CAGE

Working Paper

802/2026
May 2026

Information Shocks, Attitudes toward Immigrants, and Hate Crime

Jake Bradley,®

Facundo Albornoz,®

Silvia Sonderegger, ®

Jesús Rodríguez, ®

Devesh Rustagi ®

® *Author names randomised using the AEA random author tool. See page 1 for full information. Corresponding authors are Devesh Rustagi and Silvia Sonderegger.*

ISSN: 2978-0276
Grant number: ES/7504701/1

UNIVERSITY OF WARWICK

UKRI Economic and Social Research Council

Information Shocks, Attitudes toward Immigrants, and Hate Crime*

Jake Bradley[1] ⓡ Facundo Albornoz[1] ⓡ Silvia Sonderegger[1] ⓡ

Jesús Rodríguez[1] ⓡ Devesh Rustagi[2] ⓡ

[1]University of Nottingham [2]University of Warwick

Abstract

There are concerns over the rise in populism and hate crimes targeting minorities in democracies. We examine whether national information shocks triggered by political events play a role. Focusing on two UK events that revealed nationwide anti-immigrant sentiment, we document counterintuitive results: large persistent surges in hate crimes in the post-event periods in areas with pro-immigrant, rather than anti-immigrant, attitudes. We show that the xenophobic minority residing in pro-immigrant areas experience stronger belief shocks from these events, inducing them to update their beliefs about social acceptability of hate. Our findings highlight how heterogeneous priors interact with national events to amplify xenophobic behavior.

JEL: C72, D80, P0

Keywords: Information shocks, attitudes towards immigrants, hate crimes, United Kingdom

*The author names were randomized using the AEA random author tool. The confirmation code is EibU4ZXnVGZZ. Corresponding authors: Devesh Rustagi, devesh.rustagi@warwick.ac.uk; Silvia Sonderegger, silvia.sondergger@nottingham.ac.uk

I. Introduction

There are global concerns over the rise in populism and hate crimes targeting minorities in democratic societies (Guriev and Papaioannou, 2022). Yet, our understanding of the mechanisms underlying this surge is still evolving. There is growing interest in studying the role of information shocks from political events in fueling surges in hate crimes by revealing previously unknown support for anti-immigrant sentiments.[1] In a recent study, Bursztyn, Egorov and Fiorin (2020) leveraged *local* information shock on Donald Trump's victory in 2016 presidential election to show that it triggered rapid shift in *local* social norms, making public display of xenophobic behavior more acceptable in places with stronger xenophobic views. In this paper, we document for the first time, a counterintuitive result: when information shocks from *nationwide* political events reveal anti-immigrant sentiment in the entire country, they result in large surges in hate crimes not in places with xenophobic views but in those with *pro-immigrant* views.

How can we explain this result? We develop a conceptual framework that has two key ingredients. First, the scope of the information shock is not the same for all individuals, but depends on heterogeneous priors (see Cantoni et al., 2019). We argue that these priors are not randomly distributed across individuals but are rooted in local areas where the individuals reside. This means that individuals form beliefs about nationwide sentiments toward immigrants by extrapolating from sentiments in their local area. In pro-immigrant areas, individuals believe that the nation is friendly toward immigrants, but in anti-immigrant areas they believe xenophobia to be widespread. Consequently, when a national political event publicly reveals countrywide anti-immigrant sentiment, individuals in areas with similar views experience smaller belief shocks, as the signal largely confirms their priors. In contrast, individuals in pro-immigrant areas experience larger belief shocks, as the signal is counter to their priors. The xenophobic minority in pro-immigrant areas updates its beliefs and engages in xenophobic behavior like hate crimes in the post-event periods. Second, we argue that this behavioral shift occurs because individuals seek to comply not only with local norms but also with national norms, driven by considerations such as national identity (see Akerlof and Kranton, 2000; Shayo, 2009). When national political events reveal countrywide normalization of anti-immigrant sentiments, they make it socially more acceptable to engage in hate crimes.

Our study takes place in the context of the United Kingdom, which experienced two prominent information shocks that revealed previously hidden anti-immigrant sentiment at the national level. The first information shock was the unexpected victory of the UK Independence Party (UKIP) in the 2014 European Parliament elections. UKIP not only increased its vote share by over 66 percent, but became the first party outside of

[1] See for instance the literature on the role of information shocks in dispelling pluralistic ignorance pioneered by Kuran (1991) and Lohmann (1994).

