RedNIE
Red Nacional de Investigadores en Economía

Updating the Social Norm: the Case of Hate Crime after the Brexit Referendum

Facundo Albornoz (University of Nottingham/CONICET/CEPR)

Jake Bradley (University of Nottingham/IZA)

Silvia Sonderegger (University of Nottingham)

DOCUMENTO DE TRABAJO N° 203

Diciembre de 2022

Citar como:
Albornoz, Facundo, Jake Bradley y Silvia Sonderegger (2022). Updating the Social Norm: the Case of Hate Crime after the Brexit Referendum. *Documento de trabajo RedNIE N°203*.

Updating the social norm: the case of hate crime after the Brexit referendum[a]

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

November 13, 2022

Abstract

Within the context of the Brexit referendum, we show that changes in the perception of social norms impact behavior. The referendum revealed new information about views over immigrants at country level. This new information caused a shift in the social norm which made xenophobic expressions more acceptable. At the margin, some of these expressions involve hate crime. We argue that the post-referendum behavioral change increased in the level of surprise at the referendum result, and that observed geographical variations of the effect depend on underlying local views on immigrants. Survey data corroborate these uncovered facts and support our theoretical mechanism.

Keywords: social norm; social acceptability; xenophobia; value of information; social interactions; referendum

[a]We thank Roland Bénabou for his valuable insights on a previous version of this paper. We are also grateful to Fabrizio Adriani, Gilat Levy, Daniel Seidmann, Francesco Squintani, and specially and Leonardo Bursztyn, Antonio Cabrales, Giacomo Corneo, Guillermo Cruces and Carlo Schwarz, for valuable comments and suggestions, as well as seminar respondents at University of Bergen (NHH), 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, Nada Abdelghany and Marco Vaitilingam provided outstanding research assistance.

[b]University of Nottingham, CONICET and CEPR; 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 overt attitudes (more generally, behavior) 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 (Bénabou and Tirole, 2011b; Bursztyn, Egorov and Fiorin, 2020). Therefore, events that change perceptions about others can lead to updates in what people believe to be socially acceptable. In turn, these updated perceptions translate into observable shifts 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. We show how the post-referendum observed patterns of hate crime and expressions of anti-immigrant views are explained by people updating their perceptions of the country-level social norm.

There is broad agreement (Meleady, Seger and Vermue (2017), Clarke and Whiteley (2017), Goodwin and Milazzo (2017), Becker, Fetzer and Novy (2017)) that animus towards immigrants played an important role in shaping support for the leave option in the referendum.[1] In the immediate aftermath of the vote, social media outlets started denouncing episodes of intolerance and abuse towards immigrants. 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 released by the referendum triggered a change in overt attitudes towards immigrants.

We consider different expressions of this change. One is in terms of people's inclination to vocalize their views on immigrants. The second involves expressions that are so extreme to the

[1] Fetzer (2019), for example, 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.

point of being classified as hate crime. The advantage of using hate 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. The sharp rise in hate crime in the aftermath of the Brexit referendum was widely denounced by the media and was documented more rigorously by Cavalli (2019) and Devine (2021). At the same time, hate crime is clearly only the tip of the iceberg when it comes to overt anti-immigrant and anti-minority attitudes. For this reason, we collected additional survey data from a sample of the British public, in order to gather a more nuanced understanding of the effect of the referendum. With this data, we confirm a shift in overt attitudes towards immigrants consistent with the post-referendum spike in hate crime and go deeper in understanding the mechanisms behind behavioral changes after the referendum.

In general, publicly expressed (or overt) 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 change in overt attitudes after the referendum is that British people changed their views on immigrants overnight. This would imply that the observed changes in behavior reflect a shift in preferences. An alternative and competing interpretation is that the referendum result legitimized previously sanctioned views towards immigrants to be expressed publicly. In this interpretation, the surge in racially or ethnically motivated crime and more generally in overt attitudes towards immigrants 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.

Our empirical analysis is facilitated by the unique features of the Brexit referendum. First, as mentioned, the referendum result was informative of the country's privately held views on immigrants. Second, and most importantly, a crucial aspect of the vote is that, by itself, it had no immediate impact on legislation and policies.[2] In spite of this, the referendum had an abrupt

[2]However, the referendum did have an immediate effect on the exchange rate (e.g. Costa, Dhingra and Machin, 2019) and, even if not immediately, on the prospects of the British economy (Bloom, Bunn, Chen, Mizen, Smietanka, Thwaites and Young, 2018). While an economic-based complementary explanation could in principle explain a jump in hate crime, it can hardly replicate the geographical variation that we uncover and explain.

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.

Our analysis reveals that the increment in hate crime was more pronounced in areas with a higher share of remain votes, in spite of these areas having generally lower incidents of hate crime per immigrant. In line with existing literature as well as data from the British Election Study (BES) collected shortly before the referendum, we argue that the share of remain and leave vote in an area is informative of its inhabitants' personal views on immigrants. Our results thus suggest that the increment in hate crime triggered by the referendum was more pronounced in areas which are, ceteris paribus, more pro-immigrant. Directly regressing the increment in hate crime post-referendum against pre-referendum views on immigration from the BES paints very much the same picture, as do the data collected in our own survey. Among our respondents, those who reported becoming more comfortable in expressing their views on immigrants after the referendum tended to personally dislike immigrants and to live in areas where immigrants are generally well-liked. This sounds almost counter-intuitive but in fact is consistent with an information-based mechanism. 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 expressing anti-immigrant views is not socially acceptable, 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, this will shift your perception of what is socially acceptable at the country level more than if you lived in a very anti-immigrant area. The change in your overt attitudes is thus going to be larger. This 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 very public revelation that anti-immigrant views across the country were more widespread than was previously believed, this caused the perception of the country-level norm to shift, rendering anti-immigrant attitudes more acceptable. Our own survey data as well as the data collected by the British Election Study (BES) 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 shift in overt attitudes) to be larger in those areas.

Of course, our analysis of hate crime 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. Using data from the Crime Survey for England and Wales, we show tha, indeed, 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, Mitraille and Sonderegger (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 their perception of the social norm (average behavior). The relevant social norm depends on average behavior nationally as well as locally. Individuals know the views of people in their own area but not at the country level. Since local views are only partially informative of what people think in other areas, a potential discrepancy between what people *believe to be* the average viewpoint at country level and the *real* viewpoint emerges. An informational

update on the latter can then trigger a change in behavior at local level. This is particularly true when all of a sudden it becomes publicly clear – e.g., through a national referendum – that in the country as a whole many agree with a behavior that was previously thought to be socially unacceptable. At that point, the perception of the norm changes abruptly, and people become considerably more likely to adopt that behavior. The direction and the extent of the behavioral change is a direct function of the discrepancy between previous and updated beliefs about average views at national level. This result can explain sudden changes in overt attitudes that vary across different regions within the same country.

Our model shows that the theoretical predictions on the effect of the referendum across regions differ for the possible mechanisms identified, namely preference-driven and a norm-driven. We show that the greater effect of the referendum in areas with a higher share of remain votes is consistent with the theoretical predictions of the model, and in particular with the norm-driven interpretation of the evidence, as explained above. To investigate the different implications of norm-driven versus preference-driven changes in overt attitudes, 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 animus (preference channel). The key finding here is that the increment in hate crime following the attacks was more pronounced in areas with a higher share of leave vote, where people are less pro-immigrant, 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.[3] This finding is consistent with existing literature. For instance, Müller and Schwarz (2019b) and Bursztyn, Egorov, Enikolopov and Petrova (2019), find that exposure to social media and to Trump anti-immigrant tweets increases hate crime, but only in areas with high pre-existig prejudice. This suggests that areas with stronger anti-immigrant

[3]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, Fetzer and Novy (2017)).

preferences offer more fertile ground for further preference shifts fueled by exposure to xenophobic messages.[4] Based on these observations, we argue that a change in preferences is unlikely to have triggered the post-referendum change in behavior.

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 a little less than 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 the role of misplaced prior beliefs and the resulting the surprise effect of the referendum in generating the observed behavioral change. The estimates allow us to look, in a semi-quantitative way, at a series of thought experiments which may be informative to policymakers. 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. This underscores the importance of shared national narratives promoted for instance by public figures and cultural influencers for actual behavior.

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

[4]Similarly, Adena, Enikolopov, Petrova, Santarosa and Zhuravskaya (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.

redistribute and invest in public goods (Alesina, Baqir and Easterly (1999)). Algan, Hémet and Laitin (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. Kuran (1991) and Lohmann (1994) present early game theoretic analyses of how individuals may adopt public positions that differ from their private preferences, and how, in this context, new information may trigger a "revolutionary bandwagon" – as in the case of the Eastern European Revolution of 1989, which was largely unexpected and took everyone by surprise.

A recent paper by Bursztyn, Egorov and Fiorin (2020) studies the difference in the willingness of experimental subjects from the Pittsburgh metropolitan area to make a donation to a xenophobic organization when the donation is kept private and when it is made public to other individuals living in the same area. Their treatments vary the information people receive about local Trump support, by exploiting the fact that Trump won the election in Pittsburgh metropolitan area while Clinton won the election in Pittsburgh's county. They find that the respondents in the "Clinton won" treatment are more likely to make the donation in the private compared to the public treatment, while those in the "Trump won" treatment are equally likely to donate independently of whether this is kept private or made public. The driver of this effect is the desire by individuals to signal that their preferences are aligned with those of the people around them (the "*observers*"). We offer a complementary perspective by emphasizing conformity with the imperfectly observed national norm.[5] Even in a context where local preferences are constant and (locally) well-known, updates of the perceived national norm can have an impact on expressed attitudes towards immigrants. Highlighting the importance of "national norms" is useful to understand the geographical variation of changes in attitudes after a preference-revealing event such as a referendum.

More broadly, we contribute to the literature on the economics of norms and culture by docu-

[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.

menting 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, Giuliano and Nunn (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, Dustmann and Preston (2012)) and the nascent literature that focuses specifically on hate crime (Bursztyn, Egorov, Enikolopov and Petrova, 2019; Müller and Schwarz, 2019a,a,b). 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 the studies by Cavalli (2019), Devine (2021), and Schilter (2018). Using aggregate UK data as well as data for Manchester and London, these studies document a substantial 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. Finally, we follow the approach of previous empirical studies that attempt to understand the causes of surges in hate crime, be they terrorist attacks (Hanes and Machin (2014) and Ferrin et al. (2020)), emotive crimes (Frey (2020)), or a pandemic Dipoppa et al. (2020).

2 Research strategy

We have two main sources of data: (i) administrative data on hate crime and referendum results across UK regions (community safety partnerships, CSP henceforth) and (ii) our own survey data.

Administrative data allow us to establish facts on the relationship between the increment in hate crime and voting patterns across the United Kingdom. As suggested by commentators and the associated literature, there is a strong connection between the referendum vote and anti-immigration sentiment. However, there is no data on immigration views at the exact moment of the referendum.

For this reason, we deploy different strategies to mitigate this concern. First, we exploit a question from the 2015 wave of the British Election Study (BES) – a long-running UK-wide panel study with approximately 30,000 respondents – which asks respondents how much they think immigration should be decreased or increased. This question allows us to build an index of pre-referendum anti-immigration sentiment at the CSP level. As expected, we find a very strong negative correlation (-0.73) between the index of anti-immigration sentiment and the share of Remain votes. We also re-run our main regression replacing the share of Remain votes with the anti-immigration sentiment index. Our results confirm that hate crime increased more in areas with relatively more favorable views on immigration.

To complement our first set of findings, we develop a theoretical model and show how a positive correlation between voting and immigration views explains all our facts and produces new predictions associated with the underlying mechanisms. To further test these mechanisms as well as the connection between post-referendum changes in overt attitudes towards immigrants and views on immigration, we also ran our own survey. As we will see, the data we obtained provide further support to the facts uncovered and the mechanisms proposed by the theory.

Finally, to have a sense of how the theory can quantitatively replicate the patterns we observe, we estimate the parameters of the model and discuss their relevance and implications.

Accordingly, the paper is organized as follows. In section 3 we start by stating the main empirical facts about hate crime and the changes induced by the Brexit referendum. Section 4 presents and solves a model capable of reconciling these facts with theory by assuming that voting patterns are positively correlated with immigration views. Section 5 presents supportive evidence of the underlying mechanisms of the model. In section 6, we demonstrate that a quantitative version of the model can generate the sizes of the phenomena discussed with a *sensible* calibration. Section 7 concludes.

3 Hate crime and referendum voting patterns

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.8. We now turn to discuss our data.

3.1 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 safety 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. To better understand the size of a CSP and the spatial distribution of the vote, see Figure A.1 in the Appendix.

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	3.6×10^3	3.3×10^3
Remain Votes (share)	cross-sectional	2016q1	ONS	0.457	0.101
National Insurance Registrations (EU)	quarterly	2002q1 - 2017q2	DWP	2.7×10^2	4.6×10^2
National Insurance Registrations (non-EU)	quarterly	2002q1 - 2017q2	DWP	1.8×10^2	3.4×10^2
Population (1000s)	annual	2002 - 2016	ONS	1.8×10^2	1.2×10^2
Gross Disposable Household Income (millions of £)	annual	2002 - 2016	ONS	2.9×10^3	1.9×10^3
Gross Value Added by Sector⋆ (millions of £)	annual	2002 - 2016	ONS	-	-
Social Benefits Received (millions of £)	annual	2002 - 2016	ONS	7.7×10^2	5.3×10^2

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

Included in our data are any criminal offense reported to the police and which 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, Lader and Smith (2015) estimate that approximately 48 per cent of all incidents of hate crime come to the attention 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 3.4.1. 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. It is also smaller than the 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.

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 (2019) and Norris (2018)) and a large literature on the economic motivation

[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.

of crime in general (see Freeman (1999) and the references therein). To control for changes in economic conditions, we include the following Variables at CSP level: 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.

