundo.albornoz@nottingham.ac.uk

[c]University of Nottingham and IZA; email: jake.bradley@nottingham.ac.uk

[d]University of Nottingham; email: silvia.sonderegger@nottingham.ac.uk

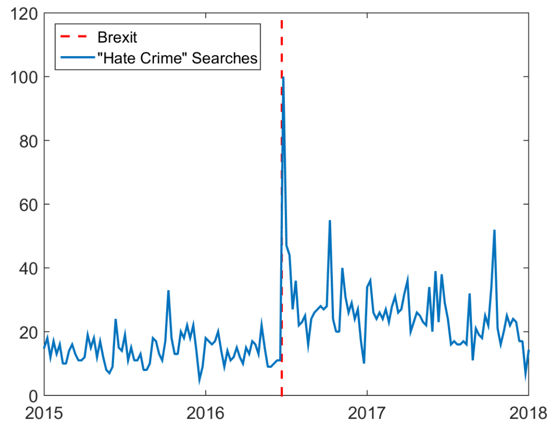
# 1 Introduction

Economists traditionally view attitudes (more generally, decisions) as a direct expression of underlying preferences. In the last decades, however, there has been a growing recognition of the social dimension of individual decisions. Starting from the early contributions by Akerlof (1980) and Elster (1989), much ground has been covered in acknowledging the role played by social norms in shaping individual behavior. Social norms depend on underlying preferences in society, as well as on the beliefs that people hold about what others think and do. These elements are however often imperfectly observed. Therefore, events that update perceptions about others can lead to changes in social norms. In turn, shifts in social norms translate into observable changes in social behavior. In this paper, we study this general abstract question within the context of the United Kingdom's Brexit referendum that took place in 2016 and the associated change in hate crime.

There is broad agreement (Meleady et al. (2017), Clarke and Whiteley (2017), Goodwin and Milazzo (2017), Becker et al. (2017)) that animus towards immigrants played an important role in shaping support for the leave option in the referendum. Fetzer (2019) shows that the rise of the UK Independence Party (UKIP) – which arguably gained consensus by capitalizing on concerns about rising immigration – was the strongest correlate of the Leave vote in the Brexit referendum. In the immediate aftermath of the vote, social media outlets started denouncing episodes of intolerance and abuse towards immigrants. This is reflected in a sharp spike in UK-originating internet searches for "*hate crime*" that coincides with the date of the referendum, as illustrated in Figure 1. The increase appears persistent as the average search in the 18 months after the referendum is 50% larger than the 18 months before. Traditional media also picked up on the phenomenon, with headlines such as "*Polish media shocked by post-Brexit hate crimes*" (BBC news, 28 June 2016) and "*UK 'more racist' after Brexit*" (The Times, 12 May 2018) to cite but a few.

The focus of this paper is to understand the extent to which the new public information about prevailing private views on immigration (which was released by the referendum) triggered a change in the norms that govern behavior towards ethnic and religious minorities. We see the rise in hate crime as an extreme expression of a more general shift in attitudes. The advantage of using hate

Figure 1: Weekly UK Google Searches for 'Hate Crime'




**Note:** Data are downloaded using 'Google trends', searches are normalized to the maximum number of searches in the period running from the first week of January 2015 until the last week of December 2017.

crime is that, precisely because it is extreme and violent behavior and thus against the law, it is documented in crime statistics and is therefore easily measurable. In general, attitudes are affected by shifts in individual preferences or by changes in social norms. A fundamental challenge is that neither of these are directly observable. One possible interpretation of the surge in hate crime after the referendum is that British people changed their views over the presence of immigrants. This would imply that the observed change in hate crime reflects a shift in preferences towards immigrants. An alternative and competing interpretation is that the referendum result legitimized previously sanctioned views towards immigrants to be expressed publicly. In this case, the surge in racially or ethnically motivated crime is compatible with a fixed share of xenophobes in the population, who feel emboldened by the referendum results and thus become more likely to act according to their personal views. Clearly, an informed policy reaction requires identifying which one of these broad two interpretations prevails and their relative influence.

Fortunately, the task of empirical identification is greatly facilitated by the unique features of the Brexit referendum. First, as mentioned, the referendum result was informative of people's privately held views on the presence of immigrants. One can identify cross borough differences in voting patterns and link them to differences along other measures such as conflict, racial attacks and survey answers. Second, an most importantly, a crucial aspect of the vote is that, by itself, it had no immediate impact on legislation and policies.[1]

In spite of this, the referendum had an abrupt effect on behavior, as exemplified by the spike in hate-crimes which took place in its immediate aftermath. This change cannot be explained by different policies or other economically-relevant variables, and, therefore, must be attributed to alternative explanations, namely a shift in preferences and/or a change in the prevailing social norm. For this reason, we argue that the referendum provides an opportunity to study the effect of new information on social norms and allows for a difference-in-difference analysis in a very natural way.

We start by assessing and quantifying the post-referendum changes in hate crime, using data collected from local police forces in England and Wales. Although this is mainly a descriptive task, our findings are interesting in their own right. Our data confirm that there was an abrupt rise hate crime in the aftermath of the referendum. Moreover, the phenomenon was not necessarily short-lived: the data we collected run up to mid-2017 and are consistent with a longer term effect.

Our analysis reveals that the change in behavior towards towards immigrants, as expressed by the increment in hate crime, was more pronounced in remain rather than leave areas, in spite of these areas having generally lower incidents of hate crime per immigrant. This sounds almost counter-intuitive but in fact is consistent with an information-based mechanism. The intuition relies on what we call the referendum's *surprise effect*, namely the notion that the referendum outcome created new common knowledge about the true extent of anti-immigrant sentiment in the country. The leave camp victory was, by and large, unexpected. On the day preceding the referendum, for instance, betting markets placed an 86% probability on a remain victory. We argue that, to the extent that the leave victory served as a public revelation that anti-immigrant views across the country were more widespread than was previously believed, this caused the norm to shift, rendering anti-immigrant attitudes more acceptable. We use data collected by the British

[1] However, the referendum did have an immediate effect on the exchange rate (e.g. Costa et al., 2019) and, even if not immediately, on the prospects of the economy (Bloom et al., 2018). While an economic-based complementary explanation could explain a jump in hate crime, it can hardly replicate the geographical variation that we uncover and explain.

Election Study (BES) – a long-run panel survey featuring around 30,000 respondents – to support the hypothesis that the referendum's surprise effect was stronger in areas with a higher share of remain votes. In turn, this caused the behavioral adjustment (reflected in the increment in hate crime) to be larger in those areas. Intuitively, if you live in an area of the country where you are surrounded by people who are very supportive of diversity, you will tend to believe that there is a strong norm against expressing anti-immigrant views, and therefore you will refrain from doing so, even if you are not a big fan of immigrants yourself. If you then discover that, actually, the country at large shares your views, you are going to adjust your behavior more than if you lived in a very anti-immigrant area, where you did not feel the presence of a strong pro-immigration norm to start with.

Of course, our analysis is based on reported hate crime, which naturally raises the concern about whether our findings are driven by changes in reporting or police recording practices. We find that there is a potential upswing in reporting rates at the time of the Brexit referendum. More specifically, an area with a ten-percentage point larger remain share reported hate crimes with a greater propensity, somewhere between 0.12 and 0.18 percentage points. However, this effect, even if interesting in its own right, vanishes after controlling for force-year fixed effects. Since potential changes in reporting are driven by variation in police force characteristics (e.g. changes in recording), our results on hate crime must be robust to the inclusion of police-year fixed effects. Reassuringly, we find that this is indeed the case.

To formally explore the underlying mechanism at play, we build a theoretical framework that draws on the so-called value of information literature (Angeletos and Pavan (2007), Grout et al. (2015)), which originates from the literature on Global Games. The underlying premise is that individuals are not only concerned with economically relevant factors or personal preferences, but also want to conform to what they think is average behavior across the country – the country-wide social norm. In our model, average behavior is only imperfectly observed. This grants a fundamental role to people's perception of the private views held by others. If people believe that everybody else finds a certain behavior to be unacceptable, then very few adopt that behavior and this becomes the norm. Importantly, in a context where the social norm is imperfectly observed, perceptions are influenced by what people can observe, such as prevalent views in geographically close areas e.g. districts or neighborhoods. This creates a potential discrepancy between observed views and the real country-wide preferences. Any update in terms of information about the latter can then result in a change in behavior. This is particularly true in events where all of a sudden it becomes publicly clear – e.g., through a referendum – that actually many agree with the behavior that was previously thought to be socially unacceptable. At that point, the norm changes abruptly, and people become considerably more likely to adopt that behavior. This mechanism explains changes in attitudes that simply follow from updates on the perception of the social norm. The direction and the extend of the behavioral change will be a direct function of the discrepancy between previous and updated perceptions – what we call the "surprise effect". This result can explain sudden changes in attitudes that vary across different regions within a same country.

Our model derives contrasting theoretical predictions on the effect of the referendum across regions for the possible mechanisms identified, namely preference-driven and a norm-driven. We show that the fact of a greater effect of the referendum in remain areas is consistent with the theoretical predictions of the model, and in particular with the norm-driven interpretation of the evidence, as explained above. To illustrate the different predicted patterns of norm-driven or preference-driven behavioral changes, we analyze the geographical distribution of the rise in hate crime triggered by a different event – namely the 7/7 terrorist attacks by Islamic fundamentalists that took place in London in 2005 – which arguably fostered an increment in anti-immigrant preferences. We show that the increment in hate crime following the attacks was more pronounced in leave areas, namely the *opposite* of what we observed in the aftermath of the Brexit referendum. This suggests that preference shocks (such as that caused by the terrorist attack and subsequent media frenzy) will generate a stronger behavioral response in areas where anti-immigrant sentiment is already widespread.[2] This finding is broadly consistent with existing studies such as Müller and Schwarz (2019b) and Bursztyn

[2] Of course using voting pattern in 2016 to infer anti-immigrant preferences in 2005 implicitly assumes persistence across time in anti-immigrant sentiment. This is in line with existing evidence (see e.g. Becker and Fetzer (2016) and Becker et al. (2017)).

et al. (2019), who find that exposure to social media and to Trump anti-immigrant tweets increases hate crime, but only in areas with high pre-existig prejudice.[3] Based on these observations, we argue that a change in preferences is unlikely to have triggered the post-referendum increase in hate crime.

Finally, we unify the theoretical model and empirical evaluation by structurally estimating a version of the model to be presented. In order to do this, me make parametric assumptions regarding voting behavior and how attitudes manifest into hate crime. Estimates suggest that an individual's propensity to conform to a societal norm determines approximately a quarter of their overall behavior. This conformity parameter is something that, to our knowledge, has not been estimated outside of a laboratory environment. Secondly, our estimates confirm that the size of the informational shock generated by the outcome of the Brexit referendum – as measured by the difference between prior and posterior beliefs about underlying views in society – was indeed large. In this respect the referendum really acted as a *bolt from the blue.*

The estimates allow us to look, in a semi-quantitative way, at a series of thought experiments which may be informative to policymakers in the future. We are able to quantify the role of shared narratives, national identity and stereotypes in shaping aggregate behavior. For instance, Australians are expected to be laid-back and relaxed, and the Swiss to be precise and punctual. We argue that these "*national stereotypes*" actually feed back into individual behavior and may result in people in separate societies behaving quite differently, even if their underlying preferences are the same. The behavioral difference will correspond to approximately 23% of the difference in prior beliefs concerning preferences in the two societies ("*stereotypes*"). This underscores the importance of shared national narratives promoted for instance by public figures and cultural influencers for actual behavior. It also shows that history can cast a long shadow in shaping norms. By revealing new information about underlying preferences, the Brexit referendum affected the dominant social norm and contributed to shaping a new national stereotype. In this form, it may continue to influence behavior for a long time, and even if true underlying preferences were eventually to change.

[3] Similarly, Adena et al. (2015) and Voigtländer and Voth (2015) show that pre-World War II Nazi propaganda in Germany had a stronger effect in areas with strong pre-existing anti-semitic sentiment.

In addition to pointing to the importance of national stereotypes, our analysis also highlights the role played by the *strength* of the national stereotype for shaping aggregate responses to informational shocks such as the referendum. When the national stereotype is weak, regional responses to the shock cancel each other out, resulting in little or even no aggregate behavioral change. On the other hand, strong stereotypes will generate a strong aggregate reaction. As mentioned above, this is what, we argue, happened in the case of the Brexit referendum.

**Related Literature.** This paper is related to a number of different literatures. First, we contribute to the literature on on social attitudes. The standard neoclassical economics approach (Becker (1957)) takes the view that individuals are naturally endowed with an innate "*taste for discrimination*" and this fully determines their attitudes towards migrants. This hypothesis is corroborated by studies on the effects of ethnic diversity on economic and social outcomes, which typically show that higher ethnic diversity correlates with lower cohesion measured by social capital (Alesina and La Ferrara (2000), Alesina and La Ferrara (2002), Putnam (2007) and willingness to redistribute and invest in public goods (Alesina et al. (1999)). Algan et al. (2016) exploit a natural experiment of exogenous residential allocation in France and find that higher fractionalization leads to higher neglect and vandalism.

Another approach – emphasized initially by sociologists –sees people primarily as social actors. In this framework, the same individual may behave differently depending on the social context they find themselves in. Bursztyn et al. (2017) provide a recent example in support of this hypothesis. Their paper exploits the election of Donald Trump to US president to produce experimental evidence showing that the election result increased people's willingness to publicly express xenophobic views.[4] The authors show that their findings are consistent with an image-based model in which individuals want to signal that their preferences are aligned with those of the people around them (the "*observers*"). Our analysis complements their lab-experimental evidence by exploring widespread actual variation in attitudes. This allows us to measure the impact of spatial differences in shaping individual responses – an aspect that is central in our analysis, which is based on

[4]Arguably, this episode is similar to the Brexit referendum in many respects.

regional differences in objective measures such as hate crime. An additional contribution is that our theoretical framework emphasizes the role played by the desire to conform to the country-wide social norm and, in particular, the notion that people form expectations about the country-wide norm by partially extrapolating from their immediate surroundings. In turn, this feature generates geographical differences in the reaction to new public information about underlying preferences in the population. We clarify how geographic variations in the surprise effect – namely the difference between pre- and post-referendum beliefs about average preferences within the country – is key to explain the fact that the hike in hate crime was more pronounced in relatively more pro-immigrant areas.[5] More broadly, we contribute to the literature on the economics of norms and culture by documenting and analyzing a norm change, thus taking a road less traveled with respect to the large literature on cultural transmission and norm persistence (see e.g., the survey by Bisin and Verdier (2011), but also Fernandez (2007); Giuliano (2007); Alesina et al. (2013)).

We also connect with the literature studying how attitudes towards immigrants are formed and how they evolve over time and across geographical areas (Mayda (2006), Facchini and Mayda (2009) and Card et al. (2012)) and the nascent literature that focuses specifically on hate crime Bursztyn et al. (2019) and Müller and Schwarz (2019a, 2019b). We contribute to this literature by capitalizing on the unique episode of the Brexit referendum to identify the effect of new information on social norms governing behavior towards immigrants and minorities. Identifying whether the observed behavioral patterns reflect preferences or social norms clearly bears important implications for the design of policies and interventions.

The effect of the Brexit referendum on hate crime is the focus of two recent studies by Devine (2018) and Schilter (2018). Using aggregate UK data as well as data for Manchester and London, these studies confirm our finding of a post-referendum increment in hate crime. Our analysis adds to their work by uncovering a clear geographical pattern of the the phenomenon, which, through our theoretical analysis, allows us to shed light on the underlying mechanism at play.

[5] In the Appendix, we show that a direct application of a framework where individuals want to signal that their preferences are aligned with those of people in their vicinity would deliver the opposite result to the one observed in our data.

The paper is organized as follows. In section 2 we start by stating the main empirical facts about hate crime and the changes induced by the Brexit referendum. Section 3 presents and solves a model capable of reconciling these facts with theory. In section 4, we demonstrate that a quantitative version of the model can generate the sizes of the phenomena discussed with a *sensible* calibration. Section 5 concludes.

# 2 Empirics

The goal of this section is to establish four main facts on hate crime before and after the Brexit referendum that we observe in the data. The police force in England and Wales records a criminal offense in the category of "religiously or racially motivated hate crime" if it is perceived to be motivated by hostility or prejudice towards someone based on their race, ethnicity or religion. These crimes are defined by statute and will typically be subject, if prosecuted, to stricter sentencing than the equivalent crime, absent the racial or religious motivation. The list of criminal offenses included is reported in Table A.6. We turn now to discuss the data and uncover the main facts of this paper.

## 2.1 Data on Hate Crime

Data are publicly available and taken from the Office of National Statistics (ONS) and the Department for Work and Pensions (DWP). For brevity, we list each variable used in this section in Table 1. The data are in panel form, with the exception of the referendum result. The cross-sectional unit is a community support partnership (CSP) area. There are 315 of these in England and Wales. Data not associated with crime are reported at the local authority area, these can be aggregated up without ambiguity allowing a common cross-sectional unit across variables.