Labour and the Conservatives to win a national election since 1910. The Guardian, a prominent UK newspaper, described the event as a "political earthquake." The second information shock was the unexpected Brexit referendum outcome in June 2016, which ultimately led to the UK's withdrawal from the European Union in January 2020. The BBC characterized the result as a "huge surprise." Both events were marked by explicit anti-immigrant rhetoric and were followed by large and persistent surges in hate crimes in the post-event periods.

We consider hate crimes motivated by race, ethnicity, and religion, which in the UK are subject to stricter sentencing than equivalent crimes without such motivation.[2] We collect quarterly data on hate crimes at the level of local administrative units called Community Safety Partnerships (CSP), which are empowered to formulate and implement strategies to tackle local crime. Our sample covers 97% of the CSPs in England and Wales over an 18-year period spanning 2002–2019.[3] We measure attitudes towards immigrants in economic and cultural domains using data from British Election Study (BES), conducted prior to the events. The respondents were asked to indicate their support for immigrants on two questions — "Are immigrants good or bad for the British economy?" and "Do immigrants undermine or enrich Britain's cultural life?". Data from other waves reveal that these attitudes are remarkably stable over time. We refer to CSPs with stronger attitudes as 'immigrant friendly or pro-immigrant' and those with weaker attitudes as 'anti-immigrant'.

We begin by documenting four facts in support of our conceptual framework, using data from BES. First, our framework requires variation in attitudes within CSPs. We show that this is indeed the case. Crucially, in immigrant friendly CSPs, a non-trivial share of individuals hold strong anti-immigrant attitudes. Second, we argue that people form beliefs about national sentiments by extrapolating from local sentiment toward immigrants. We verify empirically that people are aware of local sentiments. Third, we argue that belief shocks from political events depend on heterogeneous priors. To this end, we show that individuals with pro-and anti-immigrant attitudes residing in immigrant friendly CSPs experienced a larger belief shock than their counterparts residing in anti-immigrant CSPs. Finally, national identity can operate as a mechanism for compliance with national norms only in environments where it matters to individuals. We document a strong and stable sense of national identity across all CSPs.

We examine whether CSPs with stronger attitudes saw larger surges in hate crime in the post-event periods than CSPs with weaker attitudes using a difference-in-difference design. In our setting, all CSPs experienced UKIP and Brexit events at the same point in time, but there is a heterogeneous response depending on attitudes toward immigrants.[4]

[2]Examples include racist graffiti sprayed on someone's home or property, verbal abuse in the form of using slurs, and physical assault accompanied by racist language.

[3]We exclude data after 2019 because of confounding with COVID-19.

[4]Though UKIP and Brexit events occur at different points in time, every CSP is treated in each of these

Our specification includes CSP level and quarter fixed effects, as well as region specific time trends. This allows us to account for unobserved time-invariant differences across CSPs and common shocks that affect all CSPs and regions similarly in a given quarter. In addition, we control for CSP level covariates that change over time including population size, flow of EU immigrants, and flow of non-EU immigrants. Conditional on fixed effects and controls, the identifying assumption is that, absent the UKIP and Brexit events, hate crimes would have evolved similarly in CSPs with different attitudes toward immigrants. The event study plots show a large surge in hate crime in the post-event periods, but clear absence of pre-trends, lending credibility to our identifying assumption.

We foresee two empirical challenges. First, previous studies note that economic, cultural, and social factors led to support for UKIP and Brexit (Becker, Fetzer and Novy, 2017; Fetzer, 2019; Guriev and Papaioannou, 2022). To separate the effect of the information revealed by the events from that of the factors that gave rise to them, we include post-event interactions with austerity, education deprivation, generalized trust, and trust in politicians. A related concern is that poor economic conditions arising from the events themselves led to surges in hate crime. However, we observe an immediate rise in hate crime in the post-event periods, whereas economic changes appeared with a lag of several years (see for instance, Bakker et al., 2022).

Second, our results could reflect changes in the recording of hate crimes, particularly due to increased sensitivity among the police and directives to record hate crimes more accurately in the post-event periods. We address this concern by controlling one at a time for police-force specific time trends and an interaction between post-event and police-force indicators. A related concern is that reported hate crimes may have risen in the post-event periods not because of an increase in their frequency, but because of an increase in reporting. Addressing this concern fully is very difficult, but we attempt to assuage its scope by using data from the Crime Survey of England and Wales on reporting rates of different types of crimes that also have hate counterparts. Using an event study, we show that reporting rates did not change in the post-event periods. We then leverage CSP level variation in the type composition of hate crime to adjust our measure of hate crime using type-specific reporting rates.