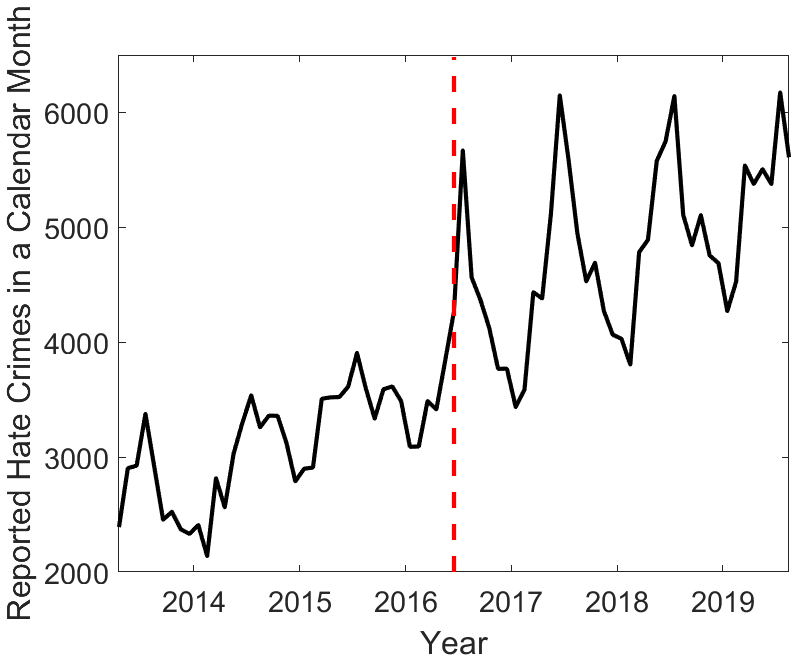
3.2 The evolution of hate crime in the UK

The sharp increase in hate crime in England, Wales and Northern Ireland right after the referendum vote is well-established. In May 2018, the UN Special Rapporteur on Racism stated *"...the environment after the referendum has made racial and ethnic minorities more vulnerable to racial discrimination and intolerance"* (Human Rights, 2017). Such statement echoed substantial evidence documented by the press.

A number of academic papers substantiate these claims. For example, Cavalli (2019) finds that hate crime "unexpectedly" increased by 19.2% after the referendum. Devine (2021) runs a time-series analysis confirming that the Brexit referendum caused the observed spike in hate crimes (between 19% to 23%). The evolution of hate crime across time is depicted in Figure 1, with the red dashed line indicating the date of the referendum. Cavalli (2019) and Devine (2021) also document a spike in hate crime immediately after the Manchester Arena bombing on May 22nd, 2017 (the so-called "Summer of Terror" of 2017).

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

Figure 1: Police Reported Hate Crime



Note: Data are provided by the Home Office, detailing all crimes reported to the 43 police forces of England and Wales and run from April 2013 to August 2019, inclusive. The solid black line indicates the number of hate crimes reported to the police in a calendar month and the red dashed line indicates the time of the referendum.

Before moving on to analyze the geographical variation of these spikes. 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.[8]

3.3 Observation 1: the post referendum surge in hate crime was higher in areas with a larger remain vote

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 with a

[8] 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.

higher share of Remain votes. In order to be as transparent as possible, we propose the following simple regression specification.

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

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 (τ_t) and CSP fixed effects (η_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 3.4, 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. All specifications include quarter dummies to remove any seasonality from the data. Crimes in general happen more frequently in the summer months. Appendix A.4 argues that there do not appear to be any systematic differences in seasonality that would confound our results. Subsequent columns include a proxy for the overall criminality of an area, a police force level trend component, and finally in the specification reported in column 6, 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 percent. While the estimates are statistically significant to visualize the economic significance we plot

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

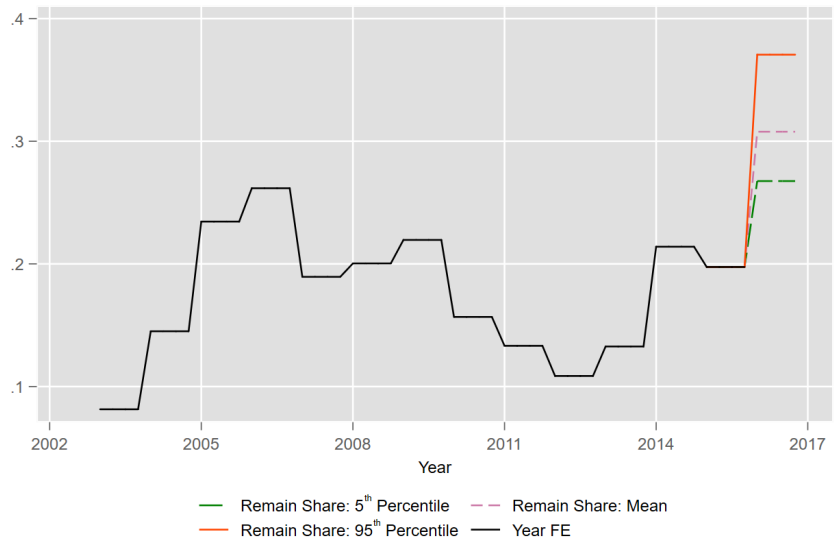
as estimated in column (4), in Figure 2. The change in the fixed effect in the first financial quarter

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

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 percent increase in hate crime for the average CSP. This increase is approximately equivalent to the gap between the fifth and ninety-fifth percentile.

Figure 2: $\hat{\tau}_t + \hat{\beta}\left(1_{\{\text{Post Brexit}\}} \times \text{remain}_i\right)$



One broad concern of our approach is that an individual's voting decision is based on a multitude of competing factors, rather than just immigration or xenophobic attitudes. To alleviate these concerns and support the mechanism put forward in the theoretical model, we ran an identical specification to that outlined in (1) replacing our variable of interest. Rather than the share of the population who voted remain we constructed a measure of anti-immigration sentiment by CSP based on wave 6 of the British Election Study (BES). The study was conducted on a sample of 30,073 respondents in May of 2015, just over a year before the referendum. Hence, it represents an approximate snapshot of preferences at the time of the referendum. Respondents were asked whether they believed the level of immigration to be too high, too low, or correct. We geocoded the location of the respondent to our unit of observation (CSP) and computed the share of respondents in each CSP that find the current level of immigration to be too high. Inference of these shares are based on approximately one hundred individuals per unit of observation and are well correlated (-0.73) with the share voting remain. The advantage of constructing our index as a proportion allows us to estimate the same quasi-elasticity as in the baseline specification. The estimated coefficients represent the percentage increase in the rate of hate crime from a one percentage point increase in the share averse to immigration. The parameter estimates are presented in Table A.3. Encouragingly,

not only are the estimates similar qualitatively, in that more anti-immigration areas saw larger increases following the referendum. But they are also similar quantitatively. The responsiveness of hate crime to individual preferences to migration are in the same order of magnitude to the referendum voting patterns.

Finally, 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 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.

3.4 Robustness

To further 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.5 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 A.1 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. As can be seen in Table A.6, our findings are robust to a reduced sample omitting London. (iii) The final robustness check concerns the specification used. Firstly, 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.7 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. Finally, in section 3.4.2, we implement a different empirical strategy. We analyze the change in hate crime relative to similar crimes before and after the referendum. Our results are quantitatively similar to the baseline presented previously.

3.4.1 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. Following Soares (2004) we attempt to correct for misreporting by comparing administrative data on reported crimes with a victimization survey. 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.[9]

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

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 crime, thus increasing the propensity to report said crime. In addition, the composition of crimes

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

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 ω_{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.8.

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

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 9 , 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.

Under the hypothesis that reporting rates vary primarily by the police force and not victim reporting we compare the force-year fixed effects in the baseline model with those looking at all other crimes, Table A.5. Appendix A.9 plots the distribution of these force-year fixed effects. The first panel shows the evolution of the mean effect by force over the estimation window. In the lead up to the referendum and at the implementation, both series are relatively stable implying not much systematic changes in how police forces categorize hate crime. This is consistent with our findings on data corrected for misreporting. However, at the start of our window we see an increase over time for hate crime and declining in other crimes. One potential reason for this is hate crime legislation was in its infancy at this point and forces took some time to correctly categorize crimes.[10] The second point of interest is the dispersion of these fixed effects. The remaining two panels in Appendix A.9 show the distribution of the force-year fixed effects with log of hate crime and log of all other crimes as the dependent variables. By construction, both distributions are centered around zero but exhibit very different second moments. The second moment when log hate crime is the dependent variable is twice that of other crimes — the associated standard deviations are 0.419 and 0.292, respectively. Since we control for permanent local area effects one could conjecture that this difference implies that there is some variation in whether or not a crime is categorized as a hate crime that depends on which force's jurisdiction it falls in. However, as we have seen this variation

[10]What constitutes a hate crime was first documented in sections 28-32 of the Crime and Disorder Act, 1998.

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

appears orthogonal to changes around the referendum.

3.4.2 Difference-in-difference

To add further credence to our main empirical finding, we consider an alternative empirical strategy. Under the identifying assumption that crimes absent a hate component are unaffected by the referendum we can implore a difference-in-difference approach. We take the same category of crime and compare the different trajectory of those deemed to be motivated by hate by those without a hate motivation and evaluate the change either side of the referendum.

Table 5: Dependent Variable: Reporting Rates

	(1)	(2)	(3)	(4)	(5)	(6)
Lag Dep. Variable						0.002
						(0.004)
Post Brexit ×	**0.012***	**0.014****	**0.018*****	**0.018*****	**0.005**	**0.005**
Remain Share	**(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

Table 6: Dependent Variable: Log Hate Crime

Crime category	Mean		Stand. dev.	
	$h = 0$	$h = 1$	$h = 0$	$h = 1$
Assault without injury	733.4	13.5	680.7	19.0
Criminal damage	2517.4	11.3	2275.4	18.8
Assault with injury	1240.5	12.4	1143.0	21.2
Public order/harassment	614.0	77.5	673	101.8

Note: Mean and standard deviations are across financial year and community safety partnership area. Years from 2002 to 2016, inclusive and there are 315 areas, making each moment based 4,725 observations. Numbers have been rounded to one decimal place.

To fix ideas, suppose a crime varies in two discrete dimensions, its category c and whether or not it is deemed a hate crime $h \in \{0, 1\}$. Where h takes the value one for a hate crime, and zero otherwise. We are now looking at more disaggregated data, in that we differentiate by crime category. Therefore in order to avoid excessive sparsity we switch from a quarterly to an annual frequency. The first two moments of the number of crimes in a financial year across CSP are reported in Table 6. Across the four crime categories, those relating to criminal damage and assault see a low proportion associated with hate — between 0.44% and and 1.8% of all crimes. By contrast, a larger proportion of public order and harassment offenses are deemed to be hate crimes — 11% on average.

To understand the differential impact across the remain share of a region we implement, as before, a heterogeneous treatment approach. Unlike before, we additionally distinguishing between categories of crime and take differences across whether the crime was motivated by prejudice. Equation (2) describes the specification, where α is the shift parameter, common to all areas. It describes the increase in hate crime following the referendum, irrespective of the voting outcome of the particular CSP. The parameter β is the parameter of interest and represents the percentage increase in hate crime following a 1% increase in the remain share within a CSP. The second line of the expression are the controls and include: a matrix of covariates; a quadratic time trend; CSP fixed effects; and dummies for whether the crime is motivated by prejudice; and the specific category of crime.

$$\begin{aligned} y_{icht} = \alpha \left(1_{\{\text{Post Brexit}\}} \times 1_{\{h=1\}} \right) + \beta \left(1_{\{\text{Post Brexit}\}} \times 1_{\{h=1\}} \times \text{remain}_i \right) \\ + \gamma \mathbf{X}_{it} + \tau(t) + \eta_i + \theta_h + \psi_c + \epsilon_{it} \end{aligned} \tag{2}$$

Table 7 reports the parameter of interests, sequentially including the same controls as were used previously. The first thing to notice is that on average an area can expect an uptake in hate crime following the referendum.[11] Across all four specifications, the parameter of interest β is positive and significant. Our preferred specification is estimated at 0.77. This implies that an area with a

[11] The mean vote of an area in our sample is 0.457. Hence, based on our parameter estimates, the average increase in hate crime as a consequence of the referendum can be computed as $\alpha + 0.457\beta$.

10% higher remain vote can expect a 7.7% higher increase in hate crime following the referendum. Although larger, the numbers are not quantitatively that different to the estimated parameter in the baseline specification reported in Table 2.

Table 7: Dependent Variable: Log of Crime

	(1)	(2)	(3)	(4)
α (shift parameter)	-0.296	-0.408	-0.279	-0.313
	(0.111)	(0.111)	(0.112)	(0.112)
β (heterogeneous treatment)	0.925	1.061	0.759	0.773
	(0.235)	(0.235)	(0.237)	(0.238)
Migration controls	NO	YES	YES	YES
Economic controls	NO	NO	YES	YES
Sectoral composition	NO	NO	NO	YES
Observations	37,800	37,800	37,800	37,800
R-squared	0.883	0.884	0.884	0.885

Note: In addition to the controls listed each specification includes: dummies for crime categories; a dummy for whether the crime was motivated by hate; CSP fixed effects; and a quadratic time trend. The total number of observations are comprised of the product of: the number of CSPs; the number of years; crime categories; and hate component. There are $315 \times 15 \times 4 \times 2 = 37{,}800$.

A further interesting feature that can be explored with this approach is how hate crime changes by crime category. The specification below takes equation (2) and conditions each parameter on the crime category. In practice the same regression is run for each category listed in Table 6.

$$y_{icht} = \alpha_c \left(1_{\{\text{Post Brexit}\}} \times 1_{\{h=1\}}\right) + \beta_c \left(1_{\{\text{Post Brexit}\}} \times 1_{\{h=1\}} \times \text{remain}_i\right) + \gamma_c \mathbf{X}_{it} + \tau_c(t) + \eta_{ci} + \theta_{ch} + \epsilon_{cit}$$

Parameter estimates are presented in Table 8. Across all crime categories and specifications we find that the level of hate crime rose most in remain areas. However, there is considerable dispersion across crime category. We find that assault on an individual, with or without physical injury, are broadly consistent with our baseline findings. These crimes represent almost a quarter of all hate crimes and the parameter β is estimated as 0.62 (without injury) and 0.69 (with injury). These are comparable to the baseline specification estimate of 0.58, see Table 2. Interestingly, arguably the two less serious crimes of criminal damage and public order offenses show a quite different relationship

with the referendum vote. For criminal damage, there is a much stronger relationship with areas with a higher remain share seeing a much larger rise. By contrast, for public order offenses, while the effect is still there, hate crime rises more uniformly.