Included in our data are any criminal offense reported to the police and need not result in a later charge or prosecution. Clearly, this is not an exhaustive list of all hate crimes committed, as not all may be reported to the police. Based on estimates from the Crime Survey for England and Wales (CSEW), a victim survey conducted by the Office of National Statistics, Corcoran et al. (2015) estimate that approximately 48 per cent of all incidents of hate crime come to the attention

Table 1: List of Variables

| Variable | Frequency | Coverage | Source | Mean | SD |
|---|---|---|---|---|---|
| Hate Crime | quarterly | 2002q1 - 2017q2 | ONS | 29 | 39 |
| Total Crime | quarterly | 2002q1 - 2017q2 | ONS | $3.6 \times 10^3$ | $3.3 \times 10^3$ |
| Remain Votes (share) | cross-sectional | 2016q1 | ONS | 0.457 | 0.101 |
| National Insurance Registrations (EU) | quarterly | 2002q1 - 2017q2 | DWP | $2.7 \times 10^2$ | $4.6 \times 10^2$ |
| National Insurance Registrations (non-EU) | quarterly | 2002q1 - 2017q2 | DWP | $1.8 \times 10^2$ | $3.4 \times 10^2$ |
| Population (1000s) | annual | 2002 - 2016 | ONS | $1.8 \times 10^2$ | $1.2 \times 10^2$ |
| Gross Disposable Household Income (millions of £) | annual | 2002 - 2016 | ONS | $2.9 \times 10^3$ | $1.9 \times 10^3$ |
| Gross Value Added by Sector⋆ (millions of £) | annual | 2002 - 2016 | ONS | - | - |
| Social Benefits Received (millions of £) | annual | 2002 - 2016 | ONS | $7.7 \times 10^2$ | $5.3 \times 10^2$ |

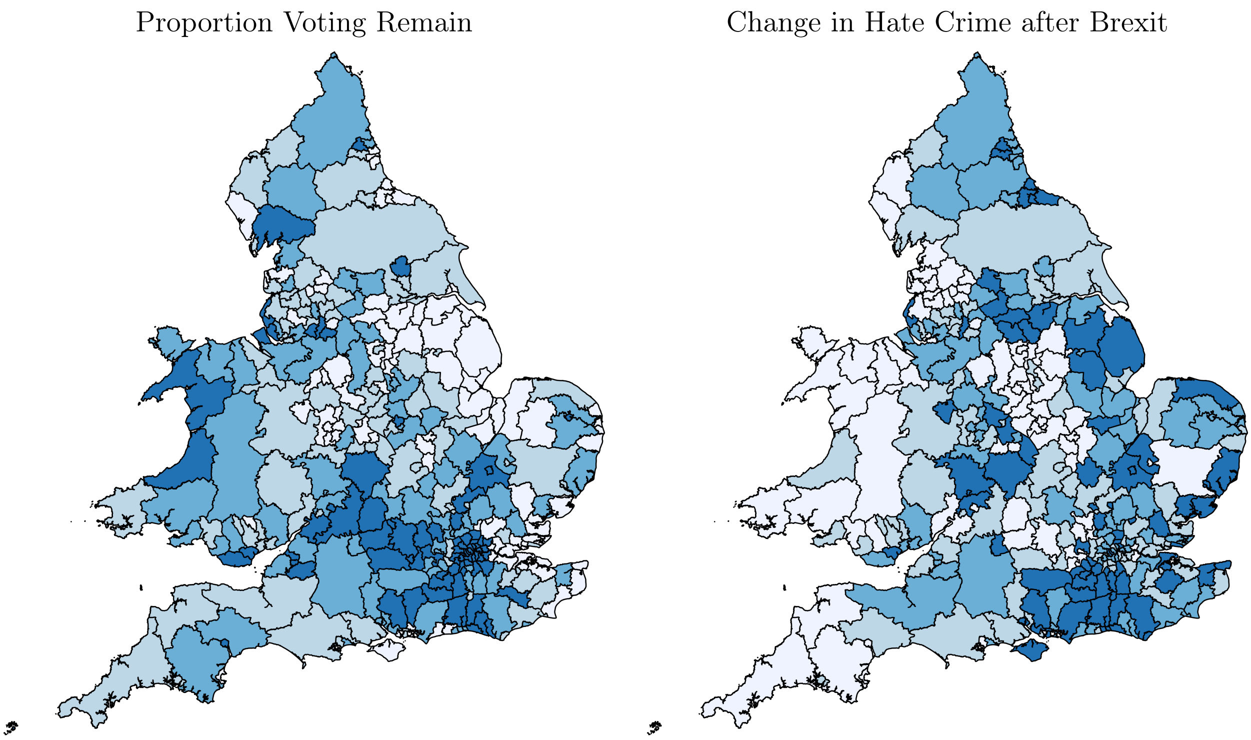
⋆ The economy is split up into ten sectors defined as: production; manufacturing; construction; distribution; information; finance; real estate; professional; public services; and other services.

of the police. Of course, our analysis has to to deal with the possibility of changes in the reporting patterns. We address this potential issue in Section 2.4.1. Total crime, the number of offenses reported in a CSP for a given financial quarter is retained to garner information relating to the overall criminality of an area.[6]

We define the remain share as the proportion of people in a given CSP who voted remain, as a proportion of eligible votes cast. The share of remain is lower than the overall result of the referendum (48.1%), primarily because of the omission of Northern Ireland and Scotland, who both had a majority voting remain. Further, it is smaller than 46.7% who voted remain in England and Wales, as the rural CSP areas have on average a smaller population and these more leave areas are thus over-weighted in our mean. Figure 2 shows the spatial distribution of the vote and how that corresponds to the percentage change in hate crime before and after Brexit in our sample. The figure is helpful to take a pithy first look at the principal empirical fact we wish to document, and a more systematic examination is provided in Section 2.3. Overall, the correlation between the two variables in Figure 2 is 0.16 indicating, not in any causal manner, that more remain areas, did on average, see a larger proportional increase in the prevalence of hate crime. To see this more clearly, Figure A.1 in the Appendix presents a scatter plot of the two panels of Figure 2.

[6] Please note in addition to the 43 regional police forces of England and Wales, criminal offenses can also come under the jurisdiction of the British Transport Police, the Civil Nuclear Constabulary or the Ministry of Defence Police, such crimes are disregarded in our analysis as cannot necessarily be geocoded, these account for a negligible share of total crime.

Figure 2: England and Wales by Vote Share and Dynamics in Hate Crime




**Note:** Change in hate crime is computed as $\log\left(E\left(\text{hate crime}|\text{post Brexit}\right)\right) - \log\left(E\left(\text{hate crime}|\text{pre Brexit}\right)\right)$. CSPs are split into quartiles on each metric and those in the darkest shades represent the largest number.

Turning to the other variables used in the analysis, these fall into two categories, relating to migration flows and economic indicators. We use administrative data provided by the Department for Work and Pensions that record the number of newly registered National Insurance Numbers. The National Insurance Number (NINO) is used in the administration of social security and the tax system and is a requirement to finding legal employment. The data we use records the number of NINOs issued to migrants in a given CSP area in a given financial quarter. We further distinguish between migrants coming from one of the other 27 EU countries and non-EU migrants. Population estimates are constructed using the 2001 and 2011 censuses and are updated annually by the ONS to account for births, deaths and migration flows.

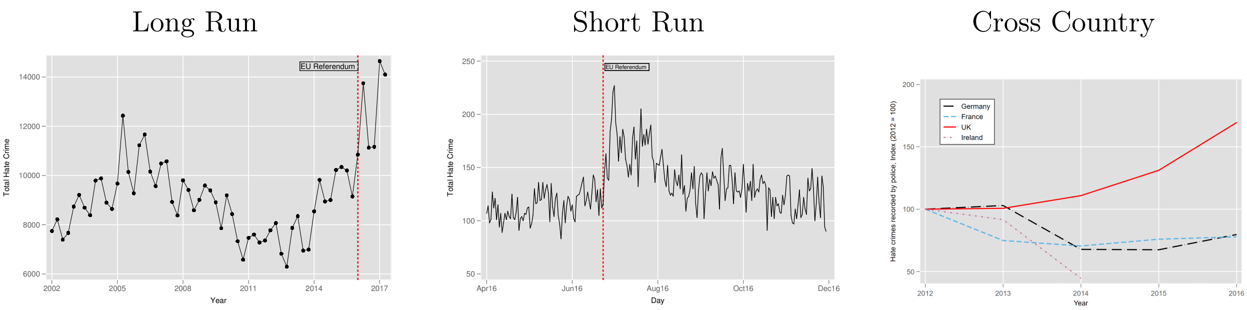
There is evidence that the economic conditions of a local area had an impact on the Brexit referendum (see Fetzer (2018) and Norris (2018)) and a large literature on the economic motivation of crime in general (see Freeman (1999) and the references therein). Therefore it is clearly important to control for changes in economic conditions. Variables we include at the CSP level are: gross disposable household income (GDI); the level of social benefits paid out; and the total value added produced by ten sectors that collectively constitute the whole economy. GDI is the amount of money that all of the individuals in the household sector have available for spending or saving. It is intended to measure the *material welfare* of the household sector at large. It is calculated by summing the primary and secondary incomes net of taxation of each household in a local area. Gross value added of an industry in a specific region are calculated via the income approach.[7]. The sector gross value added (GVA) is the sum of income generated by UK residents and corporations in the production of goods and services in a particular sector. It is measured gross of fixed capital and taxes, less subsidies. The particular CSP area is the area of economic activity rather than residence of the employees. Finally, social benefits received is the sum of government redistribution of income for a particular area, and includes for example, disability payments, state pension and job seeker's allowance.

## 2.2 Observation 1

In the aftermath of the Brexit vote, there were many claims from the UK media of a sharp rise in hate crime. Looking into these claims in a more systematic manner, we find that they are broadly borne out in the data and that, despite an international rise in populism, within the European context the increase in hate crime appears to be distinctive to the United Kingdom (see Figure 3).

[7] Details of which can be found in the UK National Accounts: Blue Book, 2016

Figure 3: Police Reported Hate Crime




**Note:** Data for the first two panels are provided by the Home Office and detail all crimes reported to the 43 police forces of England and Wales. The first panel gives a longer run interpretation of the trend. Hate crime is measured as instances of police reported hate crime per financial quarter. The second panel is a daily measure of reported hate crime and shows the short run impact of the Brexit referendum. Data used for the final panel are provided by the EU's Agency for Fundamental Rights (FRA) and the OSCE Office for Democratic Institutions and Human Rights (ODIHR) and made available at http;//hatecrime.osce.org/.

Figure 3 graphically documents the spike of hate crime following the Brexit referendum. This finding has already been well documented. For example, Devine (2018) provides a time-series analysis of the impact of the referendum on the aggregate level of hate crime. For completeness, the figure includes a shorter run trend at the daily frequency to show that the increase started at the date of the referendum, not before.

It seems appropriate to discuss the exception of Scotland. Although Scotland is not included in our sample, we were able to collect hate crime data for Scotland from separate sources. The data indicates that Scotland was the only region in the UK where the number of recorded racially and religiously motivated hate crimes actually fell after the Brexit referendum. According to the data collected by the Crown Office and Procurator Fiscal Service (COPFS), Scotland's prosecution service, the number of racially and religiously aggravated crimes in the period April 2015-April 2016 was 4315, while the equivalent for the period April 2016-April 2017 was 4022. The theoretical appendix discusses how this decline is actually consistent with our story. Intuitively, Scots have a strong separate identity from the rest of the UK, and it thus seems plausible that they should be more concerned with conforming to the the behavior of other Scots (the social norm in Scotland) rather than with the rest of the country. In Scotland, the Remain camp actually won (by 62%

against 38%), and hence, if anything, the information revealed by the referendum outcome about the underlying preferences of Scots goes in the opposite direction from the rest of the United Kingdom.

### 2.3 Observation 2

This section takes the increase in hate crime after the referendum for granted and focuses on documenting a novel observation, namely that this increase was most substantial in areas that voted remain. In order to be as transparent as possible we propose the following simple regression specification.

$$\text{hate}_{it} = \beta \left( \mathit{1}_{\{\text{Post Brexit}\}} \times \text{remain}_i \right) + \gamma \mathbf{X}_{it} + \tau_t + \eta_i + \epsilon_{it} \tag{1}$$

The dependent variable in equation (1) is the natural log of hate crimes reported in CSP area $i$ during financial quarter $t$. The first term on the right hand side contains a dummy variable taking the value one if Brexit falls in that financial quarter, or any subsequent quarter, multiplied by the share of the electorate in that CSP who voted remain in the referendum. In addition, included in the specification is a vector of time varying area specific controls ($\mathbf{X}_{it}$) which includes the variables described in Table 1 and police force-year fixed effects, year fixed effects ($\tau_t$) and CSP fixed effects ($\eta_i$). Implicit in this specification is the assumption that there is a constant elasticity of hate crime with respect to the referendum result. In section 2.4, we show this parameterization proves suitable. Results are presented in Table 2.

The specifications presented in the table progressively add more and more detailed controls, starting with measures of migration flows and population followed by economic indicators. Included in the economic indicators are the log value added of specific sectors in a CSP, the coefficients of which are presented in Appendix A.2. Subsequent columns include a proxy for the overall criminality of an area, a police force level trend component, and finally, to further control for inertia in hate crime, we include one lag of the dependent variable. The quasi elasticity of interest oscillates around a half, depending on the parameterization, implying that a one percentage point increase in the proportion voting remain in a given CSP increases the level of hate crime by half of one per

Table 2: Dependent Variable: Log Hate Crime

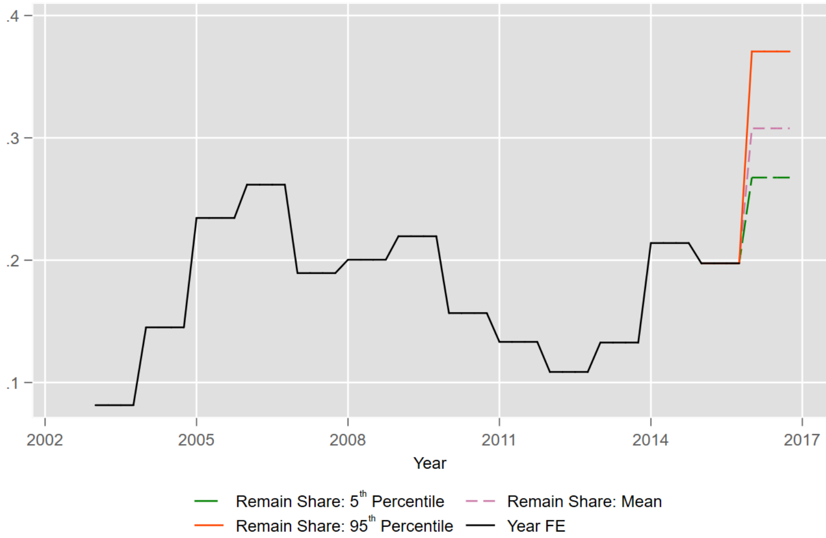
| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| Lag Dep. Variable | | | | | | 0.137*** |
| | | | | | | (0.007) |
| **Post Brexit ×** | **0.543*****  | **0.469***** | **0.318**** | **0.463***** | **0.661***** | **0.579***** |
| **Remain Share** | **(0.101)** | **(0.122)** | **(0.126)** | **(0.125)** | **(0.156)** | **(0.154)** |
| Log NiNo EU | | 0.017** | 0.016** | 0.015** | -0.004 | -0.004 |
| | | (0.008) | (0.008) | (0.007) | (0.007) | (0.008) |
| Log NiNo RoW | | -0.004 | -0.006 | -0.011 | 0.003 | -0.004 |
| | | (0.009) | (0.009) | (0.009) | (0.008) | (0.008) |
| Log Population | | 1.109*** | 0.777*** | 0.418** | -0.057 | -0.059 |
| | | (0.147) | (0.210) | (0.208) | (0.235) | (0.236) |
| Log GDI | | | 0.041 | 0.033 | 0.588*** | 0.572*** |
| | | | (0.138) | (0.136) | (0.164) | (0.165) |
| Log Social Benefits | | | 0.087 | 0.054 | -0.402** | -0.387** |
| | | | (0.115) | (0.114) | (0.159) | (0.158) |
| Log Other Crime | | | | 0.599*** | 0.512*** | 0.450*** |
| | | | | (0.027) | (0.030) | (0.030) |
| Observations | 19,530 | 18,900 | 18,900 | 18,900 | 18,900 | 18,585 |
| R-squared | 0.0277 | 0.509 | 0.506 | 0.670 | 0.885 | 0.888 |
| Number of CSPs | 315 | 315 | 315 | 315 | 315 | 315 |
| CSP FE | YES | YES | YES | YES | YES | YES |
| Season Dummies | YES | YES | YES | YES | YES | YES |
| Sectoral Composition | - | - | YES | YES | YES | YES |
| Year FE | YES | YES | YES | YES | - | - |
| Force-Year FE | - | - | - | - | YES | YES |

cent. While the estimates are statistically significant to visualise the economic significance we plot

$$\hat{\tau}_t + \hat{\beta}\left(\mathit{1}_{\{\text{Post Brexit}\}} \times \text{remain}_i\right)$$

as estimated in column (4). The change in the fixed effect in the first financial quarter of 2016 represents the aggregate impact of Brexit on hate crime, with 2002 as a baseline. Three hypothetical CSPs are plotted that are equivalent to the fifth percentile, the mean and the ninety fifth percentile. These voted 31.0%, 45.7% and 68.6% remain, respectively. Controlling for everything listed in column four, Brexit resulted in a more than ten per cent increase in hate crime for the average CSP. This increase is approximately equivalent to the gap between the fifth and ninety fifth percentile.

Figure 4: $\hat{\tau}_t + \hat{\beta}\left(\mathit{1}_{\{\text{Post Brexit}\}} \times \text{remain}_i\right)$




Although not the focus of our paper, there are three potentially interesting implications of the coefficients of the control variables. Firstly, the flow of new migrants into the country, either from the EU or outside, appears to have little effect on the level of hate crime. Across all six specifications the coefficients seem to be suggest a precisely estimated negligible impact. Secondly, public spending appears to reduce the amount of hate crime. This is reflected both in the negative effect of the social benefit expenditure in columns five and six and the negative coefficients on the GVA of public services. Finally, unsurprisingly, the level of hate crime is linked to the overall level of criminality in an area.

## 2.4 Robustness

To establish the robustness of our findings we perform three distinct exercises. (i) To show that our finding does not reflect a change in all crime that happened to coincide with the referendum, we perform the same analysis replacing hate crime with total crime, and this time we find no differential effects based on the EU referendum. Columns one to three in Table A.3 show the opposite relationship to hate crime, but with the inclusion of further controls the effect dissipates. (ii) We want to ensure this is not simply a London phenomenon, but concerns England and Wales as a whole. This is particularly important since Greater London comprises 15% of the total population of England and Wales and voted overwhelmingly to remain, almost 60% voted remain. Further, inspection of Figure 2 shows the CSP areas in London are geographically smaller than the average. Ignoring London therefore helps mitigate issues with commuter criminals - people living in one borough and committing hate crimes in another. Encouragingly our findings are robust to a reduced sample omitting London, see Table A.4. (iii) The final robustness check concerns the specification used. Implicit in the regression is an assumption that there is a linear relationship between the proportion voting remain and the proportional change in the level of hate crime. This assumption is evaluated in Appendix A.5 and the linear assumption seems viable. That said the elasticity is less at the extremes of the distribution. In other words, comparing CSPs who voted either overwhelmingly leave or remain will have less of a differential in the change in hate crime compared with comparing marginal voting CSPs.

### 2.4.1 Change in Reporting Rates

The data we have are '*reported*' crimes. It is therefore important that it is the mechanism described in our model that generates the phenomenon reported in this section - that is a change in '*actual*' hate crime. The Crime Survey for England and Wales (CSEW) is a victimization survey, which questions people whether they or a member of their household has been a victim of crime. Further, since the survey asks whether a given crime was reported to the police it is possible to compute the reporting rates for crimes, this is shown in Table 3, which is taken from a Home Office statistics bulletin.[8]

Inspection of Table 3 reveals two important phenomena to account for. Firstly, reporting rates for hate crime exceed that of other crimes by approximately 25%. As will be seen, there are two drivers of this. On the one hand, a hate component may increase the perceived severity of the

[8] Available online, https://www.gov.uk/government/statistics/hate-crime-england-and-wales-2017-to-2018, published 16$^{th}$ October 2018.