We find a strong positive association between attitudes toward immigrants and surges in hate crimes in the post-event periods, which is statistically significant at the 1% level. This result holds when we include full set of economic, cultural, and social controls interacted with post-event period, or use adjusted measures of hate crime. Our estimates suggest that one standard deviation increase in attitudes is associated with a rise in hate crime in the post-event period by close to 0.2 SD or 20% of the mean. The increase in hate

events. So, this is not a case of staggered treatment in which different units are treated at different points in time. Our strategy is similar to that of Korovkin and Makarin (2023) who study the effect of Russia-Ukraine conflict on trade patterns depending on the share of Ethnic Russians in the districts of Ukraine.

crime does not diminish over time. This means our findings cannot be due to busloads of xenophobic people traveling to immigrant friendly areas to commit hate crimes, as this can explain short-lived surges over a quarter, but not a sustained effect over 16 quarters. Our results are not due to rise in polarization wherein partisan group members engage in hate crimes against each other in the post event-periods. To this end we show that our results hold when we consider different types of hate crimes. A falsification test shows no effect of attitudes on murder and burglary in the post-event periods, which is expected as these crimes are unlikely to arise from information shocks generated by the UKIP and Brexit events.

We now turn to understanding the role of belief shocks as a mechanism underlining our results. We construct belief shock as the difference between the event realization and prior belief about it, separately for each event. Both the event study and difference-in-differences estimate show that CSPs with stronger belief shocks experienced significantly higher surges in hate crimes by 0.07 – 0.14 SD in the post-event periods. Since it is not possible to get data on hate crime at the individual level, we carry out an exploratory survey with close to 1500 participants. We collect data on their attitudes, immigrant sentiment in their area, and their willingness to express views over immigrants. We find that individuals with anti-immigrant attitudes living in immigrant-friendly CSPs are significantly more willing to express their views about immigrants in the post-event periods. This is consistent with the notion that the information shock — and the subsequent belief adjustment — was stronger in immigrant-friendly areas.

Our paper contributes to the literature on rise in populism and hate crimes in democracies (Guriev and Papaioannou, 2022). Several studies show that areas with historically persistent xenophobia are more likely to experience surges in hate crimes (Voigtländer and Voth, 2012), especially in response to persuasion campaigns and coordination facilitated by radio (Adena et al., 2015), and social media (Bursztyn et al., 2019). A closely related paper by Bursztyn, Egorov and Fiorin (2020) shows higher xenophobic behavior in response to local information shocks revealing xenophobia in local population. A common theme across these impressive studies is that surges in hate crimes and xenophobic behavior occur in areas that were already xenophobic. We are the first to show that when events reveal nationwide anti-immigrant sentiment and there are heterogeneous priors rooted in local areas, then this can result in counterintuitive increases in hate crime in pro-immigrant areas. Our study is also related to Cantoni et al. (2019) who examine the role of heterogeneous priors for participation in anti-authoritarian protests in Hong Kong. By leveraging how priors are formed in the context of a community, we can predict not only for whom but also where the change in behavior occurs.

Our paper also contributes to the emerging literature on hate and toxicity on social media (Bursztyn et al., 2019; Jiménez-Durán, 2023; Jiménez Durán, Müller and Schwarz, 2024; Kalra, 2024; Beknazar-Yuzbashev et al., 2025). These studies show that exposure to extremist or hateful content can reveal latent preferences, reduce the perceived social costs of expressing xenophobia, and facilitate coordination among like-minded individuals, often leading to increases in hateful speech. Fages and Martínez-Pozo (2026) document a temporary rise in sexist and homophobic speech on social media in the aftermath of Bolsonaro's election in areas where he received weaker vote share. However, this study does not posit any mechanisms underlying this result. Our paper complements this literature by focusing on offline hate crime and by identifying national political events—rather than platform-level or algorithmic shocks—as a source of information. Also, by emphasizing heterogeneous priors formed from local environments, we show that information shocks can disproportionately activate xenophobic minorities in otherwise pro-immigrant areas.

II. Conceptual Framework

In this section we first propose a theoretical model of how individuals choose their behavior towards immigrants based on their personal preferences and their beliefs about the behavior of other individuals. We then discuss how the insights we obtain can be applied to the specific case of hate crime.

Background.– We consider a country of measure 1 with a continuum of individuals who are divided into geographical areas of equal size (for simplicity). All individuals move simultaneously. Each individual i in area d selects an *action* $a_i \in \mathbb{R}$ to maximize expected utility. This depends on (i) how closely the action matches a preference parameter $\alpha_i \in \mathbb{R}$ reflecting the individual's intrinsic preferences, and (ii) how closely the action conforms to a reference action $\overline{a}_R^d$, defined as $\overline{a}_R^d \equiv \lambda\overline{a} + (1-\lambda)\,\overline{a}^d$, where $\overline{a}^d$ and $\overline{a}$ represent mean actions in area d and countrywide, respectively, and $0 < \lambda \leq 1$.