Table 8: Dependent Variable: Log of Crime

	(1)	(2)	(3)	(4)
Assault without injury				
α (shift parameter)	-0.215	-0.319	-0.349	-0.389
	(0.124)	(0.125)	(0.125)	(0.125)
β (heterogeneous treatment)	0.347	0.514	0.581	0.618
	(0.263)	(0.264)	(0.266)	(0.266)
R-squared	0.96	0.961	0.961	0.961
Criminal Damage				
α (shift parameter)	-0.447	-0.577	-0.455	-0.493
	(0.128)	(0.127)	(0.128)	(0.128)
β (heterogeneous treatment)	1.612	1.818	1.534	1.563
	(0.27)	(0.27)	(0.271)	(0.271)
R-squared	0.975	0.976	0.976	0.976
Assault with injury				
α (shift parameter)	-0.358	-0.483	-0.302	-0.334
	(0.123)	(0.122)	(0.122)	(0.122)
β (heterogeneous treatment)	0.956	1.104	0.684	0.69
	(0.262)	(0.259)	(0.259)	(0.259)
R-squared	0.969	0.97	0.971	0.971
Public order/harassment				
α (shift parameter)	-0.163	-0.256	-0.011	-0.035
	(0.123)	(0.12)	(0.119)	(0.119)
β (heterogeneous treatment)	0.785	0.809	0.239	0.22
	(0.261)	(0.255)	(0.253)	(0.253)
R-squared	0.909	0.915	0.917	0.918

Note: Specifications in the columns represent the same controls as in Table 7. In each case the number of observations is 9,450.

3.5 Observation 2: the surge in hate crime post 7/7 attacks followed the opposite pattern to the post referendum surge

The 7/7 Islamic fundamentalist terrorist attacks on London in 2005 also generated a spike in hate crime, similar to 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 the spike was more pronounced in more pro-remain areas, quite the opposite. The results presented in Table 9 show that, across all specifications, hate crime increased more in areas with a higher share of *leave* votes, thus exhibiting the opposite pattern to what we found for the case of the Brexit referendum.

To cement this point further, we run a battery of placebo trials. Postulating the same specification, column (6) at every possible time period in our sample. A histogram of the estimates of $\hat{\beta}$ are presented in Figure 3. Amongst the 50 estimated coefficients none are as large nor as significant as the 0.579 estimate associated with the Brexit referendum.

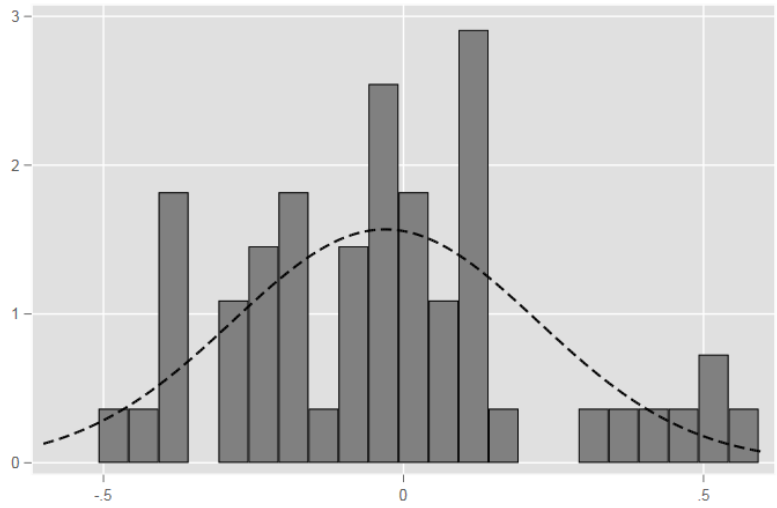
Table 9: Dependent Variable: Log Hate Crime

	(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

3.6 Observation 3: local variations in expected referendum result

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

Figure 3: A Battery of Placebo Trials



Note: The histogram plots the coefficient of interest from column (6) of Table 9 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.

(BES). 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 10. The coefficients of the additional controls are reported in Appendix A.10.

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.

Table 10: 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

Taking stock of the empirical observations

To recapitulate, besides the spike in hate crime in the aftermath of the Brexit referendum documented by the literature, we can organize our original observations in the following way:

Observation 1. *The rise in hate crime was more pronounced in areas with a higher share of remain votes in the referendum.*

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

(b) This result holds omitting London.

Observation 2. *The rise in hate crime in the aftermath of the 7/7 terrorist attack was more pronounced in areas with a higher share of leave votes in the referendum.*

Observation 3. *Before the referendum, people living in different areas of the country held different expectations about the referendum outcome. In more pro-leave areas, people expected higher country-wide support for Brexit.*

In what follows, we develop a framework that can explain these three facts with a minimal structure which can be estimated to inform counterfactual exercises about the effect of changes in preferences and perceptions of the national norm.

4 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. To connect with the previous empirical analysis, we work under the premise that the Brexit vote is correlated with views on immigrants and, therefore, the referendum reveals information about these views at the country level. In our model, individuals care about conforming to their underlying preferences (which arise from their socio-economic characteristics as well as other regional and idiosyncratic components) but also want to conform with the behavior of other individuals – the social norm.

4.1 The Environment

Background We consider a country of measure 1 which contains a continuum of individuals, represented by a real coordinate i on the unit interval $[0, 1]$, who are divided into geographical districts of equal size (for simplicity). All individuals move simultaneously. Each individual i belonging to district d selects his *overt attitude* $a_i \in \mathbb{R}$ to maximize his expected payoff. This depends on (i) how closely the individual's overt attitude matches a preference parameter $\alpha_i \in \mathbb{R}$ reflecting his intrinsic preferences, and (ii) how closely it conforms to a reference behavior $\overline{a}^{n_d}$ defined as $\overline{a}^{n_d} \equiv \lambda \overline{a} + (1 - \lambda)\, \overline{a}^{d}$, where the parameter λ satisfies $0 < \lambda \leq 1$ and $\overline{a}^{d}$ and $\overline{a}$ represent aggregate overt attitudes in district d and country-wide, respectively. We think of the reference behavior $\overline{a}^{n_d}$ as the social norm in district d: it determines which overt attitudes are considered acceptable. Intuitively, adopting a behavior that is faraway from the social norm is susceptible to social pressure/stigma, while a behavior that is close to the social norm attracts social approval. The social norm is shaped by behavior locally as well as country-wide, with the relative importance of the latter being parameterized by λ. The role of country-level aggregate behavior in shaping local norms reflects the exposure of local individuals to country-wide influences through social or traditional media as well as social interactions with people from outside their local area. Note that concern for conforming to the norm need not necessarily arise from the desire to obtain *external* social

approval. It may equally well arise from the internal desire to adopt a behavior that is consistent with one's identity – as e.g. in Akerlof and Kranton (2000); Bénabou and Tirole (2011a).[12] In this interpretation, λ captures the extent to which individuals identify with the country as opposed to their local area.

Importantly, although aggregate overt attitudes are observed *ex-post*, when payoffs are realized, they are not observed *ex-ante*, when agents select their actions. People must therefore use the information at their disposal to form beliefs about the aggregate overt attitudes that will prevail in their district as well as the country as a whole. [Payoffs and information structure are further discussed below.]

Individual preferences Individual preferences are given by the sum of two components: a district-specific component P^d, and an idiosyncratic component, ε_i. The district-specific component is in turn given by the sum of a mean preference parameter μ 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 μ corresponds to mean preferences in the whole country, when specific district characteristics are averaged out. To sum up, therefore, the preference parameter α_i^d of an individual i belonging to district d is equal to

$$\alpha_i^d = \mu + e^d + \varepsilon_i \tag{3}$$

where μ, the (unobservable) mean preference in the whole country, 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 ε_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 aggregate preferences within his district, $P^d = \mu + e^d$, but he is unable to discriminate between μ and e^d, the country–wide component and the district–

[12] We follow e.g., Akerlof (1980); Michaeli and Spiro (2017); Grout, Mitraille and Sonderegger (2015) and Bicchieri, Dimant and Sonderegger (2020) in adopting a consequentialist approach, in the sense that social esteem or self-esteem follows directly from individual behavior and its relationship with the norm. Another branch of the literature, such as Bernheim (1994); Bénabou and Tirole (2011a) or Adriani and Sonderegger (2019), focuses instead on the case where behavior is not approved or stigmatized *per se*, but only to the extent to which it reveals information about an individual's underlying type.

specific shock affecting his preferences. Observing P^d implies that, in equilibrium, people can perfectly anticipate the aggregate overt attitudes that will prevail in their district.[13] This is however not the case for aggregate overt attitudes at the level of the whole country. As far as these are concerned, people form conjectures that depend on their beliefs about μ, aggregate preferences at the country level. Consider an individual i who has observed P^d. Given the normality assumptions, his expectations of μ 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

$$E(\mu \mid P^d) = \frac{\Theta}{1+\Theta} P^d + \frac{1}{1+\Theta} \overline{\mu}. \tag{4}$$

Payoffs The payoff of an individual i belonging to district d is equal to

$$u_i = -\theta \left(a_i - \alpha_i\right)^2 - (1-\theta) \left(a_i - \overline{a}^{n_d}\right)^2. \tag{5}$$

where, as mentioned, $\overline{a}^{n_d} \equiv \lambda \overline{a} + (1-\lambda)\, \overline{a}^d$. The parameter $\theta \in (0,1)$ captures the concern for aligning own behavior with personal preference relative to conforming with the social norm.

4.2 The Equilibrium

Our equilibrium concept is Perfect Bayesian Equilibrium. Each individual i in district d chooses his overt attitude a_i to maximize his expectation of (5), where the expectation is taken with respect to $\overline{a}^{n_d}$. Differentiating the objective function and rearranging delivers i's best-reply,

$$a_i = \theta \alpha_i + (1-\theta)\, E\left(\overline{a}^{n_d}\right) \tag{6}$$

Each individual chooses an overt attitude that is a weighted average of his own personal preferences and what he expects to be the social norm. As is standard in the literature, we focus on linear symmetric strategies. For analytical convenience, we consider the computationally easier case of a continuum of districts.

[13] That's because, in this simultaneous-move game, equilibrium overt attitudes at district level can be perfectly inferred from P^d. In Appendix B we consider the more general case where individuals do not perfectly observe aggregate preference in their district, and instead receive an i.i.d. signal about it. There, we argue that this more general setup may help rationalize the "Scottish anomaly" identified in our data.

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

$$a_i = \theta\alpha_i + \gamma_0 P^d + (1 - \theta - \gamma_0)\overline{\mu} \tag{7}$$

where $\gamma_0 \equiv \frac{\theta(1-\theta)(\Theta+1-\lambda)}{\theta+\lambda(1-\theta)+\theta\Theta} < 1 - \theta$.

An immediate implication of Proposition 1 is that,

Corollary 1. *When μ is unobservable, aggregate overt attitudes are $\overline{a} = (\theta + \gamma_0)\mu + (1 - \theta - \gamma_0)\overline{\mu}$ across the whole country and $\overline{a}^d = (\theta + \gamma_0)\, P^d + (1 - \theta - \gamma_0)\overline{\mu}$ in district d.*

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

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

$$a_i = \theta\alpha_i + \gamma_1 P^d + (1 - \theta - \gamma_1)\mu \tag{8}$$

where $\gamma_1 \equiv \frac{\theta(1-\theta)(1-\lambda)}{\theta+\lambda(1-\theta)} < 1 - \theta$.

Proofs for Proposition 1 and 2 are provided in the Appendix.

Corollary 2. *When μ is publicly observable, aggregate overt attitudes are $\overline{a} = \mu$ across the whole country and $\overline{a}^d = (\theta + \gamma_1)\, P^d + (1 - \theta - \gamma_1)\mu$ in district d.*

4.3 The effect of the referendum on overt attitudes

By itself, the referendum 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. We believe that a key feature of the referendum is that it revealed information about the electorate's private preferences, above and beyond previous information available. Intuitively, suppose that individuals obey the following voting strategy: for some constant $\widehat{\alpha}$, all those with preferences $\alpha_i \leq \widehat{\alpha}$ vote remain while all those with $\alpha_i > \widehat{\alpha}$ vote leave. Clearly enough, once the shares of remain and leave votes across the nation become public information, this perfectly reveals the true value of μ and makes it common knowledge, as in Proposition 2. In contrast, we can think of the pre-referendum environment as one

where, as in Proposition 1, true average preferences are unobserved, although people have access to publicly available information that they use to form beliefs about the private views of others (captured by the common prior $\overline{\mu}$).[14]

Consider first society as a whole. From Corollaries 1 and 2, the change in aggregate overt attitudes towards immigrants before and after the referendum is

$$\overline{a}_{after} - \overline{a}_{before} = (1 - \theta - \gamma_0)(\mu - \overline{\mu}). \tag{9}$$

In what follows, without loss of generality we adopt the convention that higher values of a (resp., α) correspond to more anti-immigrant overt attitudes (preferences).

Result 1. (Change in overt attitudes at country level) *The necessary and sufficient condition for the referendum to induce average overt attitudes across the country to become more anti-immigrant is that*

$$\mu - \overline{\mu} > 0 \tag{10}$$

The condition implies that the referendum increases overt anti-immigrant attitudes if true average preferences (revealed through the referendum) are more anti-immigrant than the ex-ante prior. This result addresses the post-referendum spike in hate crime 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.[15] 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 (9) the magnitude of the change in overt attitudes depends on the size of $1 - \theta - \gamma_0$, which, after substituting for γ_0, can be shown to be decreasing in θ 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 overt attitudes.

[14] Clearly enough, results extend to the case where the referendum simply provides *more* information about μ than was previously available, without perfectly revealing it.

[15] This shares similarities with the phenomenon known in social psychology as *pluralistic ignorance*, namely a situation in which the majority privately disagree with a given behavior, but adopt it since they incorrectly assume that others agree with it (Katz, Allport and Jenness, 1931).