Table 3: Proportion of CSEW crime incidents reported to the police

| | 2007/08 to 2008/09 | | 2009/10 to 2011/12 | | 2012/13 to 2014/15 | | 2015/16 to 2017/18 | |
|---|---|---|---|---|---|---|---|---|
| | Percentage reported | Unweighted base | Percentage reported | Unweighted base | Percentage reported | Unweighted base | Percentage reported | Unweighted base |
| Hate crime | 51 | 516 | 49 | 666 | 48 | 409 | 53 | 377 |
| All crime | 39 | 24,935 | 39 | 34,314 | 40 | 20,718 | 40 | 17,019 |

crime, thus increasing the propensity to report said crime. In addition, the composition of crimes that are racially or religiously motivated are different from crimes in general. For example, they are more likely to involve violence against a person and less likely to involve theft or fraud, the latter having much lower reporting rates in general. The second thing to note is there is a potential upswing in reporting rates at the time of the Brexit referendum, which falls in the final period reported in the table. We would like a model of reporting propensity that takes into account both the compositional difference in crimes and a time varying component in the propensity to report hate crime in particular.

We assume the level of racially or religiously motivated crime of type $c$ in period $t$ is given by $h_{ct}$. The true level, reported or not, is unobservable and given by $h^{\star}_{ct}$. True and reported crime are linked by a constant factor $\omega_{ct}$ that varies by crime type and time, so that

$$h_{ct} = \omega_{ct} h^{\star}_{ct} \qquad \text{where} \qquad \omega_{ct} = \omega^0_t \omega^1_{ct}.$$

The share of hate crime that is reported to the police has two components, one that is specific to the particular crime, $\omega^1_{ct} \in [0, 1]$, and a hate crime reporting premium, $\omega^0_t \geq 0$. This latter term reflects whether the same crime is more or less likely to be reported, given that it was motivated by hate. To identify these weights, we use data from the CSEW on reporting rates of all crime of type $c$ in time $t$, which we label $\mathbf{c}_{ct}$, and the reporting rates of hate crime in aggregate reported in Table 3. The reporting rates can be identified by the equalities below, further details are provided in Appendix A.6.

$$\omega^1_{ct} = \frac{\mathbf{c}_{ct}}{\mathbf{c}^{\star}_{ct}} \quad \text{and} \quad \frac{\sum_c h_{ct}}{\sum_c h^{\star}_{ct}} = \omega^0_t \frac{\sum_c h_{ct}}{\sum_c \frac{h_{ct}}{\omega^1_{ct}}}$$

Our estimated reporting rates vary across time and space. These changes come from two sources:

firstly the estimated increase in propensity to report all hate crimes following the referendum; and secondly the ever changing composition of reported crimes. We run the same regressions as the previous section on the *true* level of hate crime and find, qualitatively, the same result. Results are reported in Table 4, and show, that across every specification, that remain areas exhibited a larger increase in hate crime than their leave counterparts. The same specifications are estimated, replacing log hate crime with the estimated proportions of hate crime reported to the police, results of which are reported in Table 5. The results reported in the first four columns seem to suggest that, after the referendum, an area with a ten percentage point larger remain share reported hate crimes with a greater propensity, somewhere between 0.12 and 0.18 percentage points, depending on the specification. However, when police force-time fixed effects are included, this increased propensity disappears. Our interpretation of this is that the increase in reporting reflects an increased propensity by the police forces to classify crimes as motivated by prejudice, rather than members of the public reporting hate crimes with greater frequency. As shown in Tables 2, 4 or 6 , including force-year fixed effects in our regression models of hate crime carries no effect on the significance or magnitude of the estimated effect of the Brexit referendum.

### 2.5 Observation 3

Examination of the first panel of Figure 3 shows a cyclical property to hate crime. The first peak corresponds to the 7/7 Islamic fundamentalist terrorist attacks on London in 2005, and the latter peak corresponds to the aftermath of the Brexit referendum. In Section 3, we discuss why we believe that the mechanism behind the latter (an information shock revealing existing views over immigrants) is quite different from the former (a shift in preferences due to increased mistrust/distaste of immigrants). Replicating the previous empirical approach to the context of the 7/7 attack we find no evidence that hate crime in remain areas increased more than their leave counterparts. In fact, the reverse seems to be true. Results presented in Table 6 show that across all specifications, leave areas increased more than their remain counterparts.

To cement this point further, we run a battery of placebo trials. Postulating the same specifica-

Table 4: Dependent Variable: Log Hate Crime (adjusted for under-reporting)

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| Lag Dep. Variable | | | | | | 0.122*** |
| | | | | | | (0.007) |
| **Post Brexit ×** | **0.558*****| **0.432***** | **0.298**** | **0.456***** | **0.659***** | **0.622***** |
| **Remain Share** | **(0.135)** | **(0.137)** | **(0.141)** | **(0.139)** | **(0.166)** | **(0.173)** |
| Log NiNo EU | | 0.018** | 0.017** | 0.015* | -0.008 | -0.002 |
| | | (0.008) | (0.008) | (0.008) | (0.009) | (0.008) |
| Log NiNo RoW | | -0.006 | -0.006 | -0.012 | 0.007 | -0.007 |
| | | (0.010) | (0.010) | (0.010) | (0.009) | (0.009) |
| Log GDI | | | -0.008 | -0.017 | 0.525*** | 0.531*** |
| | | | (0.154) | (0.152) | (0.169) | (0.186) |
| Log Population | | 1.146*** | 0.799*** | 0.410* | -0.011 | -0.026 |
| | | (0.163) | (0.234) | (0.232) | (0.241) | (0.266) |
| Log Social Benefits | | | 0.179 | 0.143 | -0.425*** | -0.426** |
| | | | (0.129) | (0.127) | (0.153) | (0.178) |
| Log Other Crime | | | | 0.651*** | 0.592*** | 0.501*** |
| | | | | (0.030) | (0.031) | (0.034) |
| Observations | 18,900 | 18,900 | 18,900 | 18,900 | 18,900 | 18,585 |
| R-squared | 0.839 | 0.839 | 0.840 | 0.844 | 0.852 | 0.877 |
| CSP FE | YES | YES | YES | YES | YES | YES |
| Season Dummies | YES | YES | YES | YES | YES | YES |
| Year FE | YES | YES | YES | YES | - | - |
| Force-Year FE | - | - | - | - | - | YES |

Table 5: Dependent Variable: Reporting Rates

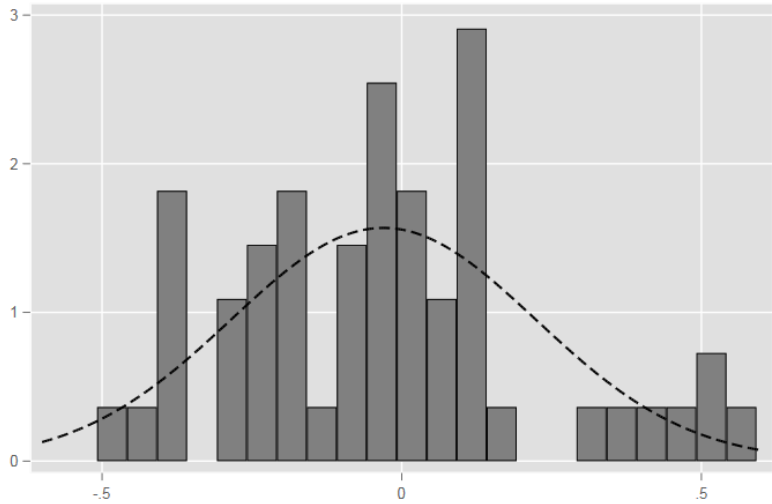
| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| Lag Dep. Variable | | | | | | 0.002 |
| | | | | | | (0.004) |
| **Post Brexit × Remain Share** | **0.012*** | **0.014**** | **0.018***** | **0.018***** | **0.005** | **0.005** |
| | **(0.006)** | **(0.006)** | **(0.007)** | **(0.007)** | **(0.009)** | **(0.009)** |
| Log NiNo EU | | -0.000 | -0.000 | -0.000 | -0.001 | -0.000 |
| | | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) |
| Log NiNo RoW | | 0.000 | 0.000 | 0.000 | 0.001* | 0.001 |
| | | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) |
| Log GDI | | | -0.001 | -0.001 | 0.001 | -0.000 |
| | | | (0.007) | (0.007) | (0.009) | (0.010) |
| Log Population | | -0.019** | -0.019* | -0.017 | -0.004 | -0.004 |
| | | (0.008) | (0.011) | (0.011) | (0.014) | (0.014) |
| Log Social Benefits | | | 0.005 | 0.005 | 0.004 | 0.005 |
| | | | (0.006) | (0.006) | (0.009) | (0.009) |
| Log Other Crime | | | | -0.002 | -0.001 | -0.001 |
| | | | | (0.002) | (0.002) | (0.002) |
| Observations | 18,641 | 18,641 | 18,641 | 18,641 | 18,641 | 18,169 |
| R-squared | 0.935 | 0.935 | 0.935 | 0.935 | 0.940 | 0.942 |
| CSP FE | YES | YES | YES | YES | YES | YES |
| Season Dummies | YES | YES | YES | YES | YES | YES |
| Year FE | YES | YES | YES | YES | - | - |
| Force-Year FE | - | - | - | - | YES | YES |

tion, column (6) at every possible time period in our sample. A histogram of the estimates of $\hat{\beta}$ are presented in Figure 5. Amongst the 50 estimated coefficients none are as large nor as significant as the 0.579 estimate associated with the Brexit referendum.

Table 6: Dependent Variable: Log Hate Crime

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| Lag Dep. Variable | | | | | | 0.138*** |
| | | | | | | (0.014) |
| Post 7/7 | 0.294*** | 0.250*** | 0.197*** | 0.172** | 0.192*** | 0.178*** |
| | (0.076) | (0.075) | (0.072) | (0.069) | (0.059) | (0.054) |
| **Post 7/7 ×** | **-0.494*** ** | **-0.396** ** | **-0.279* ** | **-0.232** | **-0.275** ** | **-0.251** ** |
| **Remain Share** | **(0.160)** | **(0.156)** | **(0.151)** | **(0.146)** | **(0.121)** | **(0.110)** |
| Log NiNo EU | | 0.014 | 0.014 | 0.013 | -0.006 | -0.006 |
| | | (0.013) | (0.013) | (0.013) | (0.010) | (0.010) |
| Log NiNo RoW | | -0.003 | -0.005 | -0.010 | 0.005 | -0.002 |
| | | (0.013) | (0.013) | (0.013) | (0.010) | (0.010) |
| Log Population | | 1.159*** | 0.772 | 0.432 | -0.036 | -0.043 |
| | | (0.417) | (0.543) | (0.548) | (0.512) | (0.439) |
| Log GDI | | | 0.071 | 0.075 | 0.609 | 0.590* |
| | | | (0.344) | (0.345) | (0.400) | (0.335) |
| Log Social Benefits | | | 0.080 | 0.032 | -0.414 | -0.395 |
| | | | (0.280) | (0.275) | (0.300) | (0.253) |
| Log Other Crime | | | | 0.593*** | 0.511*** | 0.449*** |
| | | | | (0.110) | (0.130) | (0.118) |
| Observations | 19,530 | 18,900 | 18,900 | 18,900 | 18,900 | 18,585 |
| R-squared | 0.0237 | 0.507 | 0.506 | 0.666 | 0.885 | 0.888 |
| Number of CSPs | 315 | 315 | 315 | 315 | 315 | 315 |
| CSP FE | YES | YES | YES | YES | YES | YES |
| Season Dummies | YES | YES | YES | YES | YES | YES |
| Sectoral Composition | - | - | YES | YES | YES | YES |
| Year FE | YES | YES | YES | YES | - | - |
| Force-Year FE | - | - | - | - | YES | YES |

Figure 5: A Battery of Placebo Trials




**Note:** The histogram plots the coefficient of interest from column (6) of Table 6 for 50 placebo events, grouped in equally spaced bins of 0.05. The solid line plots a normal distribution with mean and variance associated with the 50 estimates coefficients. The mean and standard deviation of the 50 coefficients are -0.03 and 0.25, respectively.

## 2.6 Observation 4

As we will discuss further in Section 3, our model is based upon the premise that an individual's beliefs about society as a whole are informed by their local environment. To test the validity of this premise in the context of the Brexit referendum, we use data from the British Election Study (BES), a long-running UK-wide panel study with approximately 30,000 respondents. In the four months leading up to the referendum, the BES study collected information about individual expectations on the referendum outcome. Respondents were asked: "*[h]ow likely do you think it is that the UK will vote to leave the EU?*" and gave an integer answer between zero and a hundred, zero implying that the UK would remain with certainty and a hundred that it would vote to leave with certainty. Additional survey questions about voting intentions, socio-economic factors and geographical location, and a time stamp denoting when the individual completed the survey, were also included. In order to test the effect of geographic location on expectations about the referendum outcome (after controlling for own voting intentions), we merged the referendum result at the Parliamentary constituency level with the BES data and performed the regressions shown in

Table 7. The coefficients of the additional controls are reported in Appendix A.7.

Table 7: Dependent Variable: Perceived likelihood of leaving the EU

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| Constituency Leave Share (%) | 0.084*** | 0.084*** | 0.075*** | 0.056*** |
| | (0.007) | (0.007) | (0.007) | (0.008) |
| Vote Intention: Leave | 17.027*** | 16.959*** | 16.807*** | 16.244*** |
| | (0.164) | (0.163) | (0.176) | (0.200) |
| Vote Intention: Will not vote | 1.776* | 1.818* | 2.243** | 2.125* |
| | (0.974) | (0.972) | (1.015) | (1.133) |
| Vote Intention: Don't know | 7.167*** | 7.333*** | 7.158*** | 6.868*** |
| | (0.389) | (0.388) | (0.408) | (0.453) |
| Observations | 54,916 | 54,916 | 49,341 | 41,317 |
| R-squared | 0.180 | 0.184 | 0.189 | 0.192 |
| Time Dummies | - | YES | YES | YES |
| Socio-Characteristics | - | - | YES | YES |
| Economic-Characteristics | - | - | - | YES |

Although small – a one standard deviation increase in the per cent voting leave within one's own constituency increases an individual's perceived probability of a leave camp victory by between 0.65 and 0.97 percentage points – the coefficient of interest remains statistically significant. We take this as evidence that, conditional on a battery of socio-economic controls and personal opinions, an individual's environment still dictates prior beliefs about country-wide views.

### Taking stock of the empirical observations

To recapitulate, let us organise the previous observations in the following way:

**Observation 1.** *The level of hate crime increased abruptly in the aftermath of the Brexit referendum.*

*(a) The most substantial increase is immediately after, not before, the vote.*

*(b) An exception is Scotland, which was the only police force area in the UK where the number of recorded racially and religiously motivated hate crimes fell after the referendum.*

**Observation 2.** *The rise in hate crime was particularly pronounced in areas that voted remain.*

(a) *No such pattern exists for crime as a whole.*

(b) *This result holds omitting London.*

**Observation 3.** *The rise in hate crime in the aftermath of the 7/7 terrorist attack was more pronounced in areas that voted leave in the referendum.*

**Observation 4.** *Before the referendum, people living in different areas of the country held different expectations about the referendum outcome. In more pro-Brexit areas, people expected higher country-wide support for Brexit.*

Taken together, Observations 1-4 constitute a challenge to any theory on the formation and change of attitudes towards racial minorities. In what follows, we develop a framework that can explain these facts with a minimal structure and inform counterfactual exercises about the effect of changes in preferences and perceptions of the social norm.

# 3 Model

In this section, we build a theoretical model which helps to fix ideas by formally deriving the implications that follow from a change in information vs a change in underlying preferences. The premise of the model is that individuals care about conforming about their underlying preferences (which arise from their socio-economics characteristics as well as other regional and idiosyncratic components) but also want to conform to the social norm – modelled as average behavior in society. Since the social norm is imperfectly observed, people need to form expectations about it from the information at their disposal.

## 3.1 The Environment

**Background** We consider a society of measure 1 which contains a continuum of individuals, represented by a real coordinate $i$ on the unit interval $[0,1]$. Each individual $i$ selects his action $a_i \in \mathbb{R}$ to maximize his expected payoff. This depends on (i) how closely his action matches a taste parameter $\alpha_i \in \mathbb{R}$ reflecting his intrinsic preferences, and (ii) how closely it matches average/mean behavior in society (the *social norm*). [Payoff is further discussed below.] Importantly, average behavior in society is not directly observable.

**Individual preferences over migrants** Individuals are divided into $n$ geographical districts of equal size (for simplicity). Individual preferences are given by the sum of two components: a district-specific component $P^d$, and an idiosyncratic component, $\varepsilon_i$. The district-specific component is in turn given by the sum of a mean preference parameter $\mu$ and a random element $e^d$, common to all individuals in district $d$. We can think of $e^d$ as capturing the effect of district-specific characteristics, while $\mu$ corresponds to mean preferences when specific characteristics are averaged out. To sum up, therefore, the taste parameter $\alpha_i^d$ of an individual $i$ belonging to district $d$ is equal to

$$\alpha_i^d = \mu + e^d + \varepsilon_i \tag{2}$$

where $\mu$, the (unobservable) mean preference in society, is drawn as $N(\overline{\mu}, \Theta)$, $e^d$ represents the district-specific shock to preferences, drawn as $N(0,1)$ with $e^{d_1} \perp e^{d_2}$, and $\varepsilon_i$, the idiosyncratic shock to preferences, is drawn as $N(0, \sigma)$, with $\varepsilon_i \perp \varepsilon_j$ for $i \neq j$, and $\varepsilon_i \perp e^d$ for any $i$ and $d$. The variable $\overline{\mu}$ can be thought as a common prior.