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 is in line with the evidence which points to a sudden spike in hate crime in the immediate aftermath of the referendum.

The role of new information is thus crucial to appreciate the mechanism behind the change in overt attitudes. Indeed, condition (10) makes clear that the effect of the referendum on overt attitudes relies on what we may call a *surprise effect.* Formally, the surprise effect can be measured as the difference between true μ and mean pre-referendum beliefs about μ, which can be easily computed from Bayesian updating,

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

Expressions (9) and (11) show that there is a direct correspondence between the surprise effect and the change in overt attitudes induced by the referendum. Intuitively, the referendum generated a behavioral reaction 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. (Geographical variations in beliefs) *When μ is unobservable, average beliefs about μ in district d is*

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

increasing in P^d.

Proof: *Follows straightforwardly from Bayesian updating.*

Result 2 rationalizes our Observation 3, namely that, prior to the referendum, people in more pro-leave areas believed it more likely that the UK would vote leave. That's because standard Bayesian updating implies that people used preferences in their own area to form beliefs about preferences across the country – a sort of "rational consensus effect."[16] Result 2 has a direct implication for

[16]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

geographical variations of the surprise effect. The difference between true μ and previous average beliefs about μ in district d is

$$\mu - E_d[E(\mu \mid P^d)] = \mu - \frac{\Theta}{1+\Theta}P^d - \frac{1}{1+\Theta}\overline{\mu} \tag{12}$$

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 directly from Propositions 1 and 2.

Result 3. (Change in overt attitudes at district level) *The difference in aggregate overt attitudes in district d before and after the referendum is*

$$\overline{a}^d_{after} - \overline{a}^d_{before} = P^d(\gamma_1 - \gamma_0) + (1 - \theta - \gamma_1)\mu - (1 - \theta - \gamma_0)\overline{\mu} \tag{13}$$

where $\gamma_1 - \gamma_0 = -\frac{\theta\Theta\lambda(1-\theta)}{(\theta+\lambda(1-\theta))(\theta(1-\lambda)+\lambda+\theta\Theta)} < 0$.

In words, the model predicts that areas with stronger anti-immigrant animus experienced a *smaller* change in overt attitudes following the referendum. If, as suggested by existing evidence, the share of leave votes is positively correlated with anti-immigrants animus, the implication is that the change in attitudes should be less pronounced in areas with a larger share of leave vote, *ceteris paribus*, in line with our Observation 1. 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 private 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. In Appendix B.6, we investigate the alternative idea that, rather than being driven by the desire to conform to social norms, 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 generate the opposite prediction to Result 3.

How does the magnitude of the effect identified in Result 3 change with λ, the role played by country-wide overt attitudes in determining the social norm? Since both γ_0 and γ_1 are decreasing

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 λ, the net effect on $\gamma_1 - \gamma_0$ is ambiguous. This implies that a higher λ may strengthen or weaken the result, depending on parameter values. The same holds for stronger conformity concerns (lower θ).

4.4 Implications for hate crime

In this section we address the implications of our previous analysis for hate crime. We think of hate crime as an extreme expression of overt, publicly expressed negative attitudes towards immigrants (or more generally individuals from a minority ethnic background).[17] The simplest way of capturing this is to assume that there is a threshold level of a, denoted as β, such that, whenever overt attitude exceeds this threshold – so that $a_i > \beta$ – this is classified as hate crime. Recall that, in our setup, individual utility is

$$u_i = -\theta \left(a_i - \alpha_i\right)^2 - (1-\theta)\left(a_i - \overline{a}^{n_d}\right)^2 \tag{14}$$

$$= 2a_i \left[\theta\alpha_i + (1-\theta)\,\overline{a}^{n_d}\right] - a_i^2 + K \tag{15}$$

where $K \equiv -\theta\alpha_i^2 - (1-\theta)\left(\overline{a}^{n_d}\right)^2$. This makes clear that, although we have expressed utility as a weighted average of two loss functions, we can equivalently think of it as the result of the trade-off between the individual return from selecting overt attitude a_i — given by $2a_i\left[\theta\alpha_i + (1-\theta)\,\overline{a}^{n_d}\right]$ — and the cost arising from the expected legal sanctions associated with a_i — given by a_i^2. This implies that, if the social norm is higher (i.e., less pro-immigrant) there are higher returns (either from social esteem or self-esteem) from adopting an overt behavior that is less favorable to immigrants.

A possible micro-foundation for the expected cost arising from legal sanctions is the following: all extreme behaviors, either extremely anti-minorities ($a_i > \beta$) or extremely pro-minorities/anti-majority ($a_i < -\beta$), are classified by the police force as crimes (with the first type being classified as hate crimes), and individuals are unsure about the exact value taken by β.[18]

[17] The notion that hate crimes reflect public behavior is backed by evidence. The breakdown of hate crimes in 2016/17, for instance, reports that 56% of these were public order offenses, 33% were violence against the person, 6% criminal damage or arson and 5% were classified as "other." Source: Home Office Hate Crime Report for England and Wales 2016/17.

[18] To fix ideas, suppose that the legal sanction for $a_i > \beta > 0$ is $b\,(a_i - \beta)$, for some $b > 0$, and that β is distributed uniformly on $[0, \overline{\beta}]$, where $\overline{\beta}$ is "large." Consider $a_i > 0$ (the case $a_i < 0$ is analogous), and let $b = 2\overline{\beta}$. Then, the expected cost associated with a_i is $2\int_0^{a_i}(a_i - \beta)\,d\beta = a_i^2$ for all $a_i \leq \overline{\beta}$.

4.4.1 Change in hate crime following the referendum

From the analysis above, we know that, in a district d, the referendum changes the distribution of overt attitudes as follows.

- Before the referendum: $a_i^d \sim N(\overline{a}_{before}^d, \theta^2\sigma)$ where $\overline{a}_{before}^d = (\theta + \gamma_0)\, P^d + (1 - \theta - \gamma_0)\overline{\mu}$.
- After the referendum: $a_i^d \sim N(\overline{a}_{after}^d, \theta^2\sigma)$ where $\overline{a}_{after}^d = (\theta + \gamma_1)\, P^d + (1 - \theta - \gamma_1)\mu$.

Denoting as $F(.)$ the cdf of the standard normal distribution, the difference in hate crime generated by the referendum is then equal to $F(\beta_{before}^d) - F(\beta_{after}^d)$, where $\beta_{before}^d \equiv (\beta - \overline{a}_{before}^d)/\theta\sqrt{\sigma}$ and $\beta_{after}^d \equiv (\beta - \overline{a}_{after}^d)/\theta\sqrt{\sigma}$. Differentiating with respect to P^d we obtain

$$\frac{1}{\theta\sqrt{\sigma}}\left[(\theta + \gamma_1)\,\varkappa + f(\beta_{before}^d)\,(\gamma_1 - \gamma_0)\right] \tag{16}$$

where $\varkappa \equiv f(\beta_{after}^d) - f(\beta_{before}^d)$ and, as already observed, $\gamma_1 - \gamma_0 < 0$. There are two countervailing effects at play. One is that, as highlighted in Result 3, the shift in overt attitudes induced by the referendum is smaller in districts with more pronounced anti-immigrant preferences. In (16), this is captured by $f(\beta_{before}^d)\,(\gamma_1 - \gamma_0) < 0$. The second effect is that, ceteris paribus, in areas with more pronounced anti-immigrant preferences, a marginal worsening of outward attitudes towards immigrants will bring a bigger mass of individuals over the hate crime threshold. That's because, in those areas, the distribution of preferences is shifted to the right compared to areas where preferences are more pro-immigrant (recall that we are adopting the convention that higher values of α correspond to a stronger dislike of immigrants). In the Appendix, we prove that if the within-district variance of individual preferences is large enough, so that overt attitudes are sufficiently spread out, the first effect outweighs the second, and hence the change in hate crime is decreasing in P^d, the amount of anti-immigrants sentiment in district d.

Proposition 3. (Change in hate crime at district level) *If σ, the variance of within-district preferences, is sufficiently large, the difference in hate crime before and after the referendum is decreasing in P^d.*

4.5 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 fueled 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 μ_{before} and μ_{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 average overt attitudes in the alternative setup can be easily computed using (7),

$$\overline{a}_{after} - \overline{a}_{before} = (\theta + \gamma_0)\,(\mu_{after} - \mu_{before}) + (1 - \theta - \gamma_0)(\overline{\mu}_{after} - \overline{\mu}_{before}). \tag{17}$$

Hence, a shift in preferences caused by the referendum is consistent with an increment in overt anti-immigrant attitudes 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 average overt attitudes in district d following a shift in preferences is equal to,*

$$\overline{a}^d_{after} - \overline{a}^d_{before} = (\theta + \gamma_0)\,(P^d_{after} - P^d_{before}) + (1 - \theta - \gamma_0)(\overline{\mu}_{after} - \overline{\mu}_{before}). \tag{18}$$

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 2, 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 shift in preferences (and thus behavior) in those areas that were already anti-immigrant to start with.[19] This is consistent with recent literature such as Adena, Enikolopov, Petrova, Santarosa and Zhuravskaya (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, Egorov, Enikolopov and Petrova (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 1 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 the evidence pointing to a spike in hate crimes in the immediate aftermath of the referendum results being announced. We find this argument suggestive that an information shock explanation of Section 4.3 is a more plausible explanation of the evidence than a preference shift.

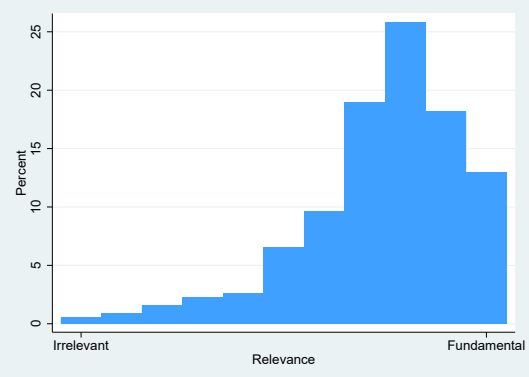
[19] Our measure of prejudice is the share of the leave vote in the referendum, namely an event that took place several years after 7/7. The underlying assumption is one of sufficient persistence anti-immigrant preferences across time. This is corroborated by existing studies such as Becker and Fetzer (2016) and Becker, Fetzer and Novy (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.

5 Survey evidence on model mechanisms and facts

The model relies on a number of assumptions and features. To provide further evidence, we carried out a survey analysis to check their plausibility. We designed a specific questionnaire to collect information about views on immigrants, overt attitudes and the Brexit referendum. We interviewed 1800 respondents during February and March, 2021. The survey was programmed on the Qualtrics web platform and was run on Prolific. To reinforce accuracy in responses, the questionnaire included attention check questions (i.e. questions with the answer included in its formulation). We report the questionnaire in Appendix C. We initially kept 1720 respondents who passed the attention check and fully completed the survey. We also removed 85 respondents who did not vote in the referendum. The final sample involves 1635 respondents. We targeted our respondents according to their vote in the referendum and their gender. In our sample, 52.1% people voted leave and 50% are female. Descriptive statistics are reported in Appendix C.1, The geographical distribution of respondents is displayed in Appendix C.2.

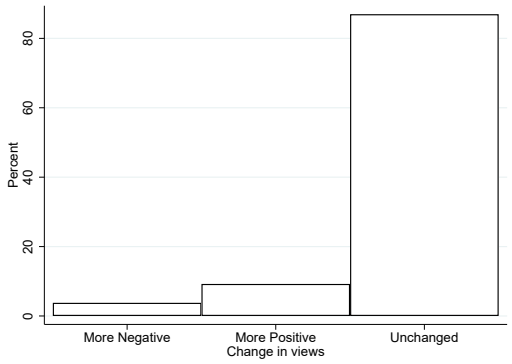
The survey responses support key elements of the model. First, our data confirm that concerns over immigration played a major role in the referendum result. We asked the respondents how important they thought concerns over immigration and erosion of British identity had been for the referendum outcome, on a scale ranging from 0 (irrelevant) to 10 (fundamental). Figure 4 shows how the distribution of responses leans towards "fundamental" with a mode of 8.

Figure 4: Role of immigration concerns in Brexit Vote



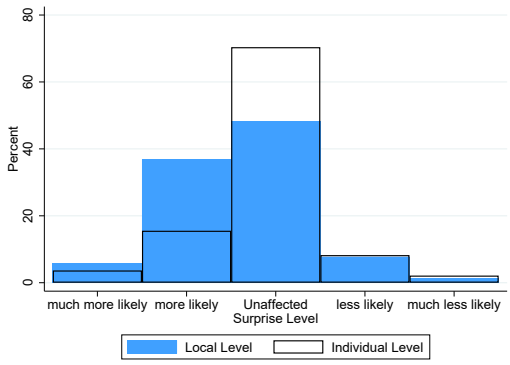
Importantly, respondents also responded that their views on immigrants were unaffected by the referendum. Figure 5 reports this finding, which supports the theoretical result according to which the change in attitudes triggered by the referendum was not driven by a shift in individual preferences.

Figure 5: Did the referendum change how people view immigrants?



Third, our evidence suggests that people became more comfortable in publicly expressing their views about immigrants after the referendum. We asked the respondents if the referendum had caused them to become more or less likely to express their views on immigrants. The variable $MoreVocal$ takes five values: much less likely (1), less likely (2), unaffected (3), more likely (4), and much more likely (5). As shown in Figure 6, around 20% of respondents became more at ease with vocalizing their views on immigrants after the referendum. This proportion increased to over 40% when respondents were asked whether they thought that, post referendum, people in their local area had become more vocal in expressing their views.