**Information** Each individual observes the mean preference within his district, $P^d = \mu + e^d$, but he is unable to discriminate between $\mu$ and $e^d$, the society–wide component and the district-specific shock affecting his preferences. Consider an individual $i$ who has observed $P^d$. Given the normality assumptions, his expectations of $\mu$ is given by linear regressions against all the relevant information available (see Morris and Shin (2002) or Angeletos and Pavan (2007)). Straightforward computations show that

$$E(\mu \mid P^d) = \frac{\Theta}{1+\Theta} P^d + \frac{1}{1+\Theta} \overline{\mu}. \tag{3}$$

**Payoffs** An individual's payoff is equal to

$$u_i = -\theta \left(a_i - \alpha_i\right)^2 - (1-\theta)\left(a_i - \overline{a}\right)^2. \tag{4}$$

The parameter $\theta \in (0,1)$ captures the concern of individuals for aligning their actions with their taste parameter (their *fundamental motive*) relative to their concern for coordination with average behavior in society (their *coordination or conformity motive*). For simplicity, our main model considers the case where the relevant social norm people want to adhere to is given by average behavior in the whole country. In the Theoretical Appendix B, we generalize our results by explicitly considering the case where the relevant norm as a weighted average of mean behavior in society and mean behavior in an individual's district.[9]

### 3.2 The Equilibrium

Each individual $i$ chooses his action $a_i$ to maximize his expectation of (4), where the expectation is taken with respect to $\overline{a}$, average behavior in society. Differentiation of the objective function delivers $i$'s best-reply action:

$$a_i = \theta\alpha_i + (1-\theta)\, E\left(\overline{a}\right) \tag{5}$$

As is standard in the literature, we will focus on linear symmetric strategies, where each player's action is a weighted average of the information at his disposal.

**Proposition 1.** *When $\mu$ is unobservable, the unique linear symmetric equilibrium of the game is given by*

$$a_i = \theta\alpha_i + \gamma P^d + (1-\theta-\gamma)\overline{\mu} \tag{6}$$

*where $\gamma \equiv \frac{n\theta\Theta(1-\theta)}{n(1+\theta\Theta)-(1-\theta)}$.*

A few remarks are in order. First, the parameter $\gamma$ is increasing in $\Theta$, the variance of the true preference mean $\mu$ around its prior $\overline{\mu}$, and decreasing in $n$, the number of districts. This reflects the fact that a higher $\Theta$ and a lower $n$ make $\overline{\mu}$ a less reliable predictor of $\mu$ (compared to $P^d$). Second, note that $\gamma + \theta$ is always increasing in $\theta$. Again, this is intuitive. The more people care about their personal preferences, the lower the weight they will place on the prior $\overline{\mu}$ when selecting their behavior.

We now compare the equilibrium described in Proposition 1 with the equilibrium that obtains in an alternative scenario, in which $\mu$ is publicly observable.

**Proposition 2.** *When $\mu$ is publicly observable, the unique linear symmetric equilibrium of the game is given by*

$$a_i = \theta\alpha_i + (1-\theta)\mu. \tag{7}$$

[9] As we argue in the Appendix, allowing for this features helps to rationalize the Scottish "*anomaly*" identified in our data.

Proofs for Proposition 1 and 2 are provided in Appendix B.2 and B.3, respectively.

### 3.3 The referendum as an information shock

A key feature of the referendum is that, by itself, it had no immediate impact on legislation and policies. In spite of this, it had a measurable effect on behavior and publicly expressed attitudes, as exemplified by the spike in hate-crimes which took place in the immediate aftermath of the vote. These changes cannot be explained by variations in policies or other economically-relevant variables, and, therefore, must be attributed to alternative explanations. We believe that a key feature of the referendum is that it revealed information about the electorate's private views, above and beyond previous information available. As a first approximation, the Brexit vote can thus be seen as an exogenous information shock that made mean preferences (in our model, $\mu$) publicly observable, as in the analysis of Proposition 2. In contrast, we can think of the pre-referendum environment as one where, as in Proposition 1, mean preferences are unobserved, although people have access to publicly available information on it (captured by the common prior $\overline{\mu}$).[10]

Consider first society as a whole. For simplicity, suppose that there are a continuum of districts. From Propositions 1 and 2, the change in average attitudes towards immigrants before and after the referendum is,

$$\begin{aligned} \overline{a}_{after} - \overline{a}_{before} &= \Delta\overline{a} = \theta\mu + (1-\theta)\mu - (\theta+\gamma)\,\mu - (1-\theta-\gamma)\overline{\mu} \\ &= (1-\theta-\gamma)\,(\mu-\overline{\mu})\,. \end{aligned} \tag{8}$$

Letting $a$ denote anti-immigrant attitude (with greater values corresponding to a more anti-immigrant attitude), a post-referendum increase in negative attitudes corresponds to $\Delta\overline{a} > 0$.[11] This leads to,

**Result 1.** *The necessary and sufficient condition for the referendum to induce mean attitudes to become more anti-immigrant is that,*

$$\mu - \overline{\mu} > 0 \tag{9}$$

[10] Results extend to the case where the referendum simply provides *more* information about $\mu$ than was previously available, without perfectly revealing it.

[11] Clearly enough, the same would hold if we let $a$ denote *pro*-immigrants attitudes. In that case, the increase in negative attitudes towards immigrants would imply $\Delta\overline{a} < 0$ and, hence, $\mu < \overline{\mu}$, indicating that true mean preferences are less pro-immigrants than what previously thought.

The condition implies that anti-immigrant attitudes rise if the true mean preferences (revealed through the referendum) are more anti-immigrant than the ex-ante prior. This result addresses Observation 1 through the lens of our theoretical model. It is important to realize that, in our account, behavior changes in spite of underlying preferences remaining *the same*. Intuitively, because of imperfect information, mean behavior before the referendum did not fully reflect true preferences.[12] People people adopted attitudes that were more tolerant towards foreigners than their true sentiment, for fear of behaving in a "politically incorrect" fashion. The referendum changed this. From (8) the magnitude of the change in attitudes depends on the size of $1 - \theta - \gamma$, which, after substituting for $\gamma$, can be shown to be decreasing in $\theta$ and thus increasing in the conformity motive – and equal to zero if the conformity motive is absent (i.e., $\theta = 1$). Intuitively, the stronger the conformity motive, the higher the amount of pre-referendum "*hypocrisy*", and thus the stronger the effect of the referendum on behavior. Finally, it is worth stressing that, here, the new information that people react to is information about aggregate preferences (revealed through the vote), *not* information about the merits of leaving the EU and curbing immigration. Hence, the model predicts that the change in attitudes should occur only *after the vote,* and should be abrupt rather than gradual – to reflect the abrupt nature of the referendum-induced information release. This corresponds to our Observation 1a), and, as we discuss below, stands in contrast with the predictions we would obtain if the rise in hate crime resulted from a shift in preferences due to exposure to campaign rhetoric.

The role of new information is thus crucial to appreciate the mechanism behind the change in attitudes. Indeed, condition (9) makes clear that the effect of the referendum on attitudes relies on what we may call a *surprise effect*. Formally, the surprise effect can be measured as the difference between true $\mu$ and mean pre-referendum beliefs about $\mu$, which can be easily computed from Bayesian updating,

$$\mu - E[E(\mu \mid P^d)] = \frac{\mu - \overline{\mu}}{1 + \Theta}. \tag{10}$$

[12] This shares similarities with the phenomenon known in social psychology as *pluralistic ignorance*, namely a situation in which the majority privately disagrees with a given behavior, but incorrectly assume that most others agree with it, and therefore adopts it (Katz et al., 1931). Formally, the result follows from the fact that $\overline{a}_{after} = \mu$ while $\overline{a}_{before} = (\theta + \gamma)\,\mu + (1 - \theta - \gamma)\overline{\mu} < \mu$, as shown in (8).

Expressions (8) and (37) show that there is a direct correspondence between the surprise effect and the change in mean behavior induced by the referendum. Intuitively, the referendum changed behavior only to the extent to which the information it revealed was unexpected. The following result describes the variations in pre-referendum beliefs across geographical areas.

**Result 2.** (Variations in beliefs) *Average beliefs about $\mu$ in district $d$ prior to the referendum were*

$$E(\mu \mid P^d) = \frac{\Theta}{1+\Theta} P^d + \frac{1}{1+\Theta} \overline{\mu}$$

*increasing in $P^d$.*

**Proof:** *Follows straightforwardly from Bayesian updating.*

Result 2 rationalizes our Observation 4, namely that, prior to the referendum, people in Leave areas believed it more likely that the UK would vote leave than people in Remain areas. That's because standard Bayesian updating implies that people used preferences in their own area to form beliefs about average preferences across the country – a sort of "rational consensus effect".[13] Result 2 has a direct implication for geographical variations of the surprise effect. The difference between true $\mu$ and previous average beliefs about $\mu$ in district $d$ is

$$\mu - E_d[E(\mu \mid P^d)] = \mu - \frac{\Theta}{1+\Theta} P^d - \frac{1}{1+\Theta} \overline{\mu} \tag{11}$$

decreasing in $P^d$. People in pro-immigrants districts underestimated more strongly the true extent of anti-immigrant sentiment across the UK. The behavioral implications of the geographical variations in the surprise effect follow immediately from Propositions 1 and 2,

**Result 3.** (Variations in behavioral changes) *The difference in anti-immigrants attitudes in district $d$ before and after the referendum is,*

$$\Delta \overline{a}^d = (1-\theta)\,(\mu - \overline{\mu}) + \gamma(\overline{\mu} - P^d)$$

*decreasing in $P^d$.*

[13] The false consensus effect is a well known concept in psychology, which refers to the tendency of people to overestimate the extent to which their opinions or preferences are normal and typical of those of others. Our analysis shows that, when mean preferences are not observed, using own preferences (or preferences in one's own area, as in our model) to predict the preferences of other individuals is actually perfectly consistent with Bayesian updating – see also Vanberg (2019) and Adriani and Sonderegger (2015) for other illustrations of this general point.

In words, the model predicts that areas with stronger anti-immigrant animus experienced a *smaller* change in attitudes following the referendum, in line with our Observation 2. The intuition for this is, again, transparent. People in more anti-immigrant areas were less surprised by the outcome of the referendum (and by what it revealed about average preferences across the country). As a result, they did not need to adjust their behavior as much in response to the new information released through the referendum.

This last result highlights the importance of introducing in our model imperfectly observable social norms to understand observed changes in attitudes. An important question in this respect concerns the relevant social norm individuals want to conform to. In a number of contexts, this might be based on the attitudes of individuals we interact with, rather than being associated with a higher geographical/geopolitical level as in the case of national norms. The possibility that people may care about conforming with the local (rather than country-wide) norm is explored in Appendix B.1. We show that, in that case, the implied prediction on the effect of the Brexit referendum would contradict Observation 2, namely the finding that post-referendum behavioral responses were smaller in areas with stronger dislike towards immigrants. In Appendix B.5, we investigate the related idea that people may want to adopt a behavior that signals to those around them that their preferences are similar to theirs, and argue that this model would also fail to predict Observation 2.

### 3.4 Behavioral effect of a shift in preferences

A key characteristic of the referendum is that the vote revealed new public information about aggregate preferences, and behavior adapted as a result of this new information. Here we take the theory further, to explore the effects of an alternative source of change in attitudes, namely a shift in preferences. Many commentators have argued that the media campaign that surrounded the referendum could have fuelled negative feelings against immigrants in the population, thus changing preferences. Other possible reasons for preference changes include tragic events such as terrorist attacks on national soil perpetrated by foreign nationals or individuals of foreign origin. It is thus instructive to analyze what would be the theoretical implications of such a shift. Let $\mu_{before}$ and

$\mu_{after}$ denote mean preferences across the country before and after the preference shift and let $\overline{\mu}_{before}$ and $\overline{\mu}_{after}$ reflect publicly available information before and after. By assumption, $\mu_{before} \neq \mu_{after}$, while $\overline{\mu}_{before}$ and $\overline{\mu}_{after}$ may or may not differ depending on the exact situation at hand. The change in aggregate attitudes across the country can be easily computed from (6),[14]

$$\Delta\overline{a} = (\theta + \gamma)\,(\mu_{after} - \mu_{before}) + (1 - \theta - \gamma)(\overline{\mu}_{after} - \overline{\mu}_{before}). \tag{12}$$

Hence, a shift in preferences caused by the referendum in consistent with an increment in anti-immigrant attitudes and thus an increase in hate crimes (Conjecture 1) provided that $\mu_{after} > \mu_{before}$ and/or $\overline{\mu}_{after} > \overline{\mu}_{before}$. Looking at geographical differences, it is easy to show that,

**Result 4.** *The change in aggregate attitudes in district $d$ following a shift in preferences is equal to,*

$$\Delta\overline{a}^d = (\theta + \gamma)\,(P^d_{after} - P^d_{before}) + (1 - \theta - \gamma)(\overline{\mu}_{after} - \overline{\mu}_{before}). \tag{13}$$

Result 4 suggests that the change in attitudes should be more pronounced in those areas that experienced a bigger shift in preferences ($P^d_{after} - P^d_{before}$ is larger). Whether these correspond to areas that were more or less pro-immigrants before the event is open to debate. On the one hand, it is possible that anti-immigrant rhetoric may be more likely to find fertile ground in areas where people hold somewhat anti-immigrant views already. On the other hand, one could also envisage the opposite preference shift as a result of media exposure or following a terrorist attack, since in low prejudice areas there is more scope for people to change their views.

To shed light on this question, we can exploit Observation 3, which looks at hate crime in the aftermath of the 2005 London bombings, often referred to as 7/7. Intuitively, a terrorist attack is likely to generate a direct shift in people's preferences, increasing their distaste for foreigners/immigrants. The pattern of increased hate crime following the 7/7 episode may thus provide useful evidence on how a shift in preferences may affect behavior in different areas. Observation 3 shows shows the increment in hate crime was more pronounced in leave areas, namely the opposite pattern than what we observed in the aftermath of the Brexit vote. This suggests that the attack triggered a bigger

[14] Again, we are assuming a continuum of districts for ease of exposition but the argument carries through more generally.

shift in preferences (and thus behavior) in those areas that were already anti-immigrant to start with.[15] This is consistent with recent literature such as Adena et al. (2015) and Voigtländer and Voth (2015) who show that Nazi antisemitic rhetoric in pre-World War II Germany had a stronger impact in areas with existing antisemitic sentiment. Bursztyn et al. (2019) and Müller and Schwarz (2019b) similarly find that exposure to social media has a stronger effect on xenophobic hate crime in more nationalistic areas. Overall, this evidence casts doubt on the idea that the increment in hate crime following the Brexit referendum may have been caused by a shift in preferences fueled by media debate – since in that case Observation 2 would be contradicted. Moreover, if the rise in hate crime had been due to changed preferences, then we should have observed a *gradual* increment, prompted by the debate, and starting some time *before* the referendum. This would however contradict Observation 1, which points to a sudden and abrupt change. We find this argument suggestive that an information shock explanation of Section 3.3 is a more plausible explanation of the evidence than a preference shift.

# 4 Quantitative exercise

In this section, we estimate the primitive parameters of our baseline model to match empirical moments computed in the data. This exercise serves two purposes. First, it allows for a normative evaluation of whether the rise in hate crime can credibly be explained by our mechanism. We argue that the parameter estimates required to match the data are plausible and in the broad ball park of similar estimates made in a very different context. Another advantage of the analysis is that it sheds further light on the specifics of the impact of the referendum as a revelation of society's true preferences. After estimating the distribution of individual beliefs about mean preferences, we find that, pre-referendum, these ascribed very small probability to true mean preferences corresponding to what was subsequently revealed through the referendum. This suggests that the British public

[15] Our measure of prejudice is the share of the leave vote in the referendum, namely an event that took place several years after 7/7. The underlying assumption is one of sufficient persistence anti-immigrant preferences across time. This is corroborated by existing studies such as Becker and Fetzer (2016) and Becker et al. (2017) , who document a strong correlation between historical regional support for anti-immigrant populist parties and the share of leave vote in the Brexit referendum.

were *blissfully unaware* of the latent prejudice present in their midst.

### 4.1 Paramaterization: linking the model with our observables

To take the model to the data we need to make three refinements to the model just presented. Firstly, since our data is at the CSP level, we consider this to be a district in the language of the model and abstract away from any heterogeneity within each district. Second, we make parametric assumptions about the functional form of attitudes and how they translate into hate crime. Finally, we embed a very simple voting rule as a function of preferences. The model is as before two periods. We consider the financial year prior to the Brexit referendum and the subsequent year and index the periods by $t \in \{0, 1\}$.

**Attitudes.** In order to link explicitly our model of individual attitudes with our data, we need to take a stance of how attitudes translate into hate crime. Consider a setup where the number of normalized hate crimes is given by,[16]

$$h = f(a) \tag{14}$$

for some function $f(\cdot)$ that translates attitude into criminal behavior. This is a reduced-form approach that focuses on attitudes, but it should be clear that, generally, the precise functional form taken by $f(.)$ will depend on a number of payoff-relevant factors such as probability of detection, penalty incurred in case of conviction, etc. For our purposes, however, it is sufficient to leave these elements in the background.

The underlying premise of our analysis is that attitudes that are more xenophobic will generally translate into higher incidence of hate crime, so $f(\cdot)$ should be strictly increasing. Moreover, according to our theory, $a$ is a normally distributed variable. Rearranging (14), we can write,

$$a = f^{-1}(h) \tag{15}$$

and, hence, following the theory, $f^{-1}(h)$ must be normal. Our data matches this assumption as the logarithm of normalized hate crime is approximately normally distributed, see Appendix C.2,

[16] Normalized hate crime is defined as the number of hate crime incidents divided by the number of foreign born individuals in an area.

suggesting that $f^{-1}(.) = \log(.)$ and, hence, $f(.) = \exp(.)$. This implies that the incidence of hate crime is a convex function of $a$ (attitudes towards immigrants). Intuitively, so long as mean attitudes are sufficiently favorable towards immigrants, the incidence of hate crime will be consistently low. Eventually, however, as attitudes towards immigrants keep worsening, hate crimes will start to rise noticeably, and further deterioration in attitudes will have increasingly large effects on hate crime.

**Voting behavior.** Our measure of attitudes in district $d$ in period $t \in \{0, 1\}$ are defined as $a_t^d$. We take the baseline model, the equilibrium of which is defined in Section 3.2, to the data. In addition to taking a stand on how attitudes manifest themselves, we also introduce a simple voting rule. Since the referendum was a secret ballot, we assume that aggregate voting in a district is motivated by the primitive district mean preferences $P^d$ and not by the outward attitude of its inhabitants, which is an endogenous outcome of the model. It is assumed that the proportion of residents in district $d$ voting for remain follows the linear relationship[17]

$$v^d = b_0 + b_1 P^d + \epsilon^d \quad \text{where, } \epsilon^d \sim N(0, \sigma_\epsilon^2). \tag{16}$$

We put no restrictions on the additional parameters $b_0$, $b_1$ and $\sigma_\epsilon$, other than non-negative variance. The relative value of $b_1$ and $\sigma_\epsilon$ determines the intrinsic importance of the primitive attitude towards immigrants in district's voting behavior. Notice, an implicit assumption in this voting rule is that while other factors can govern voting behavior, such factors are assumed orthogonal to preferences towards immigrants. While there is strong evidence that voting intention in the referendum were related to ones attitude towards immigrants (see Appendix C.1), this assumption is clearly simplistic and is made for tractability. In the model districts only vary by aggregate preferences towards migrants, while clearly in the data districts vary in a multitude of dimensions.