Figure 6: Expressing Views on Immigrants After Referendum



The survey also allows us to identify the characteristics of those who became more comfortable with expressing their views post-referendum. In our theory, from (6), the difference between an individual's overt attitudes and their personal preferences is given by,

$$a_i - \alpha_i = (1-\theta)[(\overline{a}^{n_d}) - \alpha_i]. \tag{19}$$

We are looking to identify those individuals for whom the post-referendum adjustment in overt attitudes resulted in a smaller discrepancy between attitudes and preferences. The first observation is that, following our theory (Result 2) the change in overt attitudes following the referendum was stronger in areas where people are, on average, more favorable to immigrants. In those areas, the surprise effect was more pronounced and thus there was more scope for a post-referendum behavioral realignment. Second, for this realignment to result in *less* discrepancy between own preferences and overt attitudes, we need to focus on individuals who *dislike* immigrants. This follows since, after the referendum, the norm became less immigrant-friendly. These predictions are confirmed by our data. We asked respondents to (i) state their views on immigrants on a scale from 0 (very negative) to 10 (very positive) ($MviewInd$), and (ii) to indicate their perception of the average views about immigrants in their local area ($MviewLocal$). This information allows us to estimate the after-referendum change in vocalizing immigration views ($MoreVocal$) as follows:

$$MoreVocal_i = \alpha_1 MViewLocal_i + \alpha_2 MViewInd_i + \alpha_3 MViewInd_i \times MViewLocal_i + \gamma \mathbf{X}_i + \epsilon_i, \tag{20}$$

where $MoreVocal_i$, $MviewLocal_i$ and $MviewInd_i$ are defined above and $\mathbf{X}_i$ is a vector of individual characteristics such as age, sex, level of education, income and employment status. The predictions of our model are that $\alpha_1 > 0$ and $\alpha_3 < 0$ – i.e., those who become more vocal post-referendum tend to live in areas where immigrants are well-liked and tend to personally dislike immigrants.

Table 11 displays equation (20). As expected, the coefficient associated with local support for immigrants is positive, while the interaction between local support and own views on immigrants is negative. The third column shows that $\alpha_1 > 0$ and $\alpha_3 < 0$ are robust to the inclusion of individual characteristics.

Table 11: Determinants of changes in vocalizing views on immigrants

	(1)	(2)	(3)
MViewInd	-0.063***	-0.017	-0.010
	(-7.68)	(-1.19)	(-0.67)
MViewLocal	0.019*	0.102***	0.101***
	(2.26)	(4.69)	(4.67)
MViewLocal × MViewInd		-0.012***	-0.012***
		(-3.98)	(-4.04)
Age			0.005***
			(4.20)
Sex			0.036
			(1.11)
Education			-0.012
			(-0.87)
Income			0.012
			(0.76)
Employment			-0.0159
			(-1.24)
N	1635	1635	1635

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Finally, to investigate the surprise effect generated by the referendum, we asked respondents

to indicate the level of surprise they experienced (from 0 to 10) with respect to the referendum outcome at UK level as well as at local level. In Table 12, we report mean surprise levels in both dimensions. Clearly, respondents were surprised by both results. However, surprise at the outcome of the referndum at UK level was stronger, which is consistent with individuals being better informed about underlying preferences in their local area compared to their knowledge at the country level.

Table 12: Surprise with referendum results at UK and Local level

	Surprise with UK result	Surprise with Local result	Difference
Mean	6.227	4.67	1.56***
			t=16.8

As our theoretical mechanism relies on the surprise at the country-level support for Brexit being stronger in more pro-remain areas, we estimate the following relationship:

$$\textit{Referendum Surprise UK}_i = \alpha_1 \textit{Remain Share Local}_j + \gamma \mathbf{X}_i + \epsilon_i, \tag{21}$$

where *Referendum Surprise UK*$_i$ reflects the level of surprise of respondent i and *Remain Share Local*$_j$ is the share of the Remain vote in area j. Table 13 reports the results. As predicted by our model, the level of surprise increases with the share of Remain vote and this is robust to the inclusion of individual characteristics. This is also consistent with our Observation 3 based on the British Election Study survey.

Table 13: Surprise and Brexit vote

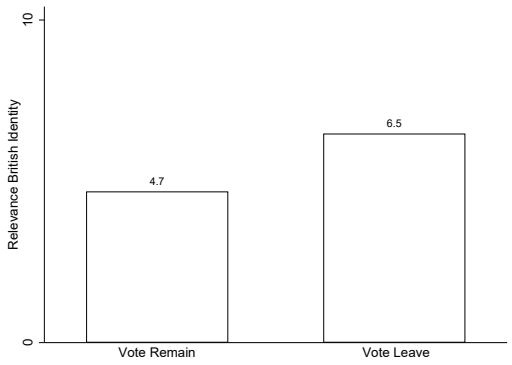
	(1) VoteSurpriseUK	(2) VoteSurpriseUK
Remain Share Local	0.0229***	0.0208***
	(-4.01)	(-3.59)
Age		0.00980
		(1.35)
Sex		0.0348
		(0.19)
Education		0.145
		(1.80)
Income		0.146
		(1.61)
Employment		-0.0841
		(-1.13)
N	828	828

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Finally, a key element of the model is that individuals wish to conform, at least to some extent, to average overt attitudes at country level. To tackle this question, we asked participates to indicate how important (on a scale from 0 to 10) is their country in defining their identity. The mean value reported by respondents is 6, which corresponds to $\lambda = 0.6$ in our model. As a sanity check, we report that the strength of British identity among our subjects was stronger among leave voters, as one would expect. This is shown in Figure 7.

Figure 7: Importance of country in defining own identity



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

6.1 Paramaterization: linking the model with our observables

Agents in the model differ in their preference parameter α. Given preferences, beliefs and geographical location they determine overt attitude a. In the data, the distributions of both preferences and beliefs are unobservable. Hence, in this exercise the preference parameter α will be treated as a latent variable and will be inferred from the level of hate crime and voting behavior. The way in which overt attitudes translate into hate crime is as described in section 4.4. Since we do not model

voting behavior explicitly in the model (as it is not needed), here we simply assume a reduced form linear voting rule. A forecaster observing the same information as the agent's in the model predicts the probability of remain winning the referendum and we target this probability to betting odds at the time of the referendum. Finally, the model features two periods. We consider the financial year prior to the Brexit referendum and the subsequent year and index the periods by $t \in \{0, 1\}$.

Hate Crime. To link the model to the hate crime data discussed earlier we define the measure of the prevalence of hate crime as the number of hate crimes in a year divided by the foreign born population. In a loose sense, one can interpret this as the approximate probability an immigrant is a victim of a hate crime in a given year. Recall that an overt xenophobic attitude becomes a hate crime if attitude a exceeds a threshold β. If one assumes the *opportunity* to commit said crime is proportional to the share of migrants in one's local area then the model counterpart to this is as described in section 4.4. Hence, the level of hate crime per immigrant in district d is the survival function of a standard normal distribution.

$$h_t^d = 1 - F(\beta_t^d) \quad \text{where,} \quad \beta_t^d = \frac{\beta - \overline{a}_t^d}{\sigma\theta}$$

Voting behavior. We take the baseline model, described in section 4.2, to the data. The theory does not take a precise stance on how individuals vote, since the precise voting rule is irrelevant – all that matters is that people are aware of how preferences affect voting decisions and can therefore infer aggregate preferences from voting shares. Here, however, we need to model voting explicitly. For concreteness, we assume that an individual's voting decision is a linear function of primitive preferences α_i. The probability of voting remain for individual i is thus given by

$$\text{Prob(vote remain)} = b_0 + b_1\alpha_i + \epsilon_i^v \quad \text{where, } \epsilon_i^v \sim N(0, \sigma_v^2).$$

Under the assumptions that α_i and ϵ_i^v are independent and identically (normal) distributed and that an agent is atomless in a given district, this voting rule can be aggregated so that the share of remain votes in district d is given by

$$v^d = b_0 + b_1 P^d. \tag{22}$$

We put no restrictions on the additional parameters b_0 and b_1. The size b_1 determines the intrinsic importance of aggregate preferences towards immigrants in a district for the 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 personal view on immigrants (see Appendix D.1), this assumption is clearly simplistic and is made for tractability. In the model, districts very only with respect to aggregate preferences towards immigrants (which can result from a number of economic and social factors), while in the data districts clearly vary along a multitude of dimensions. The linear voting rule is selected primarily for its simplicity. However, it does provide additional advantages. Firstly, it is used frequently in statistical models of voting, see Auld and Linton (2019) for the case of the referendum. Secondly, as has been shown, it allows for simple aggregation. Finally, in the model an individual's overt attitude (a_i) is a linear function of their preferences (α_i). The linear voting rule is therefore also consistent with voting decisions being governed by overt attitudes rather than being a direct expression of individual preferences as in sincere voting. All results of the two exercises will be identical with the parameters b_0 and b_1 adjusting accordingly.

The forecaster. The final addition to the model is the introduction of a forecaster. Since the model's mechanism is through agents' beliefs, we include in the estimation a moment related to expectations. In period 0 the forecaster 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 preferences regarding migrants, μ. Assuming a constant population in each district, the forecaster's prior distribution of remain votes is given by equation (23).

$$\frac{1}{n}\sum_d v^d \sim N\left(b_0 + b_1\overline{\mu}, \Theta + \frac{b_1^2 \text{Var}_d(P^d)}{n}\right). \tag{23}$$

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

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

$\Phi_v(\cdot)$ is the associated cumulative distribution function defined in equation (23). 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 D.2 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.

Calibrated parameters. In a pre-estimation step we calibrate a number of parameters. Since an agent's innate preferences are ordinal in nature, we normalize the mean in society μ to be zero. Thus for $\alpha > 0$ an individual is more xenophobic than the average in society. Similarly, for $\overline{\mu} > 0$ agents believe that society is more xenophobic than it actually is. In the theory section we normalized the variation in P^d across district to unity. Since our estimation will rely on data dissagregated to the local area, here we instead define $\text{Var}_d(P^d) \equiv \sigma_d^2$ and specify the dispersion of within location attitudes (σ) to be unity. Hence, all dispersion parameters can be thought of as relative to the dispersion within a local CSP. Finally, without more dissagregated data it is difficult to get a handle on the importance of conforming to the national relative to local norm, governed by the parameter λ. Here we rely on our survey and set $\lambda = 0.6$, see section 5. This leaves a vector of structural parameters $\vec{\theta}$ to be estimated, where

$$\vec{\theta} := (\overline{\mu}, \beta, \Theta, b_0, b_1, \sigma_d, \theta) .$$

6.2 Estimation

The vector $\vec{\theta}$ is estimated by simulating the model M times and computing a vector of moments $m(\vec{\theta})$ each iteration. The simulated data is made to look as similar to the empirical data described

before, for two periods with 314 districts.[20] 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 14. Parameters are estimated minimizing the function below.

$$\hat{\vec{\theta}} = \arg\min_{\vec{\theta}} \left(\frac{1}{M} \sum_{s} m_s(\vec{\theta}) - \hat{m} \right) \Omega \left(\frac{1}{M} \sum_{s} m_s(\vec{\theta}) - \hat{m} \right)' \tag{24}$$

The moments used in estimation are listed in Table 14 along with the fit of the model. The model is over-identified in the sense that the number of moment conditions (nine) exceed the number of parameters to be estimated (seven). The weighting matrix Ω is set as $\text{diag}(m^{-2})$ to put equal weight to each moment, irrespective of a moment's magnitude.[21] Overall, the simulated model fits the data well with the exception of the variance of the referendum result across local areas. Since there are clearly a multitude of other sources that determine regional voting patterns, this is unsurprising.

The estimated parameters are presented in Table 15. The desire to conform $1 - \theta$ is arguably small. Personal preferences determine approximately 78% of individual behavior, while the desire to conform to average attitude in society determine the remaining 22%. We are not aware of any 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, Malmendier and Weber (2012) and and List (2007), Krupka and Weber (2013) estimate that social appropriateness determines 57% to 74% of individual behavior. Our smaller figure of 22% 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

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

[21] 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, Monfort and Renault (1993). Further M is set to one thousand to ensure stability of the simulated moments.

Table 14: Targeted Moments in Estimation

	Simulated Moment	Empirical Moment
Distribution of hate crime		
Mean of the log of hate crime per immigrant (pre-referendum)	-5.288	-5.287
Variance of the log of hate crime per immigrant (pre-referendum)	0.448	0.452
Mean of the log of hate crime per immigrant (post-referendum)	-5.069	-5.068
Variance of the log of hate crime per immigrant (post-referendum)	0.425	0.422
Referendum result		
Mean remain share	0.457	0.457
Variance of remain share	0.001	0.010
Covariance of referendum result and:		
The log of hate crime per immigrant (pre-referendum)	-0.019	-0.021
The log of hate crime per immigrant (post-referendum)	-0.019	-0.017
Betting markets		
Implied probability of a remain victory	0.860	0.860

Note: All means and variances are across community safety partnerships. Data come from the ONS and is as described in section 3.1 with the exception of the implied probability of a remain victory which comes from bookmakers odds shortly before the referendum and is reported in Appendix D.2.

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 individuals. 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 15: Parameter Estimates

$\overline{\mu}$	β	Θ	b_0	b_1	σ_d	θ
−0.428	1.946	0.010	0.457	−0.139	0.211	0.783
(0.048)	(0.040)	(0.002)	(0.005)	(0.024)	(0.008)	(0.016)

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 estimated parameters suggest that the shift in beliefs from prior to posterior was quite

large, as measured by the relative size of $\overline{\mu}$ to the variance of the prior Θ.[22] 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 was something of a "*bolt from the blue*" which changed people's perspectives about British views on immigrants considerably.

The between district standard deviation σ^d is one fifth of the dispersion within district. This coupled with the large value of β making hate crime a tail event puts us in a world outlined in Proposition 3 whereby hate crime increases most in areas voting remain. The parameters b_0 and b_1 relate to the exogenous voting rule (equation (22)). A positive estimate of b_0 and a negative b_1 imply that the higher the propensity to hate crime of an area the less likely the area is to vote remain. 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

$$a_i - a_j = 0.78(\alpha_i - \alpha_j),$$

a little more than 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 $\Delta E(\overline{a}^{n_d})$, this will induce a behavioral shift corresponding to approximately $\frac{1}{4}\Delta E(\overline{a}^{n_d})$.