The final addition to the model is the introduction of a forecaster. Since the model's mechanism is through agent's beliefs, we include in the estimation a moment related to expectations. The forecaster in period 0 assigns a probability of remain winning the vote. The forecaster's information set contains all parameters of the model with the exception of the population's *true* primitive

[17]We follow a simplified version of the statistical model of district voting in the referendum of Auld and Linton (2019). We too assume a Gaussian distribution but abstract away from uncertainty over turnout.

preferences regarding migrants, $\mu$. Assuming a constant population in each district, the forecaster's prior distribution of remain votes is given by equation (17).

$$\frac{1}{n}\sum_{d} v^d \sim N\left(b_0 + b_1\overline{\mu}, \Theta + \frac{b_1^2 \mathrm{Var}_d(P^d) + \sigma_\epsilon^2}{n}\right). \tag{17}$$

Thus the forecaster will assign a probability $\pi$ of remain winning more than a share S of the votes as

$$\pi = 1 - \Phi_v(S).$$

Where $\Phi_v(\cdot)$ is the associate cumulative distribution function defined in equation (17). In practice we specify $S$ as the share required in England and Wales for remain to win overall. We specify $S < 0.5$ as Northern Ireland and Scotland are omitted from our sample and voted predominantly remain. We therefore assume that the forecaster has zero uncertainty regarding these two regions, since they are relatively small compared to England and Wales (results will not change much quantitatively under alternative assumptions). Our empirical counterpart to this moment comes from the betting market the day of the referendum. Appendix C.3 shows the odds and associated implied probabilities of the referendum result for 18 online British bookmakers. The probability of a remain victory on the day of the referendum ranged from 83% to 87% depending on the bookmaker. We take the mean of the sample 0.86 as our targeted moment. .

## 4.2 Estimation

In the theory section we normalized the variation in $P^d$ across district to unity, here we specify $\mathrm{Var}_d(P^d) = \sigma_d^2$. Giving us the vector $\vec{\theta} := (\theta, \Theta, \mu, \overline{\mu}, b_0, b_1, \sigma_d, \sigma_\epsilon)$ to estimate. The vector of structural parameters is estimated by simulating the model $M$ times and computing a vector of moments $m(\theta)$ each iteration. The simulated data is made to look as similar to the empirical data described before, for two periods with 314 districts.[18] The mean of these are matched to a vector of the same moments computed in the data, $\hat{m}$. These moments are listed in Table 8. Parameters

[18]One CSP, the Isles of Scilly, is removed since it reported no incidence of hate crime in a year.

are estimated minimizing the function below.

$$\hat{\vec{\theta}} = \arg\min_{\vec{\theta}} \left( \frac{1}{M} \sum_s m_s(\vec{\theta}) - \hat{m} \right) \Omega \left( \frac{1}{M} \sum_s m_s(\vec{\theta}) - \hat{m} \right)' \tag{18}$$

The moments used in estimation are listed in Table 8 along with the fit of the model. In Appendix C.4 we argue that the moments listed are enough to identify the parameters of the model. The weighting matrix $\Omega$ is set as $\text{diag}(m^{-2})$ to put equal weight to each moment, irrespective of a moment's magnitude.[19]

Table 8: Targeted Moments in Estimation

| | Simulated Moment | Empirical Moment | Source |
|---|---|---|---|
| **Pre-referendum** | | | |
| Mean attitude pre-referendum | -5.28 | -5.29 | ONS |
| Variance of attitudes pre-referendum | 0.44 | 0.45 | ONS |
| Probability of remain winning referendum | 0.86 | 0.86 | Betting markets Appendix C.3 |
| **Post-referendum** | | | |
| Mean attitude post-referendum | -5.07 | -5.07 | ONS |
| Variance of attitudes post-referendum | 0.44 | 0.42 | ONS |
| Covariance of attitudes post-referendum and vote shares | -0.02 | -0.02 | ONS |
| Mean referendum result (remain share) | 0.46 | 0.46 | ONS |
| Variance of the referendum result | 0.01 | 0.01 | ONS |
| Covariance of attitudes post and pre-referendum | -0.02 | -0.02 | ONS |

**Note:** Attitudes are measured as the log of hate crimes committed in a year divided by the foreign born population in a region. Like the theory a higher number implies a worse attitude towards migrants. All means and variances are across voting district.

The estimated parameters are presented in Table 9. The desire to conform $1 - \theta$ is arguably small. Personal preferences determine approximately 77% of individual behavior, while the desire to conform to average attitude in society determine the remaining 23%. We are not aware of any

[19]Since moments are computed from multiple data sources it is not clear how to compute an *optimal* weighting matrix. Setting $\Omega = \text{diag}(m^{-2})$ ensures that the weight applied to each moment is invariant to its magnitude. Any positive semi-definite matrix will yield unbiased estimates but precision could be improved with an alternative weighting matrix, Gourieroux et al. (1993). Further $M$ is set to one thousand to ensure stability of the simulated moments.

other paper that provides a direct measure of individual conformity concerns. The closest evidence we could find to put our results in context comes from the experimental literature on norms. In an influential paper, Krupka and Weber (2013) estimated the weight that individuals put on social appropriateness when selecting their action in the dictator game – an experiment in which subjects must decide how to split a cash prize between themselves and another anonymous participant. Using their own data as well as data from different variations of the dictator game collected by Lazear et al. (2012) and and List (2007), Krupka and Weber (2013) estimate that social appropriateness determines 57% to 74% of individual behavior. Our smaller figure of 23% can be reconciled with these estimates once we bear in mind the sharp difference in contexts (hate crime vs generosity), as well as the nature of the parameters being estimated (concern for conforming with average empirical behavior vs concern for taking socially appropriate actions). Further, it is likely that in reality the structural parameters of the model, and in particular the desire to conform, are not constant across individual. In fact there is some evidence suggesting consequential dispersion. see Wilcox (2006). Although beyond the scope of this paper it would be interesting to use a fully-fledged structural model to accommodate such heterogeneity and evaluate its effects.

Table 9: Parameter Estimates

| $\Theta$ | $\mu$ | $\overline{\mu}$ | $b_0$ | $b_1$ | $\sigma_d$ | $\sigma_\epsilon$ | $\theta$ |
|---|---|---|---|---|---|---|---|
| 0.01 | −5.07 | −5.97 | 0.29 | −0.03 | 0.86 | 0.10 | 0.77 |
| (0.01) | (0.04) | (0.84) | (0.03) | (0.01) | (0.10) | (0.005) | (0.09) |

**Note:** Standard errors are computed by resampling the data with repetition and re-estimating the model on each resample (1000 times). Note, the moment taken from the betting odds is fixed in each resample.

The table suggests that the shift in beliefs from prior to posterior was large, as measured by the relative size of $\mu - \overline{\mu}$ to the variance of the prior $\Theta$. As we discuss further below, $1/\Theta$ can be interpreted as a measure of the strength of the ex-ante national stereotype $\overline{\mu}$. Our quantitative model is thus indicative of a strongly held stereotype which was then dispelled by the information shock provided by the referendum outcome. In this respect, the referendum result really was a "*bolt from the blue*" and changed everyone's perspectives about British views on immigrants considerably. The remaining parameters, the $b$'s and $\sigma$'s relate to the dispersion of attitudes across area and the exogenous voting rule (equation (16)). A positive estimate of $b_0$ and a negative $b_1$ implies that the higher the propensity of hate crime of an area the less likely the area is to vote remain. Finally, the large value of $\sigma_d$ relative to $\sigma_\epsilon$ is encouraging. It suggests the latent attitude of an area, under our simple voting rule, explains the majority of the variation in the remain share of the vote. We now further explore the implications of our calibration results by delving into three thought experiments.

**Beliefs vs true preferences.** Consider two individuals, $i$ and $j$, who hold the *same beliefs* about the dominant social norm. From our estimates, the difference in the behaviors adopted by $i$ and $j$ is

$$a_i - a_j = 0.77(\alpha_i - \alpha_j),$$

approximately 3/4 of the difference in their underlying preferences. In other words, the desire to conform with society at large induces these two individuals to adopt behaviors that are closer than if they simply followed their own inclinations.

Similarly, suppose that an individual's personal preferences change but her beliefs about the norm remain the same. Rather than fully reflecting her changed preferences, the shift in this individual's behavior will correspond to approximately 3/4 of the underlying preference change. Vice-versa, if an individual's preferences remain unchanged but her beliefs about the dominant norm change by $\Delta E(\overline{a})$, this will induce a behavioral shift corresponding to approximately $\frac{1}{4}\Delta E(\overline{a})$.

**The role of stereotypes and shared narratives.** Consider now two different societies that are characterized by different priors $\overline{\mu}_0$ and $\overline{\mu}_1$. Even if the realized mean preferences $\mu$ are *identical*, attitudes in these two societies may differ in a non-negligible way. From our previous analysis, average attitude is equal to

$$\overline{a} = (\theta + \gamma)\mu + (1 - \theta - \gamma)\overline{\mu}. \tag{19}$$

After substituting for $\theta$ and $\gamma$ (approximating $n \to \infty$), the difference in average attitudes between the two societies is equal to

$$\overline{a}_1 - \overline{a}_0 = 0.228(\overline{\mu}_1 - \overline{\mu}_0).$$

This observation underscores the importance of stereotypes in shaping collective attitudes. Although our analysis is deliberately vague about the origins of the prior $\overline{\mu}$, it is clear that beliefs about preferences in a society (as well as preferences themselves) are strongly influenced by history, culture and country-level identity. Our analysis suggests that, through their effect on beliefs, these elements may create a wedge between *true* underlying views and behavior. A strongly pro-immigrant national identity, for instance, will act as a mitigating force for latent xenophobia. In turn, this creates a role for policy interventions aimed at shaping public perception of attitudes. A case in point is the reaction of New Zealand's prime minister Jacinda Ardern in the wake of the 2019 Christchurch mosque shootings, aimed at protecting and preserving a strong sense of national identity centered around values of tolerance and inter-cultural respect.

**Stereotype strength.** Consider now two societies with the same prior $\overline{\mu}$ and the same realized mean preferences $\mu$, but different values of $\Theta$, $\Theta_0$ and $\Theta_1 > \Theta_0$. Recall that $\Theta$ is the variance of the distribution from which mean preferences are drawn. Intuitively, we can think of $1/\Theta$ as a measure of the strength of the ex-ante stereotype $\overline{\mu}$.

From (19), the difference in average attitudes in the two societies is

$$\overline{a}_1 - \overline{a}_0 = (\gamma_1 - \gamma_0)(\mu - \overline{\mu}) \tag{20}$$

where (assuming $n \to \infty$ in both societies for simplicity),

$$\gamma_1 - \gamma_0 = \frac{(1-\theta)\theta(\Theta_1 - \Theta_0)}{\theta(\Theta_1 + 1)(\theta\Theta_0 + 1)} > 0.$$

A number of remarks are in order. First, note that, from (20), if $\overline{\mu} = \mu$, then $\overline{a}_1 = \overline{a}_0$. If true preferences coincide with the ex-ante stereotype, mean behavior is the same in both societies, in spite of different stereotype strength. Second, if the stereotype is somewhat incorrect so that $\mu \neq \overline{\mu}$, then behavior *does* differ in the two societies. If $\mu > \overline{\mu}$ (i.e., true preferences are *more* anti-immigrant than the stereotype), then $\overline{a}_1 > \overline{a}_0$, and vice-versa, if $\mu < \overline{\mu}$ (true preferences are *less* anti-immigrant than the stereotype), then $\overline{a}_1 < \overline{a}_0$. If $\Theta_0 \to 0$ (very strong prior/stereotype) and $\Theta_1 \to \infty$ (diffuse prior), then the difference in behavior across the two societies is

$$\overline{a}_1 - \overline{a}_0 = (1-\theta)(\mu - \overline{\mu}) = 0.23(\mu - \overline{\mu}).$$

i.e. approximately $1/4$ of the underlying discrepancy between stereotype and true mean preferences. Note that, in the society characterized by a stronger stereotype (society 0), mean behavior is always further away from true preferences than in the society with a weaker stereotype (society 1).[20] In other words, the presence of a strong stereotype mitigates anti- immigrant attitudes when true preferences are *more* anti-immigrant than the stereotype, but acts as a countervailing force when true preferences are *less* anti-immigrant than the prior.

The strength of the ex-ante stereotype also has an impact on the surprise effect and the resulting behavioral change following a sudden information shock such as the Brexit referendum. From expression (8), the change in behavior triggered by the information shock is

$$\overline{a}_{after} - \overline{a}_{before} = (1 - \theta - \gamma)\,(\mu - \overline{\mu})$$

where $\gamma \equiv \frac{n\theta\Theta(1-\theta)}{n(1+\theta\Theta)-(1-\theta)}$, increasing in $\Theta$. When $\Theta \to \infty$, i.e. the ex-ante stereotype is extremely weak, $\gamma \to 1-\theta$, which implies that the information shock generates no change in aggregate behavior. That is because, in that case, the behavioral adjustments occurring in different areas of the country fully cancel each other out. People located in areas which turn out (once average preferences in the country are revealed) to be more pro-immigrant than the average adjust by becoming less tolerant, and people located in areas which turn out to be less pro-immigrant than the average adjust by becoming more tolerant, and these two effects fully offset each other when the ex-ante stereotype is weak. More generally, our analysis indicates that the magnitude of the effect of the information shock on aggregate behavior is inversely related to $\Theta$, and is thus increasing in the strength of the ex-ante stereotype.

## 5 Concluding remarks

A referendum is a universal vote on a particular proposal, but its result publicly reveal new information about the population's underlying preferences. If prevalent private views within society were

[20] Recall that, from (19), if $\mu > \overline{\mu}$ then both $\overline{a}_0$ and $\overline{a}_1$ are $< \mu$ (while the opposite holds if $\mu < \overline{\mu}$). When $\mu > \overline{\mu}$, $\overline{a}_1 > \overline{a}_0$ implies $\mu > \overline{a}_1 > \overline{a}_0$ (and, similarly, when $\mu < \overline{\mu}$, $\overline{a}_1 < \overline{a}_0$ implies $\mu < \overline{a}_1 < \overline{a}_0$). In both cases, $|\overline{a}_1 - \mu| < |\overline{a}_0 - \mu|$.

previously imperfectly observed, this information shock can trigger changes in individual behavior and in the social norm. We argue that this logic can explain the sharp increase in hate crime and its geographical variation in the aftermath of the Brexit referendum in June 2016.

In this paper, we show that the hike in hate crime was more pronounced in areas supporting remaining in the EU. As the leave vote is associated with concerns with immigration, we interpret remain areas as being friendlier towards immigrants. Why did hate crime increase more in these areas? Did the British public's attitudes in relatively open areas suddenly become more hostile overnight towards immigrants and ethnic minorities? Why do we not observe a same geographical pattern in other episodes of hate crime spikes? Our theoretical framework provides answers to these questions.

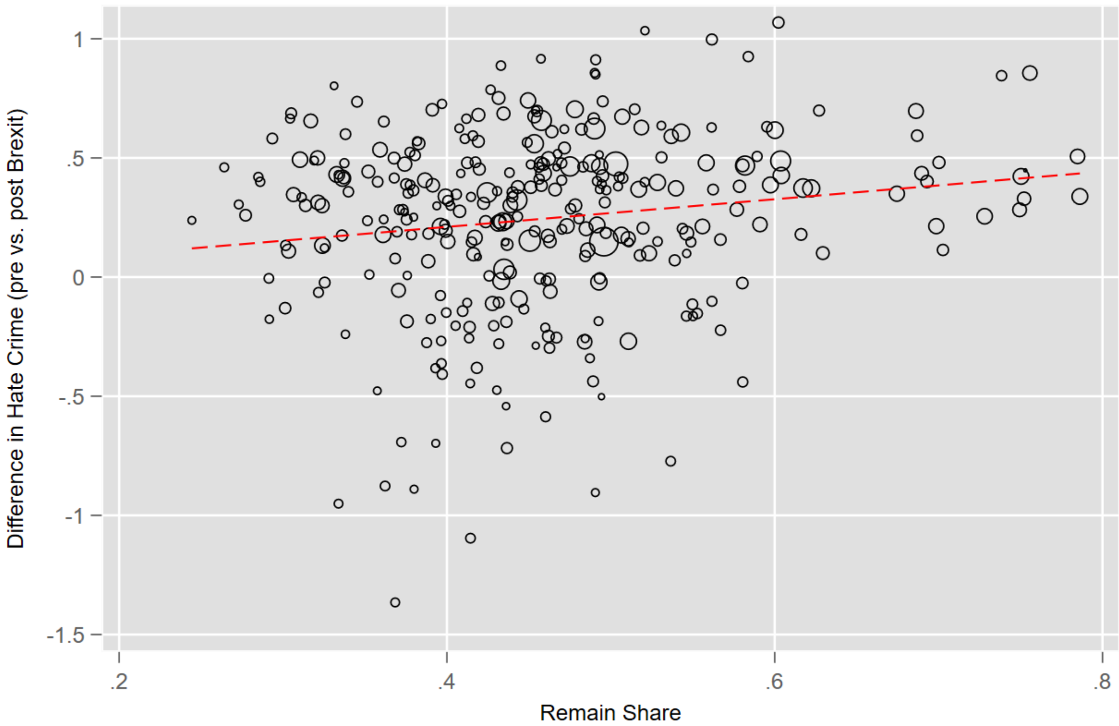
We explain the spike in hate crime and its variation across areas according to voting patterns through a theory of social norm compliance coupled with an information shock. In our framework, the referendum revealed that anti-immigrant sentiment was more widespread in the UK than was previously believed, and this triggered an update of the social norms governing behavior towards ethnic and religious minorities. Since agents' beliefs about prevalent views in society are guided by their local area, this surprise effect was particularly pronounced in remain voting areas. In those areas, the perception of an immigrant-friendly social norm made those opposed to migrants conceal or repress their private views. The referendum empowered those latent haters, who, after updating their perception about prevalent views in UK society, started adopting a behavior more in line with their true preferences. At the limit, an update of the social norm can translate into violence and intimidation on the streets against racial minorities in a way that crosses the line of the law and becomes hate crime. As the surprise effect of the national result was greater in remain areas, it follows that these areas exhibited a greater a change in the trend of hate crime.