The role of stereotypes and shared narratives. Consider now two different societies that

[22] Recall that society's mean aggregate true preference μ is normalized to zero.

are characterized by different priors $\overline{\mu}_A$ and $\overline{\mu}_B$. Even if the (unobserved) realized mean preferences μ are *identical*, overt attitudes in these two societies may differ in a non-negligible way. From our previous analysis, average attitude is equal to

$$\overline{a} = (\theta + \gamma_0)\mu + (1 - \theta - \gamma_0)\overline{\mu}. \tag{25}$$

After substituting for θ and γ_0 using the parameters in Table 15, the difference in average attitudes between the two societies is equal to

$$\overline{a}^B - \overline{a}^A = 0.14(\overline{\mu}_B - \overline{\mu}_A).$$

This observation underscores the role 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 (unobserved) realized mean preferences μ, but different values of Θ, Θ_A and $\Theta_B > \Theta_A$. Recall that Θ 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 (25), the difference in average attitudes in the two societies is

$$\overline{a}^B - \overline{a}^A = (\gamma_0^B - \gamma_0^A)(\mu - \overline{\mu}) \tag{26}$$

where,

$$\gamma_0^B - \gamma_0^A = \frac{\lambda(1-\theta)\theta(\Theta_B - \Theta_A)}{[\theta(\Theta_B + 1) + \lambda(1-\theta)][\theta(\Theta_A + 1) + \lambda(1-\theta)]} > 0.$$

A number of remarks are in order. First, note that, from (26), if $\overline{\mu} = \mu$, then $\overline{a}^B = \overline{a}^A$. 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 $\overline{a}^B > \overline{a}^A$, and vice-versa, if $\mu < \overline{\mu}$ (true preferences are *less* anti-immigrant than the stereotype), then $\overline{a}^B < \overline{a}^A$, with

$$\overline{a}^B - \overline{a}^A = \lambda(1-\theta)(\mu - \overline{\mu}) = 0.13(\mu - \overline{\mu}).$$

Note that, in the society characterized by a stronger stereotype (society A), mean behavior is always further away from true preferences than in the society with a weaker stereotype (society B).[23] 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. As we have seen, the change in aggregate behavior triggered by the information shock is

$$\overline{a}_{after} - \overline{a}_{before} = (1 - \theta - \gamma_0)\,(\mu - \overline{\mu})$$

where $\gamma_0 \equiv \frac{\theta(1-\theta)(\Theta+1-\lambda)}{\theta+\lambda(1-\theta)+\theta\Theta}$, increasing in Θ. When $\Theta \to \infty$, i.e. the ex-ante stereotype is extremely weak, $\gamma_0 \to 1-\theta$, which implies that the information shock generates no change in aggregate behavior. That's 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

[23] Recall that, from (25), if $\mu > \overline{\mu}$ then both $\overline{a}^A$ and $\overline{a}^B$ are $< \mu$ (while the opposite holds if $\mu < \overline{\mu}$). When $\mu > \overline{\mu}$, $\overline{a}^B > \overline{a}^A$ implies $\mu > \overline{a}^B > \overline{a}^A$ (and, similarly, when $\mu < \overline{\mu}$, $\overline{a}^B < \overline{a}^A$ implies $\mu < \overline{a}^b < \overline{a}^A$). In both cases, $|\overline{a}^B - \mu| < |\overline{a}^A - \mu|$.

of the information shock on aggregate behavior is inversely related to Θ, and is thus increasing in the strength of the ex-ante stereotype.

7 Concluding remarks

A referendum is a universal vote and its result publicly reveals new information about the population's underlying preferences. If prevalent private views within society were 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 shift in overt attitudes towards immigrants and minorities that followed the 2016 Brexit referendum, exemplified by a sharp increase in hate crime episodes as well as 'softer' expressions of anti-immigrant positions.

We show that the hike in hate crime was more pronounced in areas with a higher share of remain vote. As the leave vote is associated with concerns with immigration, we interpret a higher share of remain vote as an indication of being friendlier towards immigrants. This is confirmed by our survey data: the referendum outcome caused those who dislike immigrants, but live in areas where immigrants are well-liked, to become more vocal about their views. What happened in those areas? Did the British public living in relatively open areas suddenly become more hostile overnight towards immigrants and ethnic minorities? Why not before the referendum? 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 these post-referendum changes in overt attitudes 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. Our theory argues that, since agents' beliefs about prevalent views in society are guided by their local area, this surprise effect was particularly pronounced in pro-immigrant areas. Our survey data confirms this hypothesis, showing that the level of surprise felt by people upon learning the outcome of the referendum country-wide was more pronounced in areas with a higher share of remain votes.

In those areas, the perception of an immigrant-friendly social norm made those opposed to migrants conceal or repress their private views. After updating their perception about prevalent views in UK society, they 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 in 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 more pro-remain areas, it follows that these areas exhibited a greater a change in the hate crime.

While our analysis studies and dissects the post-referendum spike in hate crime, we cannot make claims about the duration of the spike or its long-term geographic trends. Our theory shows that overt attitudes respond to both shocks in preferences and changes in the perceived social norms. The evolution of hate crime reflects the evolution of these forces over time. Clearly enough, the longer the period of time elapsed since the referendum, the more likely it is that other changes will have occurred which may affect attitudes towards immigrants. A case in point is the so-called "Summer of Terror" of 2017 which saw two major terrorist attacks take place in rapid succession: the Manchester Arena bombing on the 22nd of May and the London Bridge attack on the 3rd of June. These attacks affected hate crime, generating spikes in their own right. As shown in our analysis, terrorist attacks increase hate crime more in relatively anti-immigration areas (e.g. pro-leave areas), which is the opposite of what we find following the Brexit referendum. This implies that the 2017 attacks might have introduced forces that countervail the "*surprise-driven*" mechanisms we describe. Another important countervailing factor was triggered by the reaction to the spike in hate crime and xenophobia. The international community as well as various media outlets pressured the UK government to take decisive action to combat the change in hate crime trend. For example, in 2017, the Crown Prosecution Service published a "Hate Crime Strategy 2017-2020" document, which outlined a series of measures aimed at reducing hate crime, such as increasing conviction rates and making it easier to prosecute hate crime perpetrators.[24]. A tougher stance by the government is likely to affect hate crime. Furthermore, many people felt the need to

[24] Available at https://www.cps.gov.uk/sites/default/files/documents/publications/CPS-Hate-Crime-Strategy-2020-Feb-2018.pdf

dissociate themselves from "*racist Britain*" for fear of being accused of xenophobia. This backlash is likely to have caused a further readjustment in what people perceive as social acceptable, as argued for instance by Schwartz et al. (2020).

In sum, the evolution of overt attitudes and hate crime is affected by discrete political and social events. Highlighting the importance of updates in people's perceptions of the country-level social norm is our main contribution. Providing a full account of the evolution of hate crime over time which identifies each of the different forces governing its dynamics would be a potentially useful exercise, but it is beyond the scope of this paper.

An important contribution of our work is to emphasize how superficially similar shifts in overt attitudes – exemplified for instance by hate crime spikes triggered by terrorist events or by public revelation of new information about privately held views (as in the case of a referendum) – may in fact be driven by very different mechanisms, preference-shocks driven against information-shocks driven. In turn, this generates very different geographical patterns of behavioral responses, which allows to make predictions on the likely effects of various types of shocks. For instance, the behavioral effect of a public revelation that indicates that immigrants are *more liked* across the country than what was previously thought should be stronger in anti-immigrant areas.

Though the Brexit context is unique, there are many other examples of situations where our mechanism may apply. We can test the robustness of our approach in other contexts. For example, there are concerns about how the recent sexist and homophobic rhetoric of President Bolsonaro in Brazil might have stimulated aggressive behavior against sexual minorities. Clearly, investigating the multi-faceted relationship between preferences, norms and behavior is crucial for our understanding of how societies work and evolve, and continues to be an urgent topic of research. This 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.

.

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.

———, "A theory of esteem based peer pressure," *Games and Economic Behavior* 115 (2019), 314–335.

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

Akerlof, G. A. and R. E. Kranton, "Economics and identity," *The quarterly journal of economics* 115 (2000), 715–753.

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).

Bénabou, R. and J. Tirole, "Identity, morals, and taboos: Beliefs as assets," *The Quarterly Journal of Economics* 126 (2011a), 805–855.

———, "Laws and norms," (2011b).

Bernheim, B. D., "A Theory of Conformity," *Journal of Political Economy* 102 (1994), 841–877.

Bicchieri, C., E. Dimant and S. Sonderegger, "It's Not A Lie if You Believe the Norm Does Not Apply: Conditional Norm-Following with Strategic Beliefs," *Available at SSRN 3326146* (2020).

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," *American Economic Review* 110 (2020), 3522–3548.

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

Cavalli, N., "Did hate crime double after Brexit?," *Centre for Social Investigation, CSI* 34 (2019).

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., "Discrete Events and Hate Crimes: The Causal Role of the Brexit Referendum," *Social Science Quarterly* 102 (2021), 374–386.

Dipoppa, G., G. Grossman and S. Zonszein, "Locked Down, Lashing Out: Situational Triggers and Hateful Behavior Towards Minority Ethnic Immigrants," (2020).

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.

Ferrin, M., M. Mancosu and T. M. Capiali, "Terrorist attacks and Europeans' attitudes towards immigrants: An experimental approach," *European Journal of Political Research* 59 (2020), 491–516.

Fetzer, T., "Did austerity cause Brexit?," *American Economic Review* 109 (2019), 3849–86.

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

Frey, A., "'Cologne Changed Everything'—The Effect of Threatening Events on the Frequency and Distribution of Intergroup Conflict in Germany," *European Sociological Review* 36 (2020), 684–699.

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.

Hanes, E. and S. Machin, "Hate Crime in the wake of terror attacks: evidence from 7/7 and 9/11," *Journal of Contemporary Criminal Justice* 30 (2014), 247–267.

Human Rights, U., "End of Mission Statement of the Special Rapporteur on Contemporary Forms of Racism, Racial Discrimination, Xenophobia and Related Intolerance at the Conclusion of Her Mission to the United Kingdom of Great Britain and Northern Ireland," (2017).

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.

Kuran, T., "The East European revolution of 1989: is it surprising that we were surprised?," *The American Economic Review* 81 (1991), 121–125.

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.

Lohmann, S., "The dynamics of informational cascades: The Monday demonstrations in Leipzig, East Germany, 1989-91," *World Pol.* 47 (1994), 42.

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.

Michaeli, M. and D. Spiro, "From peer pressure to biased norms," *American Economic Journal: Microeconomics* 9 (2017), 152–216.

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).

Schwartz, C., M. Simon, D. Hudson and J. van Heerde-Hudson, "A populist paradox? How Brexit softened anti-immigrant attitudes," *British Journal of Political Science* (2020), 1–21.

Soares, R., "Development, crime and punishment: accounting for the international differences in crime rates," *Journal of Development Economics* 73 (2004), 155–184.

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

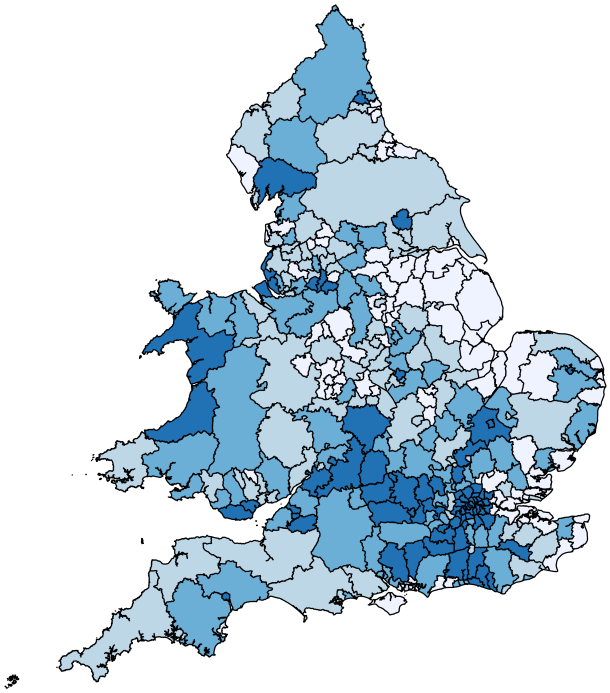
Voigtländer, N. and H.-J. Voth, "Nazi indoctrination and anti-Semitic beliefs in Germany," *Proceedings of the National Academy of Sciences* 112 (2015), 7931–7936.

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

A Empirical Appendix

A.1 England and Wales by Vote Share

Figure A.1: England and Wales by Vote Share



Note: CSPs are split into quartiles based on the proportion of remain votes in the referendum. The darker the shade the higher the remain share.