To add credence to the mechanism, we develop and estimate a quantitative version of the model. We show that, in order to match the stylized facts reported, it is sufficient that agents put relatively little weight on their desire to conform when choosing how to behave, approximately a third of the weight they apply to their own preferences. Further, our quantitative analysis suggests that the referendum was a large information shock, as measured by the difference in prior and posterior mean attitudes relative to the variance. We argue that this was probably due the existence of a strong pre-referendum national stereotype, which was not confirmed by the referendum outcome. Future research should investigate the extent to which new public information about preferences (such as that provided by the referendum) might crystallize into a "new stereotype," which may then shape attitudes for a long time. Without further information shocks, this "long tail effect" of salient past history might persist even if the underlying preferences it reflects were to evolve into a new, different consensus. More generally, investigating the multi-faceted relationship between preferences, norms and behavior is crucial for our understanding of how societies work and evolve. This research agenda has gained considerable momentum in the last few years, but there are still many unanswered questions. We hope that our paper might contribute towards filling this gap.

# A Empirical Appendix

## A.1 Correlation of Rise in Hate Crime and Referendum Result

Figure A.1: Rise in Hate Crime and Referendum Result




**Note:** Change in hate crime is computed as $\log(E(\text{hate crime}|\text{post Brexit})) - \log(E(\text{hate crime}|\text{pre Brexit}))$. The size of the scatter plot is proportional to the population of the CSP. The red line represents an ordinary least squares line of best fit.

## A.2 Sectoral Composition

Table A.2: Coefficients of log(GVA) of given sector

| | (3) | (4) | (5) | (6) |
|---|---|---|---|---|
| Production | -0.019 | -0.019 | -0.032** | -0.029** |
| | (0.013) | (0.013) | (0.013) | (0.013) |
| Manufacturing | -0.074*** | -0.051** | -0.023 | -0.016 |
| | (0.024) | (0.024) | (0.026) | (0.026) |
| Construction | 0.067** | 0.023 | 0.012 | -0.001 |
| | (0.030) | (0.029) | (0.034) | (0.034) |
| Distribution | -0.047 | 0.023 | 0.096* | 0.068 |
| | (0.044) | (0.044) | (0.051) | (0.050) |
| Information | -0.053*** | -0.064*** | 0.000 | -0.004 |
| | (0.020) | (0.019) | (0.022) | (0.022) |
| Finance | 0.075*** | 0.073*** | -0.006 | -0.008 |
| | (0.019) | (0.019) | (0.022) | (0.022) |
| Real estate | 0.159*** | 0.200*** | -0.070 | -0.042 |
| | (0.053) | (0.052) | (0.065) | (0.065) |
| Professional | 0.105*** | 0.090*** | -0.004 | 0.000 |
| | (0.029) | (0.029) | (0.031) | (0.031) |
| Public services | -0.187*** | -0.187*** | -0.123** | -0.116** |
| | (0.051) | (0.051) | (0.058) | (0.058) |
| Other services | 0.025 | 0.009 | 0.054* | 0.045 |
| | (0.027) | (0.026) | (0.031) | (0.030) |

## A.3 Other Crime Placebo

Table A.3: Dependent Variable: Log All Other Crime

| | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| Lag Dep. Variable | | | | | 0.430*** |
| | | | | | (0.007) |
| **Post Brexit × Remain Share** | **-0.172*** ** | **-0.225*** ** | **-0.242*** ** | **-0.003** | **0.006** |
| | **(0.028)** | **(0.034)** | **(0.034)** | **(0.038)** | **(0.035)** |
| Log NiNo EU | | 0.001 | 0.002 | 0.015*** | 0.012*** |
| | | (0.002) | (0.002) | (0.002) | (0.002) |
| Log NiNo RoW | | 0.008*** | 0.008*** | 0.010*** | 0.008*** |
| | | (0.002) | (0.002) | (0.002) | (0.002) |
| Log Population | | 0.697*** | 0.599*** | 0.035 | 0.015 |
| | | (0.040) | (0.057) | (0.058) | (0.053) |
| Log GDI | | | 0.015 | -0.026 | -0.015 |
| | | | (0.038) | (0.040) | (0.037) |
| Log Social Benefits | | | 0.056* | 0.171*** | 0.091** |
| | | | (0.032) | (0.039) | (0.036) |
| Observations | 19,530 | 18,900 | 18,900 | 18,900 | 18,585 |
| R-squared | 0.0501 | 0.776 | 0.751 | 0.988 | 0.990 |
| Number of CSPs | 315 | 315 | 315 | 315 | 315 |
| CSP FE | YES | YES | YES | YES | YES |
| Season Dummies | YES | YES | YES | YES | YES |
| Sectotal Composition | - | - | YES | YES | YES |
| Year FE | YES | YES | YES | - | - |
| Force-Year FE | - | - | - | YES | YES |

## A.4 Hate Crime Excluding London

Table A.4: Dependent Variable: Log Hate Crime (excluding London Boroughs)

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| Lag Dep. Variable | | | | | | 0.112*** |
| | | | | | | (0.008) |
| **Post Brexit ×** | **0.486*****| **0.380**** | **0.320*** | **0.410**** | **0.549***** | **0.504***** |
| **Remain Share** | **(0.135)** | **(0.164)** | **(0.165)** | **(0.163)** | **(0.193)** | **(0.191)** |
| Log NiNo EU | | -0.001 | -0.000 | 0.000 | 0.001 | -0.000 |
| | | (0.008) | (0.008) | (0.008) | (0.008) | (0.008) |
| Log NiNo RoW | | 0.005 | 0.006 | 0.002 | 0.008 | -0.000 |
| | | (0.009) | (0.009) | (0.009) | (0.009) | (0.009) |
| Log Population | | 1.526*** | 1.491*** | 1.255*** | 1.298*** | 1.100*** |
| | | (0.191) | (0.271) | (0.268) | (0.301) | (0.305) |
| Log GDI | | | -0.400** | -0.454** | -0.827*** | -0.620*** |
| | | | (0.191) | (0.189) | (0.237) | (0.239) |
| Log Social Benefits | | | 0.322** | 0.252* | 0.132 | 0.047 |
| | | | (0.147) | (0.145) | (0.188) | (0.188) |
| Log Other Crime | | | | 0.587*** | 0.494*** | 0.441*** |
| | | | | (0.028) | (0.032) | (0.032) |
| Observations | 17,484 | 16,920 | 16,920 | 16,920 | 16,920 | 16,638 |
| R-squared | 0.0313 | 0.522 | 0.493 | 0.609 | 0.867 | 0.869 |
| Number of CSPs | 282 | 282 | 282 | 282 | 282 | 282 |
| CSP FE | YES | YES | YES | YES | YES | YES |
| Season Dummies | YES | YES | YES | YES | YES | YES |
| Sectoral Composition | - | - | YES | YES | YES | YES |
| Year FE | YES | YES | YES | YES | - | - |
| Force-Year FE | - | - | - | - | YES | YES |

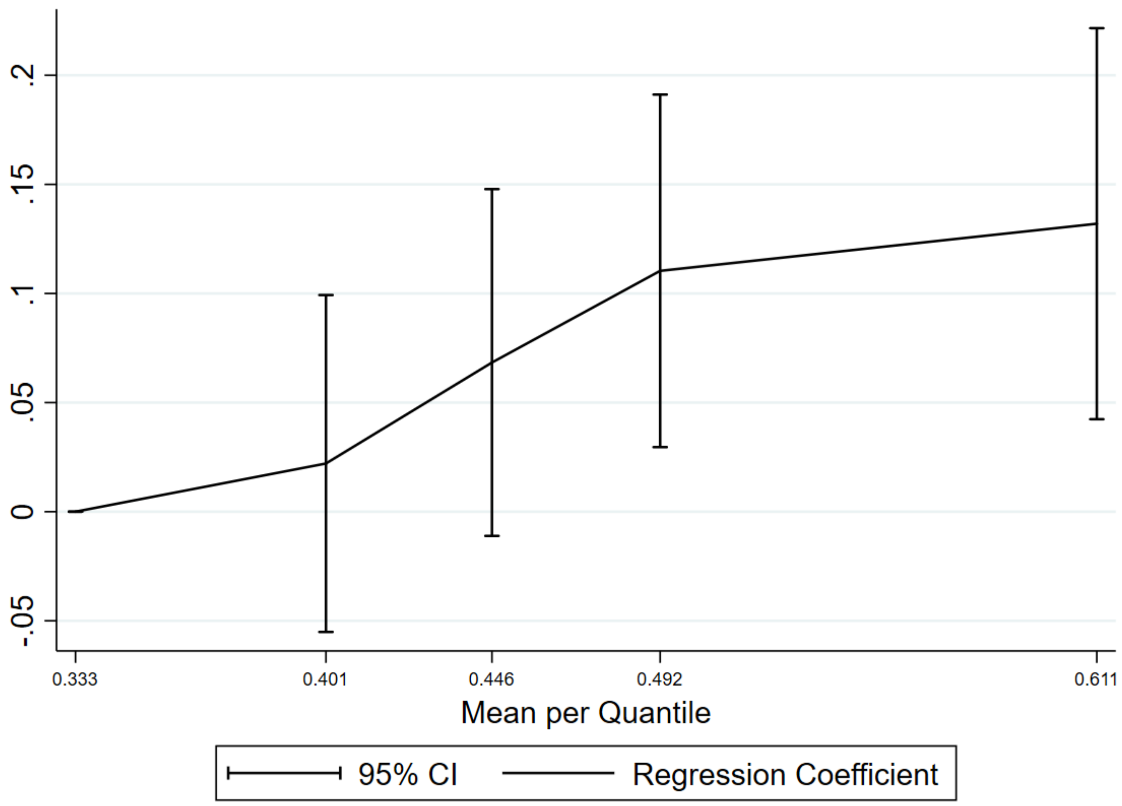
## A.5 Linearity

To understand if the linear specification is suitable we impose a slightly less parametric approach. We take the same specification as column four in the baseline results but modify the parameter of interest. Instead of using the remain share, we split the CSPs into five quintiles based on the proportion who voted remain. The variable quintile$_i^j$ below takes the value one if CSP $i$ falls in the

$j^{\text{th}}$ quintile and the value zeros otherwise.

$$\text{hate}_{it} = \sum_{t=2}^{5} \beta_j \left( \mathit{1}_{\{\text{Post Brexit}\}} \times \text{quintile}_i^j \right) + \gamma \mathbf{X}_{it} + \tau_t + \eta_i + \epsilon_{it}$$

Figure A.5: Quintile Dummies $\beta_j$




**Note:** The figure plots the coefficients $\beta^j$ against the mean voting remain in the given quintile.

The coefficients on the quintile dummies are monotonically increasing which reassures us that pro remain areas did indeed see a larger increase in the levels of hate crime post Brexit. The relationship is approximately linear in the sense that the confidence bounds are sufficiently wide not to reject linearity as a null. However, the effect appears strongest at marginal levels of voting.

## A.6 Estimation of Reporting Rates

The crime specific component, $\omega^1_{ct}$, can be identified by the estimated reporting rates each year as reported in Table D8 of "*Crime in England and Wales: Annual Trend and Demographic Tables*". [21] Crime types across the two data sources do not align perfectly. We allocate the twelve different types of hate crime into one of five crime types in the CSEW. The crime types in the CSEW we use are: violence-wounding (VWo); violence-assualt with minor injury (VMI); violence-without injury (VWI); criminal damage to a vehicle (CDV); and arson and other criminal damage (OCD).

Table A.6: Crime Equivalency Table

| Hate crime in baseline data | CSEW (*equivalent*) crime |
|---|---|
| Racially or religiously aggravated assault without injury | VWI |
| Racially or religiously aggravated Criminal damage to a dwelling | OCD |
| Racially or religiously aggravated Criminal damage to a building other than a dwelling | OCD |
| Racially or religiously aggravated Criminal damage to a vehicle | CDV |
| Racially or religiously aggravated other Criminal damage | OCD |
| Racially or religiously aggravated Criminal damage | OCD |
| Racially or religiously aggravated less serious wounding | VMI |
| Racially or religiously aggravated inflicting grievous bodily harm without intent | VWo |
| Racially or religiously aggravated actual bodily harm and other injury | VWo |
| Racially or religiously aggravated harassment | VWI |
| Racially or religiously aggravated assault with injury | VWo |
| Racially or religiously aggravated public fear, alarm or distress | VWI |

The hate crime reporting component $\omega^0_t$ is identified through the reporting rates of hate crime according to the CSEW, reported in Table 3. Due to their being relatively few instances of victims reporting hate crimes, we allow this parameter to take two values, one pre-referendum $\omega^0_0$, and one post-referendum $\omega^0_1$. Further, we divide the sample into two subperiods, $t \in \tau_0$ refers to the first three columns of Table 3, and is the period between April 2007 and March 2015, inclusive. The period from April 2015 onwards is defined as $t \in \tau_1$. Thus for $t$ pre-referendum, $t \in \tau_0$, we can compute $m_0 := \frac{\sum_{t\in\tau_0}\sum_c h_{ct}}{\sum_{t\in\tau_0}\sum_c h^{\star}_{ct}}$, directly from Table 3. Rearranging, and using our baseline data gives an estimate for $\omega^0_0$ as

$$\hat{\omega}^0_0 = \frac{m_0 \sum_{t\in\tau_0}\sum_c \frac{h_{ct}}{\omega^1_{ct}}}{\sum_{t\in\tau_0}\sum_c h_{ct}}.$$

[21] Data can be downloaded from: https://www.ons.gov.uk/peoplepopulationandcommunity/crimeandjustice/datasets/

## A.7 Omitted Covariates from Table 7

Table A.7: Perceived likelihood of leaving the EU

| | | | | |
|---|---|---|---|---|
| Age | | | 0.136*** | 0.169*** |
| | | | (0.030) | (0.035) |
| Age Squared | | | -0.001*** | -0.001*** |
| | | | (0.000) | (0.000) |
| Female | | | 1.072*** | 0.937*** |
| | | | (0.165) | (0.183) |
| British | | | -0.790** | -0.713 |
| | | | (0.397) | (0.436) |
| Non White | | | 1.505*** | 1.435** |
| | | | (0.535) | (0.590) |
| **Education** | | | | |
| GCSE D-G | | | | -1.212** |
| | | | | (0.525) |
| GCSE A*-C | | | | -2.203*** |
| | | | | (0.378) |
| A-level | | | | -3.269*** |
| | | | | (0.389) |
| Undergraduate | | | | -3.936*** |
| | | | | (0.378) |
| Postgrad | | | | -3.870*** |
| | | | | (0.378) |
| **Income Bracket (per year)** | | | | |
| 5,000 to 9,999 | | | | -0.636 |
| | | | | (0.703) |
| 10,000 to 14,999 | | | | -1.250* |
| | | | | (0.665) |
| 15,000 to 19,999 | | | | -1.187* |
| | | | | (0.666) |
| 20,000 to 24,999 | | | | -1.513** |
| | | | | (0.664) |
| 25,000 to 29,999 | | | | -1.980*** |
| | | | | (0.668) |
| 30,000 to 34,999 | | | | -1.479** |
| | | | | (0.682) |
| 35,000 to 39,999 | | | | -1.311* |
| | | | | (0.697) |
| 40,000 to 44,999 | | | | -2.029*** |
| | | | | (0.713) |
| 45,000 to 49,999 | | | | -1.190 |
| | | | | (0.746) |
| 50,000 to 59,999 | | | | -1.891*** |
| | | | | (0.716) |
| 60,000 to 69,999 | | | | -2.294*** |
| | | | | (0.762) |
| 70,000 to 99,999 | | | | -2.051*** |
| | | | | (0.738) |
| 100,000 to 149,999 | | | | -2.248** |
| | | | | (0.936) |
| 150,000 and over | | | | -2.774** |
| | | | | (1.249) |
| Prefer not to answer | | | | -1.790*** |
| | | | | (0.628) |
| Don't know | | | | -1.538** |
| | | | | (0.693) |
| Constant | 37.496*** | 37.230*** | 31.795*** | 37.447*** |
| | (0.346) | (0.731) | (1.058) | (1.342) |
| Observations | 54,916 | 54,916 | 49,341 | 41,317 |
| R-squared | 0.180 | 0.184 | 0.189 | 0.192 |

# B Theoretical Appendix

## B.1 A more general setup

The setup we have considered in the main body assumes that individuals care about conforming with the average behavior in society. This is sensible enough, but raises possible questions. For instance, it is well possible that people may not only care about society as a whole, but may also be concerned with conforming to other reference groups, perhaps at a more disaggregate level. To analyze what would happen in that case, suppose that rather than $\overline{a}$, people wish to conform to the following "reference behavior," formed as a weighted average of different components:

$$\lambda\overline{a} + (1-\lambda)\,\overline{a}^d \tag{21}$$

where $\overline{a}^d$ and $\overline{a}$ represent, as before, mean attitudes in $i$'s area of residence and mean attitudes in the country as a whole, and the parameter $\lambda$ satisfies $0 \leq \lambda \leq 1$. Individual payoff is now equal to

$$u_i = -\theta\,(a_i - \alpha_i)^2 - (1-\theta)\,(a_i - \lambda\overline{a} - (1-\lambda)\,\overline{a}^d)^2. \tag{22}$$

Another simplifying assumption of the main setup is that, before the referendum, individuals perfectly observe their district's aggregate preferences. We now allow for greater generality and assume that each individual instead observes a signal given by

$$\widetilde{P}_i^d = P^d + \varrho_i$$

where $\varrho_i$ is drawn as $N\,(0, x)$. The case analyzed in the main setup above is thus a special case of this more general model in which $x = 0$ and $\lambda = 1$. The individual best-reply action is now given by

$$a_i = \theta\alpha_i + (1-\theta)\,[\lambda E\,(\overline{a}) + (1-\lambda)\,E(\overline{a}^d)]$$

In this setup, the Brexit vote plays the dual role of revealing *both* $P^d$ and $\mu$. Before the vote, individual expectations were formed through linear regressions against all the relevant information available, in this case $\widetilde{P}_i^d$, $\overline{\mu}$ and $\alpha_i$ – the latter is relevant because it is informative about $P^d$ (and thus, indirectly, about $\mu$ as well). In what follows, for simplicity, we will assume that $n \to \infty$.

**Proposition 3.** *When $\mu$ and $P^d$ are imperfectly observed, the unique linear equilibrium of the game is given by,*

$$a_i = k_0\alpha_i + \gamma_0\widetilde{P}_i^d + (1 - k_0 - \gamma_0)\overline{\mu} \tag{23}$$

$k_0 \equiv \theta\frac{x(1+\Theta)+\sigma(x+\theta(1-\lambda+\Theta)+\lambda)}{(x+\sigma)(\theta(1-\lambda)+\lambda+\theta\Theta)+\sigma x}$ *and* $\gamma_0 \equiv \frac{\sigma(\theta(1-\theta)(1-\lambda+\Theta))}{(x+\sigma)(\theta(1-\lambda)+\lambda+\theta\Theta)+\sigma x}$.
*When $\mu$ and $P^d$ are observed, the unique linear equilibrium is,*

$$a_i = k_1\alpha_i + \gamma_1 P^d + (1 - k_1 - \gamma_1)\mu \tag{24}$$

*where* $k_1 \equiv \theta$ *and* $\gamma_1 \equiv \frac{\theta(1-\theta)(1-\lambda)}{\theta+\lambda(1-\theta)}$.
**Proof:** *provided in Appendix B.4.*

The difference in mean behavior before and after the referendum is thus,

$$\Delta\overline{a} = (1 - k_0 - \gamma_0)(\mu - \overline{\mu}) \tag{25}$$

which replicates Result 1. Result 2 is similarly replicated by noting that mean pre-referendum beliefs about $\mu$ in district $d$ are,

$$E[E_i(\mu \mid \widetilde{P}_i^d, \alpha_i)] = \frac{\Theta(\sigma + x)P^d + (x + \sigma + x\sigma)\overline{\mu}}{\sigma(1 + x + \Theta) + x(1 + \Theta)}$$

increasing in $P^d$. Consider now Result 3. The district-level change in attitudes is,

$$\Delta\overline{a}^d = (k_1 + \gamma_1 - k_0 - \gamma_0)P^d + (1 - k_1 - \gamma_1)\mu - (1 - k_0 - \gamma_0)\overline{\mu}$$

where it can be easily shown that,

**Lemma 1.** *If $x > 0$, then there exists $\widehat{\lambda} \in (0,1)$ such that,*

$$k_1 + \gamma_1 - k_0 + \gamma_0 < 0 \text{ (resp., } > 0) \Longleftrightarrow \lambda > \text{(resp., } < ) \widehat{\lambda}$$

*If $x = 0$, $k_1 + \gamma_1 - k_0 + \gamma_0 < 0$ for all values of $\lambda$.*

**Proof:** *provided in Appendix B.4.*

It then immediately follows that,

**Result 5.** *Suppose that $x > 0$. There exist $\widehat{\lambda} \in (0,1)$ such that, if $\lambda >$ (resp., $<$ ) $\widehat{\lambda}$, the change in mean attitudes caused by the referendum is decreasing (resp., increasing) in $P^d$.*

Compared to the previous analysis, here an additional effect comes into play. This arises because Brexit generated a surprise effect not only with respect to $\mu$ – a *nation-level surprise effect* – but also with respect to $P^d$ – a *district-level surprise effect.* To understand what this implies, it is instructive to compute the difference between realized $\mu$ and $P^d$ and pre-referendum beliefs about them. These are,

$$\mu - E_d[E_i(\mu \mid \widetilde{P}_i^d, \alpha_i)] = \frac{\Theta\,(\sigma + x)\,(\mu - P^d) + (x + \sigma + x\sigma)\,(\mu - \overline{\mu})}{\Lambda} \text{ decreasing in } P^d \tag{26}$$

and

$$P^d - E_d[E_i(P^d \mid \widetilde{P}_i^d, \alpha_i)] = \frac{\sigma x(P^d - \overline{\mu})}{\Lambda} \text{ increasing in } P^d \tag{27}$$

where $\Lambda \equiv \sigma\,(1 + x + \Theta) + x\,(1 + \Theta)$. There are thus two opposing forces at play. On the one hand, (26) confirms that, as already discussed above, compared to people in anti-immigrant areas, people in pro-immigrant districts tended to underestimate the true extent of nation-wide anti-immigrant animus. Ceteris paribus, this should induce them to revise their attitudes more. On the other hand, (27) shows that, compared to more pro-immigrant districts, people in anti-immigrant areas tended to underestimate the true amount of anti-immigrant sentiment in their area. Ceteris paribus, this induces *them* to react more strongly to the new information released by the referendum. Result 5 shows that, if conforming to district-level behavior is sufficiently important relative to conforming to the nation as a whole, the latter effect will dominate, and vice-versa. This implies that empirical Observation 2 can be used make inferences about the relevant reference group that people in the UK want to conform to. In particular, our findings suggest that $\lambda$ cannot be too low and, hence, that mean behavior in the country as a whole is a non-negligible component of reference behavior.

These observations can help us shed light on the apparent puzzle concerning Scotland. In the referendum, Scottish people voted for the UK to remain in the EU (by 62% against 38%). However, in Scotland, hate crime did not increase in the aftermath of the referendum – in fact, quite the opposite. This stands in contrast with what happened in all other UK areas with similar vote outcome, and appears to contradict with our empirical Conjecture 2, namely that Remain areas experienced a more pronounced increment in hate crime than Leave areas (and Result 3, which shows that this is what our theory predicts). A relevant question then is: can we reconcile these two contradictory facts?

The key to the puzzle rests in the observation that Scottish people have a strong Scottish identity and see themselves as quite distinct from the rest of the UK. According to the 2014 Scottish Social Attitude Survey, for instance, 49% of Scots see themselves as more Scottish than British, and a further 32% have an equally strong Scottish and British identity. From a theoretical viewpoint, this suggests that $\lambda^{\text{Scot}}$, the weight given by Scottish people to the UK as a whole in their reference behavior, might be rather low. The theory thus predicts that, in Scotland, the change in attitudes following the referendum should have been primarily governed by the district-level surprise effect, differently from other UK regions where nation-level surprise played a dominant role. The strong Scottish identity that characterizes Scottish people can thus reconcile the "Scottish puzzle" with Conjecture 2 and Result 3, through the following observation,

**Result 6.** *Suppose that $\lambda^{Scot} < \lambda$ (where $\lambda$ applies in the rest of the UK). There exist $t \in (0,1)$ and $D > 0$ such that, if $\lambda - \lambda^{Scot} > t$, then necessarily $\Delta\overline{a}^{Scot} < \Delta\overline{a}^{d}$ for all districts $d$ that satisfy $P^d < D + P^{Scot}$.*

## B.2 Proof of Proposition 1

Suppose that, in equilibrium, all players select $a_i = k\alpha_i + \gamma P^d + \delta\overline{\mu}$. From (5), $i$'s best reply is

$$\begin{aligned} a_i &= \theta\alpha_i + (1-\theta)\left(kE(\mu) + \gamma\frac{\left(P^d + (n-1)E(\mu)\right)}{n} + \delta\overline{\mu}\right) \\ &= \theta\alpha_i + (1-\theta)\gamma\frac{P^d}{n} + (1-\theta)\delta\overline{\mu} + (1-\theta)\left(k + \gamma\frac{n-1}{n}\right)E(\mu) \end{aligned}$$

Substituting for $E(\mu) = \frac{\Theta}{1+\Theta}P^d + \frac{1}{1+\Theta}\overline{\mu}$ we obtain

$$\theta\alpha_i + (1-\theta)\gamma\frac{P^d}{n} + (1-\theta)\delta\overline{\mu} + (1-\theta)\left(k + \gamma\frac{n-1}{n}\right)\left(\frac{\Theta}{1+\Theta}P^d + \frac{1}{1+\Theta}\overline{\mu}\right)$$

i.e.

$$\theta\alpha_i + P^d(1-\theta)\left(\frac{\gamma}{n} + \left(k + \gamma\frac{n-1}{n}\right)\frac{\Theta}{1+\Theta}\right) + \overline{\mu}(1-\theta)\left(\delta + \left(k + \gamma\frac{n-1}{n}\right)\frac{1}{1+\Theta}\right)$$

It thus follows that

$$
\begin{aligned}
k &= \theta && (28)\\
\gamma &= (1-\theta)\left(\frac{\gamma}{n}+\left(k+\gamma\frac{n-1}{n}\right)\frac{\Theta}{1+\Theta}\right) && (29)\\
\delta &= (1-\theta)\left(\delta+\left(k+\gamma\frac{n-1}{n}\right)\frac{1}{1+\Theta}\right) && (30)
\end{aligned}
$$

Substituting for $k=\theta$ and rearranging, (29) gives

$$\gamma=\frac{n\theta\Theta\,(1-\theta)}{n\,(1+\Theta\theta)-(1-\theta)}.$$

Substituting for $k$ and $\gamma$ in (30) we obtain $\delta=1-\theta-\gamma$. QED

### B.3 Proof of Proposition 2

Suppose that, in equilibrium, all players select $a_i=k\alpha_i+\gamma P^d+\delta\overline{\mu}+\tau\mu$. From (5), $i$'s best reply is

$$
\begin{aligned}
a_i &= \theta\alpha_i+(1-\theta)\left(k\mu+\gamma\frac{\left(P^d+(n-1)\mu\right)}{n}+\delta\overline{\mu}+\tau\mu\right)\\
&= \theta\alpha_i+P^d\,(1-\theta)\,\frac{\gamma}{n}+\overline{\mu}\,(1-\theta)\,\delta+\mu\,(1-\theta)\left(k+\gamma\frac{n-1}{n}+\tau\right)
\end{aligned}
$$

It thus follows that

$$
\begin{aligned}
k &= \theta && (31)\\
\gamma &= (1-\theta)\,\frac{\gamma}{n} && (32)\\
\delta &= (1-\theta)\,\delta && (33)\\
\tau &= (1-\theta)\left(k+\gamma\frac{n-1}{n}+\tau\right) && (34)
\end{aligned}
$$

It is apparent from (32) and 33) that $\gamma=\delta=0$. Substituting for these, and for $k=\theta$, in (34), we obtain $\tau=1-\theta$. QED

## B.4 Proof of Proposition 3 and proof of Lemma 1

**Proof of Proposition 3** We first solve for the pre-referendum equilibrium. Note that

$$E(\mu \mid \widetilde{P}_i^d) = \frac{\Theta\sigma}{\sigma(1+x+\Theta)+x(1+\Theta)}\widetilde{P}_i^d \tag{35}$$

$$+\frac{\Theta x}{\sigma(1+x+\Theta)+x(1+\Theta)}\alpha_i + \frac{x+\sigma+x\sigma}{\sigma(1+x+\Theta)+x(1+\Theta)}\overline{\mu}. \tag{36}$$

and

$$E(P^d \mid \widetilde{P}_i^d) = \frac{(1+\Theta)\sigma}{\sigma(1+x+\Theta)+x(1+\Theta)}\widetilde{P}_i^d$$
$$+\frac{(1+\Theta)x}{\sigma(1+x+\Theta)+x(1+\Theta)}\alpha_i + \frac{\sigma x}{\sigma(1+x+\Theta)+x(1+\Theta)}\overline{\mu}$$

Consider an equilibrium where $a_i = k\alpha_i + \gamma\widetilde{P}_i^d + (1-k-\gamma)\overline{\mu}$, and, hence, $\overline{a} = (k+\gamma)\mu + (1-k-\gamma)\overline{\mu}$ and $\overline{a}^d = (k+\gamma)P^d + (1-k-\gamma)\overline{\mu}$.

The best reply of individual $j$ living in district $d$ is

$$\begin{aligned} a_j &= \theta\alpha_j + (1-\theta)\left[\lambda E_j(\overline{a}) + (1-\lambda)E_j(\overline{a}^d)\right] \\ &= \theta\alpha_j + (1-\theta)(1-k-\gamma)\overline{\mu} \\ &\quad + (1-\theta)(k+\gamma)\Big[\lambda E_j(\mu) + (1-\lambda)E_j\left(P^d\right)\Big] \end{aligned}$$

Substituting for $E_j(\mu)$ and $E_j\left(P^d\right)$ we obtain

$$\begin{aligned} k &= \theta + (1-\theta)(k+\gamma)\left(\lambda\frac{\Theta x}{\sigma(1+x+\Theta)+x(1+\Theta)} + (1-\lambda)\frac{(1+\Theta)x}{\sigma(1+x+\Theta)+x(1+\Theta)}\right) \\ \gamma &= (1-\theta)(k+\gamma)\left(\lambda\frac{\Theta\sigma}{\sigma(1+x+\Theta)+x(1+\Theta)} + (1-\lambda)\frac{(1+\Theta)\sigma}{\sigma(1+x+\Theta)+x(1+\Theta)}\right) \end{aligned}$$

Solving out the system we obtain $k_0$ and $\gamma_0$ described in proposition 3.

Consider now the post-referendum equilibrium. Suppose that, in equilibrium, $a_i = k\alpha_i + \gamma P^d + (1-k-\gamma)\mu$ so that $\overline{a} = \mu$ and $\overline{a}^d = (k+\gamma)P^d + (1-k-\gamma)\mu$. The best reply of individual $j$ living in district $d$ is

$$\begin{aligned} a_j &= \theta\alpha_j + (1-\theta)\Big[\lambda\mu + (1-\lambda)\Big((k+\gamma)P^d + (1-k-\gamma)\mu\Big)\Big] \\ &= \theta\alpha_j + (1-\theta)\Big[(\lambda + (1-\lambda)(1-k-\gamma))\mu + (1-\lambda)(k+\gamma)P^d\Big] \end{aligned}$$

This gives

$$\begin{aligned} k &= \theta \\ \gamma &= (1-\theta)(1-\lambda)(k+\gamma) \end{aligned}$$

Solving out the system we obtain $k_1$ and $\gamma_1$ described in proposition 3.

**Proof of Lemma 1** The value of $k_1 + \gamma_1 - k_0 - \gamma_0$. is equal to

$$-\theta(1-\theta)\frac{x\sigma(1-\lambda)-(x+\sigma)\Theta\lambda}{(\theta+\lambda-\theta\lambda)(\theta(x+\sigma)(1-\lambda)+(\sigma+\lambda+\theta\Theta)x+(\sigma\lambda+\theta\Theta\sigma))} \tag{37}$$

If $x = 0$, this is negative for all $\lambda > 0$. If $x > 0$, this is positive when $\lambda = 0$ and negative when $\lambda = 1$. Note that the numerator of (37) is decreasing in $\lambda$ while the denominator is increasing. Overall, therefore, (37) is decreasing in $\lambda$. We conclude that there is a value $\widehat{\lambda} \in (0, 1)$ such that (37) is negative for $\lambda > \widehat{\lambda}$ and positive for $\lambda < \widehat{\lambda}$.

### B.5 Social Image Concerns

Consider now an alternative environment where, in addition to matching their preferences, people are concerned with their social image (rather than with following the country-wide norm). Similar to Bursztyn el al. (2019), we assume that social image increases in the proximity between an individual's inferred type and the types of those observing the individual's behavior. Formally, this can be modeled as

$$u_i = -\theta(a_i - \alpha_i)^2 - (1-\theta)\left(E(\alpha_i \mid a_i) - P^d\right)^2$$

where $P^d$ represents average preferences in $i$'s district and $E(\alpha_i \mid a_i)$ is individual $i$'s average inferred type conditional on his behavior. Intuitively, this captures the idea that the "audience" to an individual's behavior towards immigrants (and possible hate crimes) is represented by other individuals living in the same area.

The Brexit referendum revealed aggregate preferences in districts as well as the overall nation. While in the main analysis we focused on the latter, we now focus on the former (since we are now considering a setup where people which to conform to preferences in their own district). Suppose then that, prior to the referendum, people do not observe $P^d$ and have to form expectations about it. Similar to Section B.1 above, we assume that each individual $i$ observes a signal

$$\widetilde{P}_i^d = P^d + \varrho_i$$

where $\varrho_i$ is drawn as $N(0, x)$. As in the main model, the preferences of individual $i$ in district $d$ are given by $\alpha_i^d = P^d + \varepsilon_i$, where $\varepsilon$ is drawn as $N(0, \sigma)$. In the absence of further information, individual $i$'s beliefs about $P^d$ are

$$\begin{aligned} E\left(P^d \mid \alpha_i\right) &= \frac{(1+\Theta)\,x}{\sigma\,(1+x+\Theta)+x\,(1+\Theta)}\alpha_i + \frac{(1+\Theta)\,\sigma}{\sigma\,(1+x+\Theta)+x\,(1+\Theta)}\widetilde{P}_i^d \\ &\quad + \frac{\sigma x}{\sigma\,(1+x+\Theta)+x\,(1+\Theta)}\overline{\mu}. \end{aligned}$$

We consider two types of equilibria that may arise.[22] First a fully revealing equilibrium in which optimal individual action is given by $a_i = r_0\alpha_i + r_1\widetilde{P}_i^d + (1-r_0-r_1)\,\overline{\mu}$ for $r_0 > 0$ so that $E\left(\alpha_i \mid a_i\right) = E\left[E\left(\alpha_i \mid a_i, \alpha_j, \widetilde{P}_j^d\right)\right] = \frac{a_i - r_1P^d-(1-r_0-r_1)\overline{\mu}}{r_0}$ by the law of iterated expectations. Individual $i$ maximizes

$$-\theta\,(a_i-\alpha_i)^2 - (1-\theta)\left(\frac{a_i - r_1P^d - (1-r_0-r_1)\,\overline{\mu}}{r_0} - P^d\right)^2$$

Solving out, we see that the optimal action is,

$$a_i = \frac{r_0^2}{r_0^2\theta+1-\theta}\left[\theta\alpha_i + \frac{1-\theta}{r_0^2}\left((r_1+r_0)\,E\left(P^d \mid \alpha_i\right) + (1-r_0-r_1)\,\overline{\mu}\right)\right].$$

Substituting for $E\left(P^d \mid \alpha_i\right)$, it is straightforward to see that in any fully separating equilibrium the values $r_0$, $r_1$ must solve the following system,

$$r_0 = \frac{r_0^2}{r_0^2\theta+1-\theta}\left(\theta + \frac{1-\theta}{r_0^2}\,(r_1+r_0)\,\frac{(1+\Theta)\,x}{\sigma\,(1+x+\Theta)+x\,(1+\Theta)}\right), \tag{38}$$

$$r_1 = \frac{1-\theta}{r_0^2\theta+1-\theta}\,(r_1+r_0)\,\frac{(1+\Theta)\,\sigma}{\sigma\,(1+x+\Theta)+x\,(1+\Theta)}. \tag{39}$$

[22] For brevity, we focus on fully separating or fully pooling equilibria, ignoring the possibility of partial pooling. This is however inconsequential for our purposes, since $\overline{a}_{before}^d$ can be expressed as a weighted average of $\alpha_i$, $\widetilde{P}_i^d$ and $\overline{\mu}$ also in the case of partial pooling.