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 Attitude rather than votes

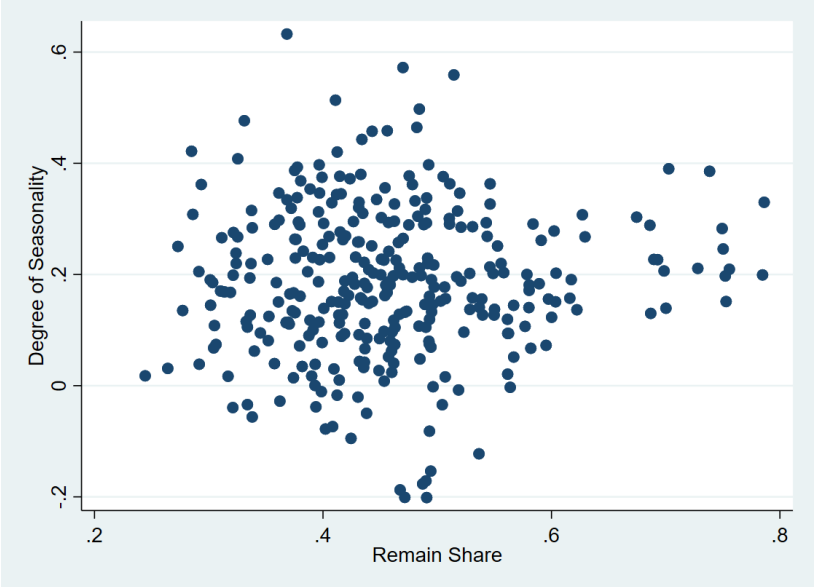
Table A.3: Dependent Variable: Log Hate Crime

	(1)	(2)	(3)	(4)	(5)	(6)
Lag Dep. Variable						0.135***
						(0.007)
Post Brexit ×	**-0.499*** **	**-0.407*** **	**-0.333*** **	**-0.501*** **	**-0.376*** **	**-0.324*** **
Anti-Immigration	**(0.117)**	**(0.118)**	**(0.121)**	**(0.119)**	**(0.133)**	**(0.131)**
Log NiNo EU		0.017**	0.018**	0.0185**	-0.007	-0.007
		(0.008)	(0.008)	(0.007)	(0.007)	(0.007)
Log NiNo RoW		-0.004	-0.005	-0.008	0.005	-0.002
		(0.009)	(0.009)	(0.009)	(0.008)	(0.008)
Log Population		1.151***	0.822***	0.305	-0.188	-0.187
		(0.146)	(0.205)	(0.203)	(0.229)	(0.230)
Log GDI			0.398***	0.394***	0.652***	0.626***
			(0.121)	(0.119)	(0.154)	(0.156)
Log Social Benefits			-0.138	-0.139	-0.449	-0.414***
			(0.107)	(0.105)	(0.152)	(0.151)
Log Other Crime				-0.716***	0.679***	0.603***
				(0.029)	(0.035)	(0.035)
Observations	18,840	18,840	18,840	18,840	18,840	18,526
R-squared	0.847	0.848	0.848	0.852	0.883	0.887
Number of CSPs	314	314	314	314	314	314
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.4 Seasonality in the data

The purpose of this subsection is to argue that there are little systematic differences in the seasonality of hate crime across place. Further, any differences in the degree of seasonality are unrelated to the share of voters voting remain. To do this, we first define our measure of seasonality. Aggregate statistics suggest that hate crime is largest in the second financial quarter of a year, July through September, and lowest in the fourth financial quarter, January through March. Our measure of seasonality is by CSP, we compute the difference in the log of the mean number of hate crimes in quarter two minus the same for quarter four. Figure A.4 is a scatter plot of our measure of seasonality against voting patterns.

Figure A.4: Seasonality by Remain Share



Seasonality is defined as the log difference in hate crime in the summer months relative to the winter months.

Inspection of the figure shows no clear systematic pattern in seasonality. Fitting an affine function to the data by ordinary least squares yields a slope coefficient of 0.06 with an associated standard error of 0.08.

A.5 Other Crime Placebo

Table A.5: Dependent Variable: Log All Other Crime

	(1)	(2)	(3)	(4)	(5)
Lag Dep. Variable					0.430***
					(0.007)
Post Brexit ×	**-0.172*** **	**-0.225*** **	**-0.242*** **	**-0.003**	**0.006**
Remain Share	**(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.6 Hate Crime Excluding London

Table A.6: 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

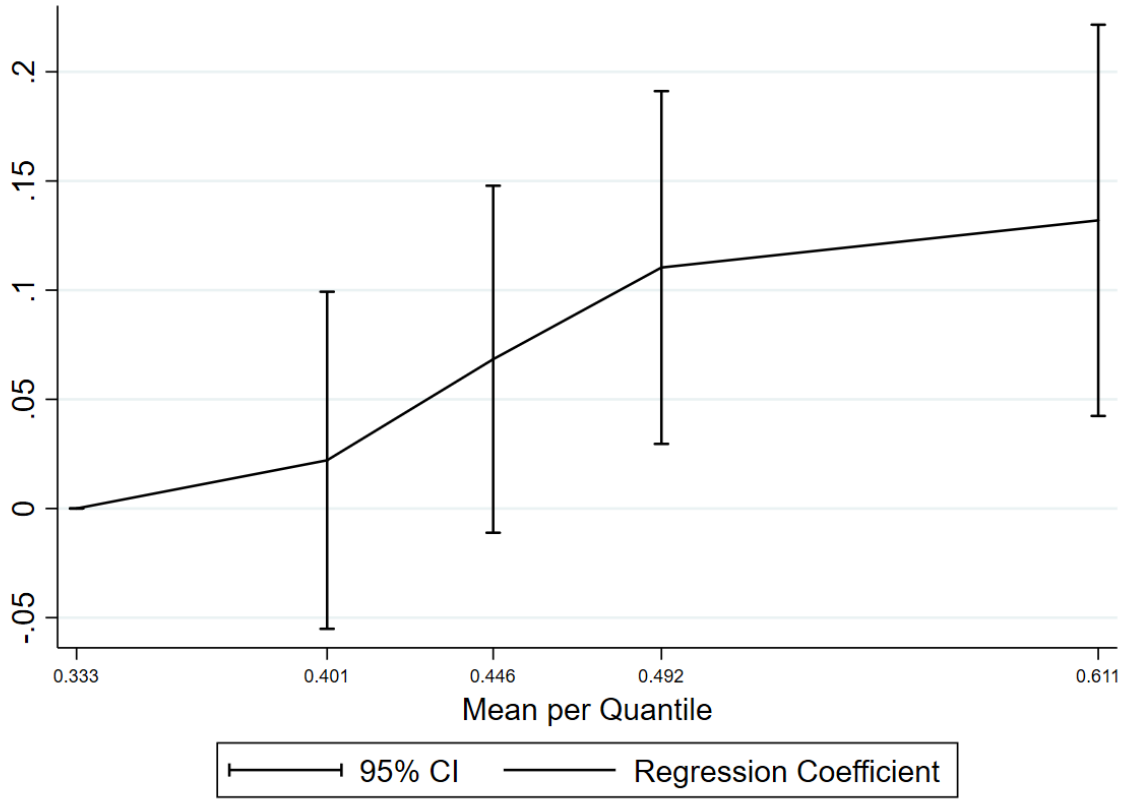
A.7 Linearity

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^{th} quintile and the value zeros otherwise.

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

Figure A.7: Quintile Dummies β_j



Note: The figure plots the coefficients β^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.8 Estimation of Reporting Rates

The crime specific component, ω_{ct}^1, 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*". [25] 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.8: 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

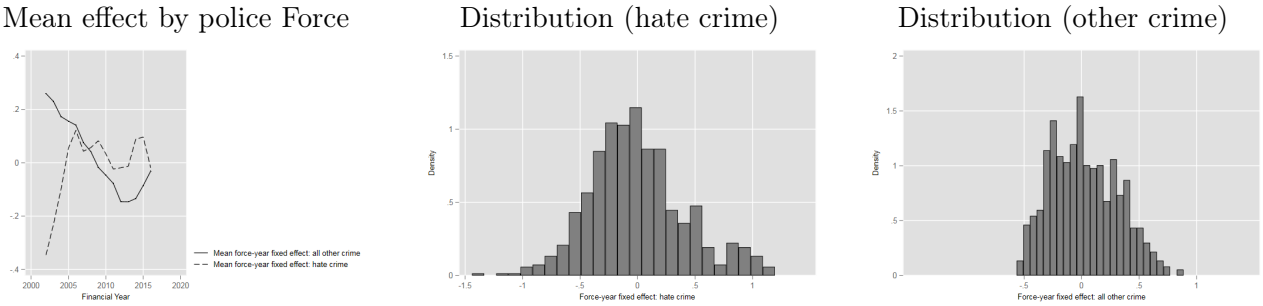
The hate crime reporting component ω_t^0 is identified through the reporting rates of hate crime according to the CSEW, reported in Table 3. Due to their being relatively few instances of victims reporting hate crimes, we allow this parameter to take two values, one pre-referendum ω_0^0, and one post-referendum ω_1^0. 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 $m_0 := \frac{\sum_{t\in\tau_0}\sum_c h_{ct}}{\sum_{t\in\tau_0}\sum_c h_{ct}^\star}$, directly from Table 3. Rearranging, and using our baseline data gives an estimate for ω_0^0 as

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

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

A.9 Force-Year Fixed Effects

Figure A.9: Force-Year Fixed Effects



Note: These plot the distribution of force-year fixed effects from the final column of Tables 2 and A.5, respectively. The first panel shows the mean (by police force) over time and the second two show the unweighted distribution.

A.10 Omitted Covariates from Table 10

Table A.10: 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 can perfectly observe average preferences in their own district. 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 ϱ_i is drawn as $N(0, x)$. The case analyzed in the main body above is thus a special case of this more general model in which $x = 0$. The individual best-reply action is now given by

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

In this setup, the Brexit vote plays the dual role of revealing *both* P^d and μ. 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 α_i – the latter is relevant because it is informative about P^d (and thus, indirectly, about μ as well).

Proposition 4. *When μ and P^d are imperfectly observed, the unique linear equilibrium of the game is given by,*

$$a_i = k_0\alpha_i + \gamma_0\widetilde{P}_i^d + (1-k_0-\gamma_0)\overline{\mu} \tag{27}$$

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

$$a_i = k_1\alpha_i + \gamma_1 P^d + (1-k_1-\gamma_1)\mu \tag{28}$$

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,

$$\Delta\overline{a} = \overline{a}_{after} - \overline{a}_{before} = (1-k_0-\gamma_0)(\mu-\overline{\mu}) \tag{29}$$

which replicates Result 1. Result 2 is similarly replicated by noting that mean pre-referendum beliefs about μ 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,

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

where it can be easily shown that,

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

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

Proof: *provided in Appendix B.4.*

Noteworthy, as shown in the proof, in order for $k_1 + \gamma_1 - k_0 + \gamma_0 < 0$ (implying greater surprise effect and greater behavioral adjustment in more pro-immigrant areas) to apply it is not necessary that $\lambda > 1/2$. In fact, so long as $x < \Theta$ (i.e., uncertainty about mean preferences at local level is smaller than at country level) then the result is consistent with $\lambda < 1/2$, i.e. country-wide behavior having a minoritarian role in determining local social norms. In general, since $\widehat{\lambda}$ is decreasing in x, for any $\lambda > 0$ we can identify a value $\widehat{x} > 0$ such that $k_1 + \gamma_1 - k_0 + \gamma_0 < 0 \Longleftrightarrow x <\widehat{x}$.

Result 5. *Suppose that $x > 0$. There exist $\widehat{\lambda} \in (0,1)$ such that the change in average overt attitudes caused by the referendum is decreasing in P^d if $\lambda > \widehat{\lambda}$ and is increasing in P^d if $\lambda < \widehat{\lambda}$. The sufficient condition for $\widehat{\lambda} < 1/2$ is that $x < \Theta$.*

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 μ – 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 μ and P^d and pre-referendum beliefs about them. These are,

$$\mu - E_d[E_i(\mu \mid \widetilde{P}^d_i, \alpha_i)] = \frac{\Theta\,(\sigma + x)\,(\mu - P^d) + (x + \sigma + x\sigma)\,(\mu - \overline{\mu})}{\Lambda} \text{ decreasing in } P^d \tag{30}$$

and

$$P^d - E_d[E_i(P^d \mid \widetilde{P}^d_i, \alpha_i)] = \frac{\sigma x (P^d - \overline{\mu})}{\Lambda} \text{ increasing in } P^d \tag{31}$$

where $\Lambda \equiv \sigma\,(1 + x + \Theta) + x\,(1 + \Theta)$. There are thus two opposing forces at play. On the one hand, (30) 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 overt attitudes more. On the other hand, (31) 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 which effect dominates depends on the value taken by λ.

Result 5 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 Observation 1, namely that areas with a higher share of remain votes experienced a more pronounced increment in hate crime (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 λ^{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 (if any), 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 1 and Result 3, through the following observation,

Result 6. *Suppose that $\lambda^{Scot} < \lambda$ (where λ 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 Proofs of Propositions 1 and 2

The setup we consider in the main body is a special case of more general setup discussed in Appendix B.1 when $x = 0$. Propositions 1 and 2 are therefore also special cases of Proposition 4. Please refer to the proof of the Proposition 4 for the general proof.

B.3 Proof of Proposition 3

Proof of proposition 3: If we perform a second order Taylor approximation, we can write

$$\begin{aligned}\varkappa &\cong f\left(\beta^d_{before}\right)\left(\beta^d_{after}-\beta^d_{before}\right)\left[\beta^d_{before}-\frac{1}{2}\left(\left(\beta^d_{before}\right)^2-1\right)\left(\beta^d_{after}-\beta^d_{before}\right)\right]\\ &= f\left(\beta^d_{before}\right)\frac{\overline{a}^d_{before}-\overline{a}^d_{after}}{\theta^2\sigma}\left[\beta-\overline{a}^d_{before}-\frac{1}{2}\left(\frac{\left(\beta-\overline{a}^d_{after}\right)^2}{\theta^2\sigma}-1\right)\left(\overline{a}^d_{before}-\overline{a}^d_{after}\right)\right]\end{aligned}$$

Substituting in (16) we get

$$\frac{f\left(\beta^d_{before}\right)}{\theta\sqrt{\sigma}}\left[\widehat{\varkappa}+(\gamma_1-\gamma_0)\right] \tag{32}$$

where $\widehat{\varkappa}\equiv\frac{\overline{a}^d_{before}-\overline{a}^d_{after}}{\theta\sigma}\left[\beta-\overline{a}^d_{before}-\frac{1}{2}\left(\frac{\left(\beta-\overline{a}^d_{after}\right)^2}{\theta^2\sigma}-1\right)\left(\overline{a}^d_{before}-\overline{a}^d_{after}\right)\right]$. Recall that $\overline{a}^d$, γ_1 and γ_0 are independent of σ. If $\sigma\to\infty$, $\widehat{\varkappa}\to 0$ and hence (32) is unambiguously negative (recall that $\gamma_1-\gamma_0<0$). By continuity, this implies that there exists a value $\widetilde{\sigma}$ such that expression (32) is unambiguously negative whenever $\sigma>\widetilde{\sigma}$.

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

Proof of Proposition 4 We first solve for the pre-referendum equilibrium. Note that

$$\begin{aligned}E(\mu \mid \widetilde{P}_i^{\,d}) &= \frac{\Theta\sigma}{\sigma(1+x+\Theta)+x(1+\Theta)}\widetilde{P}_i^{\,d}\\ &+\frac{\Theta x}{\sigma(1+x+\Theta)+x(1+\Theta)}\alpha_i+\frac{x+\sigma+x\sigma}{\sigma(1+x+\Theta)+x(1+\Theta)}\overline{\mu}.\end{aligned} \tag{33}$$

and

$$\begin{aligned}E(P^d \mid \widetilde{P}_i^{\,d}) &= \frac{(1+\Theta)\sigma}{\sigma(1+x+\Theta)+x(1+\Theta)}\widetilde{P}_i^{\,d}\\ &+\frac{(1+\Theta)x}{\sigma(1+x+\Theta)+x(1+\Theta)}\alpha_i+\frac{\sigma x}{\sigma(1+x+\Theta)+x(1+\Theta)}\overline{\mu}\end{aligned} \tag{34}$$

Consider an equilibrium where, for all i, $a_i = k\alpha_i + \gamma \widetilde{P}_i^d + (1-k-\gamma)\,\overline{\mu}$, 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

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

Substituting for $E_j(\mu)$ and $E_j\left(P^d\right)$ from (33) and (34) 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 γ_0 described in proposition 4.

Consider now the case where μ and P^d are observable. 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 now

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

This gives

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

Solving out the system we obtain k_1 and γ_1 described in proposition 4.

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

$$\theta(1-\theta)\,\frac{x\sigma(1-\lambda) - (x+\sigma)\,\Theta\lambda}{(\theta+\lambda-\theta\lambda)\left(\theta(x+\sigma)(1-\lambda) + (\sigma+\lambda+\theta\Theta)\,x + (\sigma\lambda+\theta\Theta\sigma)\right)} \tag{35}$$

This is negative for $\lambda > \widehat{\lambda} \equiv \frac{x\sigma}{x\sigma+\Theta(x+\sigma)}$ and positive otherwise, where $\widehat{\lambda}$ is increasing in x and takes value $\frac{\sigma}{\Theta+\sigma(1+R)} < \frac{1}{1+R}$ when $x = \frac{\Theta}{R}$. An implication is that, provided that x is sufficiently small,

$\lambda > \widehat{\lambda}$ is consistent with $\lambda < 1/2$ and, more generally, with λ being relatively small. For instance, if $x = \Theta$ then $\widehat{\lambda} < 1/2$ while if $x = \frac{\Theta}{2}$ then $\widehat{\lambda} < 1/3$.

B.5 Discrepancy between overt attitudes and private preferences

We model the discrepancy between the overt attitudes adopted by an individual and the individual's private preferences as $\mid a_i - \alpha_i \mid$, the absolute value of the difference between a_i and α_i. From Propositions 1 and 2, we know that both $E\left(\overline{a}^{n_d}\right)_{before}$ and $E\left(\overline{a}^{n_d}\right)_{after}$ are increasing in P^d. Moreover,

$$\begin{aligned} E\left(\overline{a}^{n_d}\right)_{after} - E\left(\overline{a}^{n_d}\right)_{before} &= \mu\left(1-(1-\lambda)\left(k_1+\gamma_1\right)\right) - \overline{\mu}\left(1-\left(k_0+\gamma_0\right)\left(1-\frac{\lambda\Theta}{1+\Theta}\right)\right) \\ &\quad + P^d\left(\left(1-\lambda\right)\left(\left(k_1+\gamma_1\right)-\left(k_0+\gamma_0\right)\right) - \left(k_0+\gamma_0\right)\frac{\lambda}{1+\Theta}\right) \end{aligned}$$

decreasing in P^d since $k_1 + \gamma_1 < k_0 + \gamma_0$. Restricting attention to P^d such that $E\left(\overline{a}^{n_d}\right)_{after} > E\left(\overline{a}^{n_d}\right)_{before}$ – which, for $\overline{\mu}$ sufficiently low/negative, applies to all districts – the change triggered by the referendum in the discrepancy between over attitudes and own preferences of individual i is,

$$\Delta(\mid a_i - \alpha_i \mid) = (1-\theta)\begin{cases} \left(E\left(\overline{a}^{n_d}\right)_{after} - E\left(\overline{a}^{n_d}\right)_{before}\right) > 0 \text{ if (i) } E\left(\overline{a}^{n_d}\right)_{before} > \alpha_i \\ \left(E\left(\overline{a}^{n_d}\right)_{after} + E\left(\overline{a}^{n_d}\right)_{before} - 2\alpha_i\right) \text{ if (ii) } E\left(\overline{a}^{n_d}\right)_{after} > \alpha_i > E\left(\overline{a}^{n_d}\right)_{before} \\ \left(E\left(\overline{a}^{n_d}\right)_{before} - E\left(\overline{a}^{n_d}\right)_{after}\right) < 0 \text{ if (iii) } \alpha_i > E\left(\overline{a}^{n_d}\right)_{after} \end{cases}$$

where $\Delta(\mid a_i - \alpha_i \mid) \equiv \mid a_i - \alpha_i \mid_{before} - \mid a_i - \alpha_i \mid_{after}$. As we move from case (i) to case (ii) to case (iii) – corresponding to progressively higher α_i – the value of $\Delta(\mid a_i - \alpha_i \mid)$ decreases and eventually turns negative. Fixing α_i, we see that, when P^d is larger, case (i) is more likely to obtain and cases (ii) and (iii) are less likely. Moreover, conditional on $\Delta(\mid a_i - \alpha_i \mid) < 0$, a larger P^d causes it to be less negative.

All together, these observations imply that individuals who dislike immigrants were most likely to experience a significant reduction in the discrepancy between overt attitudes and own preferences following the referendum, but only if they happen to live in areas where immigrants are well liked.

B.6 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 conforming to the norm). Similar to Bursztyn

et 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

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

where P^d represents average preferences in i's district and $E\left(\alpha_i \mid a_i\right)$ 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 ϱ_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 ε is drawn as $N(0, \sigma)$. In the absence of further information, individual i's beliefs about P^d are

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

We consider two types of equilibria that may arise.[26] 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 $E\left(\alpha_i \mid a_i\right) = E\left[E\left(\alpha_i \mid a_i, \alpha_j, \widetilde{P}_j^d\right)\right] = \frac{a_i - r_1 P^d - (1-r_0-r_1)\overline{\mu}}{r_0}$ by the law of iterated expectations. Individual i maximizes

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

[26] 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 α_i, $\widetilde{P}_i^d$ and $\overline{\mu}$ also in the case of partial pooling.

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,

$$r_0 = \frac{r_0^2}{r_0^2\theta + 1 - \theta}\left(\theta + \frac{1-\theta}{r_0^2}\,(r_1 + r_0)\,\frac{(1+\Theta)\,x}{\sigma\,(1 + x + \Theta) + x\,(1+\Theta)}\right), \tag{36}$$

$$r_1 = \frac{1-\theta}{r_0^2\theta + 1 - \theta}\,(r_1 + r_0)\,\frac{(1+\Theta)\,\sigma}{\sigma\,(1 + x + \Theta) + x\,(1+\Theta)}. \tag{37}$$

Second, there may also be a pooling equilibrium, in which $a_i = \overline{\mu}$, independent of α_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,[27]

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

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

$$\overline{a}^d_{after} - \overline{a}^d_{before} = (1 - r_{before})\left(P^d - \overline{\mu}\right) \tag{38}$$

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.

[27] 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 Description of Survey

C.1 Descriptive Statistics

Respondent's characteristics

	Mean	SD	Min	Max	N
Age	0.405	13.45	18	83	1635
Female	0.500				1635
Num. members of household	2.79	1.28	1	10	1635
Employed full-time	0.574				1635
Employed part-time	0.153				1635
Unemployed	0.095				1635
Student	0.050				1635
Income	£20,000 - £49,999				1635
University level	0.564				1635
Vote Leave	49.36				1635
Vote Conservative	34.92				1635
Vote Labour	36.51				1635
Vote Other	28.56				1635

C.2 Geographical Distribution

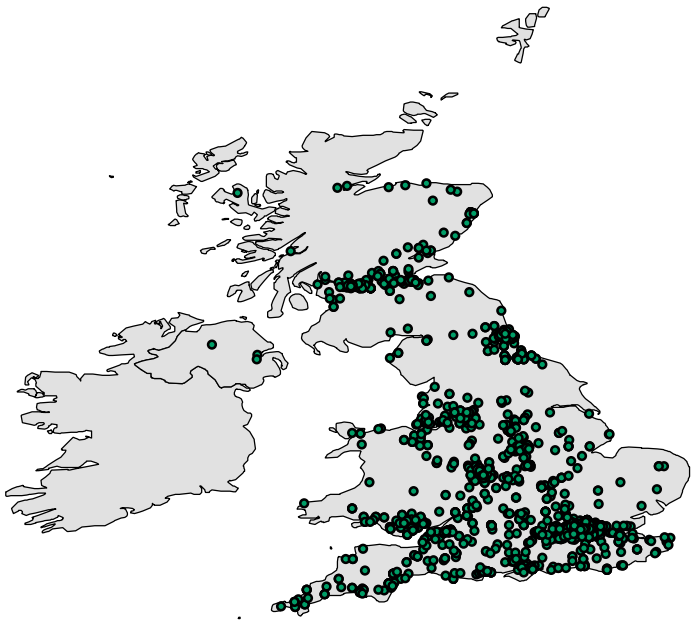


Figure C.2: Geographical distribution of survey respondents

C.3 Questionnaire

Welcome. Thank you for taking part in this study. In this study, you will be presented with information and asked to answer some questions. Please be assured that your responses will be kept completely confidential. In total, the study should take you less than 8 minutes to complete. By clicking the Continue button below, you acknowledge that your participation in the study is voluntary, that you are at least 18 years old, and that you are aware that you may choose to terminate your participation in the study at any time and for any reason, but you will not be paid if you do not finish. Please answer the following questions:

1. What is your gender? Male; Female; Prefer not to say
2. How old are you?
3. What is the highest educational level you achieved? GCSE or equivalent; A levels or equivalent; Apprenticeship; university degree and above; other: No formal qualification

4. What is your current employment status employed full-time; employed part-time; unemployed; student; other

5. Income - What was your household's pre-tax income last year? less than £10,000; £10,000 - £19,999; £20,000 - £49,999; £50,000 - £79,999; £80,000 - £149,999; More than £150,000

6. Please write the first 3 characters of your current postcode (e.g. XB9)

7. Is this the same postcode you had at the time of the Brexit referendum (23rd of June 2016)? Yes; No

8. If answered No to previous question. Please write the first 3 characters of your postcode at the time of the Brexit referendum (e.g. TB9)

9. How did you vote in the 2016 Brexit referendum? Leave; Remain; did not vote

10. In the UK, what was the share of votes in favour of leaving the EU? If you do not know the answer, please provide your best guess

11. On a scale of 0 to 10, with 0 representing not important at all and 10 representing extremely important, how important do you think that concerns over immigration and erosion of British identity were for the outcome of the 2016 Brexit referendum?

12. Going back to June 2016, when the result of the Brexit referendum came out. Can you describe how surprised you felt at the time by the outcome of the referendum across the UK? Use a scale where 0 is not surprised at all and 10 Extremely surprised

13. In the local area where you were living at the time of the 2016 Brexit referendum, what was the share of votes in favour of leaving the EU? If you do not know the answer, please provide your best guess 5o Much smaller

14. On a scale of 0 to 10 with 0 representing very negatively and 10 representing very positively, how do you view immigrants?

15. On a scale of 0 to 10 with 0 representing very negatively and 10 representing very positively, how do you think people in the UK view immigrants?

16. On a scale of 0 to 10 with 0 representing very negatively and 10 representing very positively, how do you think people in your local area view immigrants?

17. Did you change YOUR views towards immigrants after the Brexit referendum? Yes, they became more negative;Yes, they became more positive; No, they did not change

18. Whatever your views on immigrants, did the outcome of the 2016 Brexit referendum change how comfortable you are in expressing those views? After the referendum, I became "..." to publicly express my views on immigrants. Much more likely; more likely; neither more nor less likely; less likely; much less likely

19. Do you think that the 2016 Brexit referendum had an effect on how comfortable people in your local area are about publicly expressing their views on immigrants? Please select the option that corresponds most closely to your impression. After the referendum, people in my area became "..." to publicly express their views on immigrants. Much more likely; more likely; neither more nor less likely; less likely; much less likely

20. How important is your country in defining your identity? Use a scale where 0 represents "not important at all" and 10 "very important"

21. Including yourself, how many people are currently living in your household?

22. Were you born in the UK?

23. On the UK political spectrum, where do you position yourself? left-wing; centre-left; centre; centre-right; right-wing; none of the above

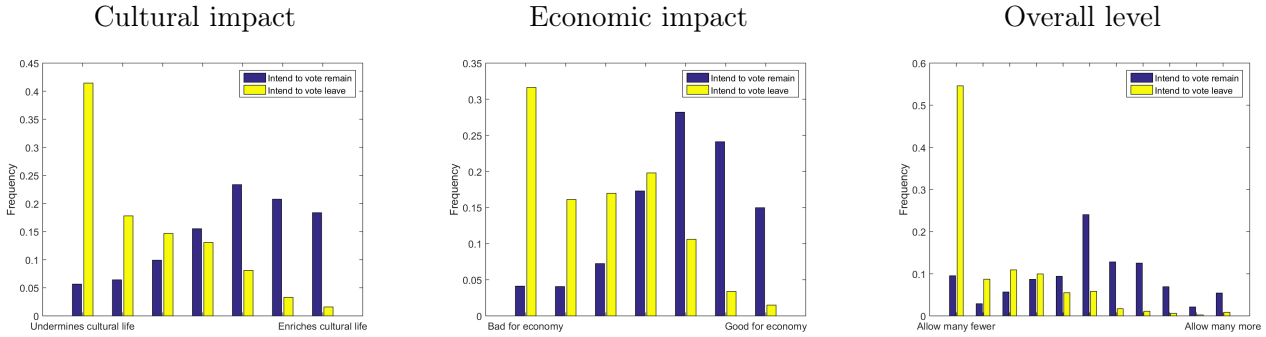
24. Which party did you vote for in the last general election? Conservative; Labour; lib-dem; SNP; other ; I did not vote

Before you leave this study, we would like to ask your feedback. In particular, was the study clear? If not, which part did you struggle to understand?

D Quantitative Appendix

D.1 Views on immigration

Figure D.1: 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.

D.2 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.