Second, there may also be a pooling equilibrium, in which $a_i = \overline{\mu}$, independent of $\alpha_i$. In both cases, district $d$'s aggregate behavior can be expressed as,

$$\overline{a}^d_{before} = r_{before}P^d + (1 - r_{before})\,\overline{\mu}$$

where $r_{before} = r_0 + r_1$ in the perfectly revealing equilibrium, and $r_{before} = 0$ in the pooling equilibrium.

After the referendum, equilibrium behavior can be expressed as $a_i = r_{after}\alpha_i + (1 - r_{after})\,P^d$ and, hence, aggregate behavior in district $d$ is,[23]

$$\overline{a}^d_{after} = P^d.$$

The behavioral change generated by the referendum in district $d$ is thus equal to,

$$\Delta\overline{a}^d = \overline{a}^d_{after} - \overline{a}^d_{before} = (1 - r_{before})\left(P^d - \overline{\mu}\right) \tag{40}$$

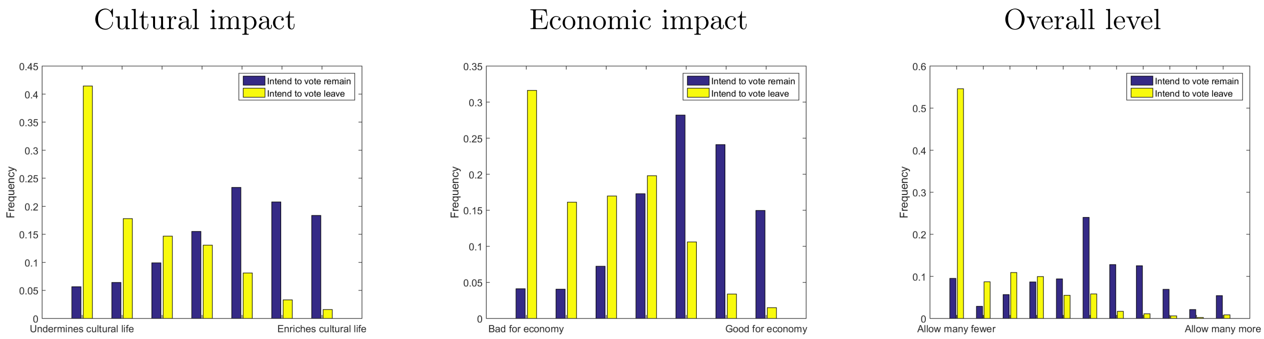
*increasing* in $P^d$. This of course contradicts our finding that districts with stronger dislike towards immigrants experienced a smaller behavioral response to the referendum (Observation 2).

[23]Similar to the pre-referendum equilibrium, the equilibrium after the referendum may be fully separating or pooling. Although it is straightforward to show that the pooling equilibrium where $a_i = \overline{\mu}$ does not survive D1 post-referendum, it may be replaced by a pooling equilibrium where $a_i = P^d$.

# C Quantitative Appendix

## C.1 Views on immigration

Figure 6: Views on immigration by voting intention



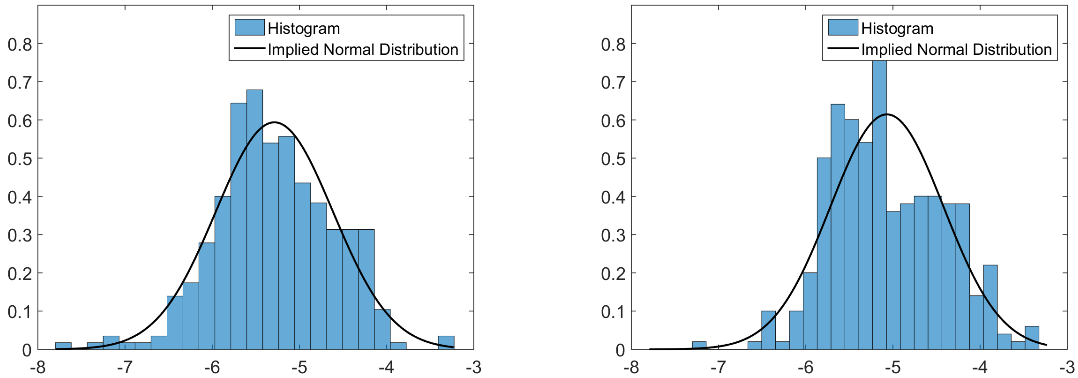

**Note:** Data are taken from the British Election Survey, waves 7 and 8, 2015-2016. Respondents' perceived level of the cultural and economic impact of immigration is rated on an integer scale between 1 and 7. Ranging from immigration '*undermines*' to '*enriches cultural life*' and are '*bad*' to '*good for economy*', respectively. For the overall level of immigration, the integer scale ranges from 0 to 10, 0 represents the country should '*allow many fewer*' and 10 '*allow many more*'. In all cases, the response '*don't know*' has been ignored, which never counted for as much as 10% of all responses.

## C.2 Distribution of Hate Crime

Figure 7: Normality of Hate Crime

Pre-Brexit Referendum

Post-Brexit Referendum




**Note:** Data are reported hate crime divided by foreign born population of a district in the financial year preceding the referendum and the financial year of the Brexit referendum. Normal curves are fiteed with the same mean and variance as the raw data. Pre-referendum data has skewness equal to -0.11 and kurtosis to 3.44. Post-referendum data has skewness equal to 0.17 and kurtosis to 2.75.

## C.3 Betting Markets

**Bookmaker's Probability of the Referendum Result**

| | **Odds** | | **Implied Probability** | | **Normalized Probability** | |
|---|---|---|---|---|---|---|
| **Bookmaker** | **Remain** | **Leave** | **Remain** | **Leave** | **Remain** | **Leave** |
| Skybet | 1/9 | 11/2 | 0.90 | 0.15 | 0.85 | 0.15 |
| Boylesports | 1/10 | 6/1 | 0.91 | 0.14 | 0.86 | 0.14 |
| Betfred | 1/9 | 11/2 | 0.91 | 0.15 | 0.86 | 0.14 |
| Sportingbet | 1/10 | 11/2 | 0.91 | 0.15 | 0.86 | 0.14 |
| BetVictor | 1/12 | 7/1 | 0.92 | 0.13 | 0.88 | 0.125 |
| Paddy Power | 1/7 | 9/2 | 0.88 | 0.18 | 0.83 | 0.17 |
| Stan James | 1/7 | 9/2 | 0.88 | 0.18 | 0.83 | 0.17 |
| 888 Sport | 2/19 | 11/2 | 0.90 | 0.15 | 0.85 | 0.15 |
| Ladbrokes | 1/10 | 6/1 | 0.91 | 0.14 | 0.86 | 0.14 |
| Coral | 1/9 | 5/1 | 0.90 | 0.17 | 0.84 | 0.16 |
| William Hill | 1/8 | 5/1 | 0.89 | 0.17 | 0.84 | 0.16 |
| Sports Winner | 1.1 | 6.5 | 0.91 | 0.15 | 0.86 | 0.14 |
| Betfair | 1/9 | 6/1 | 0.90 | 0.14 | 0.86 | 0.14 |
| Unibet | 1.11 | 6.5 | 0.90 | 0.15 | 0.86 | 0.14 |
| Marathon Bet | 9/100 | 32/5 | 0.92 | 0.14 | 0.87 | 0.13 |
| Betfair Exchange | 1.14 | 7.8 | 0.88 | 0.13 | 0.87 | 0.13 |
| Betdaq | 1.14 | 7.6 | 0.88 | 0.13 | 0.87 | 0.13 |
| Matchbook | 1.128 | 7.3 | 0.89 | 0.14 | 0.87 | 0.13 |
| **Mean** | | | | | 0.86 | 0.14 |

**Note:** Odds are given as displayed by the particular bookmaker either in the *traditional* fractional format or in the *modern* decimal form. In fractional form the profit from a bet of stake equal to the numerator equals the denominator. In decimal form the total returns of a bet (profit plus stake) equals the stake multiplied by the decimal odds. From these returns the implied probability is computed. Since the sum of implied probabilities exceeds one as the bookmaker takes a profit in expectation these are normalized by dividing by the sum of implied probabilities. Odds are taken on the day of the referendum.

## C.4 Identification

To make the case for identification we aim to show that given the vector of moments $\vec{m}$ used in estimation a unique $\vec{\theta}$ of structural parameters can be inferred. The vector of structural parameters are as described in the main text and the vector of moments are listed below, represented in the order they are listed in Table 8.

$$\vec{\theta} := (\theta, \Theta, \mu, \overline{\mu}, b_0, b_1, \sigma_d, \sigma_\epsilon)$$
$$\vec{m} := \Big(E_d(a_0^d), \mathrm{var}_d(a_0^d), \pi, E_d(a_1^d), \mathrm{var}_d(a_1^d), \mathrm{cov}(a_1^d, v^d), E_d(v^d), \mathrm{var}_d(v^d), \mathrm{cov}_d(a_0^d, a_1^d)\Big)$$

The structure of the argument is as follows. Conditional on the parameter $\theta$, the measure of conformity, we show that all other parameters are identified. Following this we make a numerical argument to show that $\theta$ itself is also identified.

From Proposition 2, $\mu$ is immediately observable as $\mu = E_d(a_0^d)$. Rearranging Result 2 gives an expression for $\Theta$ where $r := \mathrm{Var}_d(a_1^d)/\mathrm{Var}_d(a_0^d)$. Notice, this is conditional on $\theta$.

$$\Theta = \frac{(\theta - \theta\sqrt{r})(n(1-(1-\theta)))}{n\theta((1-\theta)\sqrt{r} - (\theta - \theta\sqrt{r}))}$$

From this $\overline{\mu}$ can be computed as

$$\overline{\mu} = \frac{E_d(a_0^d) - (\theta + \gamma)\mu}{1 - \theta - \gamma}.$$

Where $\gamma$ is defined as in the main body of the paper. Thus we are now in a position to define the mapping between attitudes and the innate taste parameters $\alpha$.

$$P^d = \begin{cases} \frac{a_0^d - (1-\theta-\gamma)\overline{\mu}}{\theta+\gamma} \\ \frac{a_1^d - (1-\theta)\mu}{\theta} \end{cases} \quad \rightarrow \quad \sigma_d^2 = \begin{cases} \frac{\mathrm{var}(a_0^d)}{(\theta+\gamma)^2} \\ \frac{\mathrm{var}(a_1^d)}{\theta^2} \end{cases}$$

Notice, that the relationship depends on whether we use pre- or post-Brexit data. In theory, if the model is correctly specified these will coincide. In practice it is likely for some error to occur. Given the linear voting rule $b_1$ can be inferred as

$$b_1 = \frac{\mathrm{cov}(P^d, v^d)}{\mathrm{var}(P^d)} = \frac{\mathrm{cov}(a_1^d, v^d)}{\theta\sigma_d^2}.$$

Taking the expectation and variance operators through the linear voting rule and rearranging gives us an expression for the structural parameters $b_0$ and $\sigma_\epsilon$

$$b_0 = E_d(v^d) - b_1 E_d(P^d) \qquad \rightarrow \quad b_0 = E_d(v^d) - b_1\overline{\mu}$$
$$\sigma_\epsilon^2 = \mathrm{var}_d(v^d) - b_1^2 \mathrm{var}_d(P^d) \qquad \rightarrow \quad \sigma_\epsilon^2 = \mathrm{var}_d(v^d) - b_1^2\sigma_d^2$$

Finally, the moment $\pi$, the probability the forecaster anticipates Brexit helps to pin down the parameter $\theta$. The distribution of expected remain votes is normal and hence symmetric. If the mean lies to the right of $S$ then remain winning is the more likely. The parameter $\theta$ however will impact the first two moments of this distribution. The mean, $b_0 + b_1\overline{\mu}$, is increasing in $\theta$ and so clearly that will increase the ex ante probability of remaining. The variance is composed of two terms, the second of which is tiny at our estimated parameter values. Since $\Theta$ is decreasing in $\theta$, conditional on $E_d(v^d) > S$ a mean preserving spread makes remaining less likely.

## References

ADENA, M., R. ENIKOLOPOV, M. PETROVA, V. SANTAROSA AND E. ZHURAVSKAYA, "Radio and the Rise of the Nazis in Prewar Germany," *The Quarterly Journal of Economics* 130 (2015), 1885–1939.

ADRIANI, F. AND S. SONDEREGGER, "Trust, trustworthiness and the consensus effect: An evolutionary approach," *European Economic Review* 77 (2015), 102–116.

AKERLOF, G. A., "A theory of social custom, of which unemployment may be one consequence," *The quarterly journal of economics* 94 (1980), 749–775.

ALESINA, A., R. BAQIR AND W. EASTERLY, "Public goods and ethnic divisions," *Quarterly Journal of Economics* 114 (1999), 1243–1284.

ALESINA, A., P. GIULIANO AND N. NUNN, "On the origins of gender roles: Women and the plough," *The Quarterly Journal of Economics* 128 (2013), 469–530.

ALESINA, A. AND E. LA FERRARA, "Participation in heterogeneous communities," *Quarterly Journal of Economics* 115 (2000), 847–904.

———, "Who trusts others?," *Journal of Public Economics* 85 (2002), 207–234.

ALGAN, Y., C. HÉMET AND D. LAITIN, "The social effects of ethnic diversity at the local level: A natural experiment with exogenous residential allocation," *Journal of Political Economy* 124 (2016), 696–733.

ANGELETOS, G. M. AND A. PAVAN, "Efficient use of information and social value of information," *Econometrica* 75 (2007), 1103–1142.

AULD, T. AND R. LINTON, "The behaviour of betting and currency markets on the night of the EU referendum," *International Journal of Forecasting* (2019), 371–389.

BECKER, G., "The Economics of Discrimination," *The University of Chicago Press, Chicago & London* (1957).

Becker, S., T. Fetzer and D. Novy, "Who voted for Brexit? A comprehensive district-level analysis," *Economic Policy* 32 (2017), 601–650.

Becker, S. O. and F. Fetzer, "Does Migration Cause Extreme Voting," *CAGE Working Paper* 306 (2016).

Bisin, A. and T. Verdier, "The economics of cultural transmission and socialization," in *Handbook of social economics,*volume 1 (Elsevier, 2011), 339–416.

Bloom, N., P. Bunn, S. Chen, P. Mizen, P. Smietanka, G. Thwaites and G. Young, "Brexit and uncertainty: insights from the Decision Maker Panel," *Fiscal Studies* 39 (2018), 555–580.

Bursztyn, L., G. Egorov, R. Enikolopov and M. Petrova, "Social media and xenophobia: evidence from Russia," (2019).

Bursztyn, L., G. Egorov and S. Fiorin, "From extreme to mainstream: How social norms unravel," Technical Report, National Bureau of Economic Research, 2017.

Card, D., C. Dustmann and I. Preston, "Immigration, wages and compositional amenities," *Journal of the European Economic Association* 10 (2012), 78–119.

Clarke, G. M., H. and P. Whiteley, "Why Britain voted for Brexit: An individual-level analysis of the 2016 referendum vote," *Parliamentary Affairs* 70 (2017), 439–464.

Corcoran, H., D. Lader and K. Smith, "Hate Crime, England and Wales, 2014/15," *Statistical Bulletin, Home Office* (2015).

Costa, R., S. Dhingra and S. Machin, "Trade and worker deskilling," Technical Report, National Bureau of Economic Research, 2019.

Devine, D., "The UK referendum on membership of the European Union as a trigger event for hate crimes," *mimeo* (2018).

Elster, J., "Social norms and economic theory," *Journal of economic perspectives* 3 (1989), 99–117.

Facchini, G. and A. M. Mayda, "Who is against immigration? A cross-country investigation of individual attitudes toward immigrants," *The Review of Economics and Statistics* 91 (2009), 295–314.

Fernandez, R., "Women, work, and culture," *Journal of the European Economic Association* 5 (2007), 305–332.

Fetzer, T., "Did austerity cause Brexit?," *Warwick Economics Research Papers* (2018).

———, "Did austerity cause Brexit?," (2019).

Freeman, R., "The economics of crime," *Handbook of Labor Economics* 3 (1999), 3529–3571.

Giuliano, P., "Living arrangements in western europe: Does cultural origin matter?," *Journal of the European Economic Association* 5 (2007), 927–952.

Goodwin, M. and C. Milazzo, "Taking back control? Investigating the role of immigration in the 2016 vote for Brexit," *The British Journal of Politics and International Relations* 19 (2017), 429–433.

Gourieroux, C., A. Monfort and E. Renault, "Indirect Inference," *Journal of Applied Econometrics* 8 (1993), S85–S118.

Grout, P., S. Mitraille and S. Sonderegger, "The costs and benefits of coordinating with a different group," *Journal of Economic Theory* 160 (2015), 517–535.

Katz, D., F. H. Allport and M. B. Jenness, "Students' attitudes; a report of the Syracuse University reaction study.," (1931).

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

Lazear, E. P., U. Malmendier and R. A. Weber, "Sorting in experiments with application to social preferences," *American Economic Journal: Applied Economics* 4 (2012), 136–63.

List, J. A., "On the interpretation of giving in dictator games," *Journal of Political economy* 115 (2007), 482–493.

Mayda, A. M., "Who is against immigration? A cross-country investigation of individual attitudes toward immigrants," *The Review of Economics and Statistics* 88 (2006), 510–530.

Meleady, R., C. Seger and M. Vermue, "Examining the role of positive and negative intergroup contact and anti-immigrant prejudice in Brexit," *British Journal of Social Psychology* 56 (2017), 799–808.

Morris, S. and H. Shin, "Social Value of Public Information," *American Economic Review* 92 (2002), 1521–1534.

Müller, K. and C. Schwarz, "Fanning the flames of hate: Social media and hate crime," *Available at SSRN 3082972* (2019a).

———, "From hashtag to hate crime: Twitter and anti-minority sentiment," *Available here: https://ssrn. com/abstract* 3149103 (2019b).

Norris, P., "Understanding Brexit: Cultural resentment versus economic grievances," *Harvard Kennedy School Working Paper Series* 18-021 (2018).

Putnam, R., "E Pluribus Unum: Diversity and Community in the Twenty-first Century The 2006 Johan Skytte Prize Lecture," *Scandinavian Political Studies* 30 (2007), 137–174.

Schilter, C., "Hate Crime after the Brexit Vote: Heterogeneity Analysis based on a Universal Treatment," *mimeo* (2018).

Vanberg, C., "A short note on the rationality of the false consensus effect," (2019).

Voigtländer, N. and H.-J. Voth, "Nazi indoctrination and anti-Semitic beliefs in Germany," *Proceedings of the National Academy of Sciences* 112 (2015), 7931–7936.

Wilcox, N. T., "Theories of Learning in Games and Heterogeneity Bias," *Econometrica* 74 (2006), 1271–1292.