The reference action $\overline{a}_R^d$ captures the desire to conform to local as well as countrywide behavior, with the relative importance of the latter being parameterized by λ. These conformity concerns may originate from a desire to fit in, to behave in a socially acceptable way – where social acceptability is defined by a combination of local and countrywide components – or to behave in a way that is consistent with one's identity – as e.g. in George A Akerlof and Rachel E Kranton (2000); Roland Bénabou and Jean Tirole (2011).[5] In this interpretation, λ captures the extent to which individuals identify with the country as opposed to their local area.

[5] We follow George A Akerlof (1980); Moti Michaeli and Daniel Spiro (2017); P Grout, S Mitraille and S Sonderegger (2015) and Cristina Bicchieri, Eugen Dimant and Silvia Sonderegger (2023) in adopting a consequentialist approach, in the sense that an individual's utility depends on how the individual behavior relates to the reference behavior. Another branch of the literature, such as B Douglas Bernheim (1994); Roland Bénabou and Jean Tirole (2011) or Fabrizio Adriani and Silvia Sonderegger (2019), focuses instead on the case where behavior generates social or psychological utility only to the extent to which it reveals information about an individual's underlying type.

Individual preferences.– The individual preference parameter is given by the sum of a area-specific component P^d, and an individual-specific component, ε_i. The area-specific component is in turn equal to the sum of a mean preference parameter μ and a component that captures area-specific characteristics, e^d. To sum up, therefore, the preference parameter α_i^d of an individual i in area d is

$$\alpha_i^d = \underbrace{\mu + e^d}_{=P^d} + \varepsilon_i \tag{1}$$

where μ represents (unobservable) countrywide mean preferences, $e^d \sim N(0,1)$ is a area-specific component and $\varepsilon_i \sim N(0,\sigma)$ is an individual-specific component.

Information.– Consistent with the empirical results provided in Figure 3 below, people have a good grasp of preferences at the local level. Formally, this is captured by letting them observe P^d, their local mean preferences.[6] In addition, all individuals observe a public signal $\overline{\mu} = \mu + \xi$ where $\xi \sim N(0,\Theta)$. We can think about $\overline{\mu}$ as the countrywide, publicly available information about μ. This may come from shared history and cultural traits, as well as information-revelation events at country level, such as elections or referenda.

Beliefs.– The beliefs about μ in area d are a weighted average of the information available to individuals, namely local preferences P^d and the public information $\overline{\mu}$,

$$E(\mu \mid P^d, \overline{\mu}) = \frac{\Theta}{1+\Theta} P^d + \frac{1}{1+\Theta}\overline{\mu}. \tag{2}$$

This highlights that beliefs are *heterogeneous* across areas. Intuitively, when forming beliefs about countrywide preferences, people partially extrapolate from local preferences.[7] Individuals living in pro-immigrant areas tend to believe that the country as a whole is pro-immigrant, while those living in anti-immigrant areas tend to hold the opposite belief.

Utility.– The utility of any individual i in area d with personal preference α_i is composed of two loss functions, capturing the psychological cost of deviating from one's own personal preferences as well as the social and psychological cost of deviating from the reference action. Formally,

$$u_i = -\theta \left(a_i - \alpha_i\right)^2 - (1-\theta)\left(a_i - \overline{a}_R^d\right)^2. \tag{3}$$

[6]The analysis can be easily extended to the case where P^d is not fully observable, at the cost of additional analytical complexity.

[7]This is superficially reminiscent of the false consensus effect, a well known psychological bias that refers to the tendency of people to overestimate the extent to which their preferences (in this case, preferences in their local areas) are typical of those of others (here, other areas). However, in our analysis, using preferences in one's local area to predict countrywide preferences is not due to a bias but is actually consistent with Bayesian updating – see also Christoph Vanberg (2019) and Fabrizio Adriani and Silvia Sonderegger (2015) for illustrations of this general point.

where $\theta \in (0,1)$ captures the concern for aligning behavior with personal preferences relative to conforming with the reference behavior and, as mentioned, $\overline{a}_R^d \equiv \lambda \overline{a} + (1-\lambda)\,\overline{a}^d$.

II.A. Equilibrium Characterization

Our equilibrium concept is Perfect Bayesian Equilibrium. Each individual i in area d chooses a_i to maximize the expectation of (3), where expectations are consistent with equilibrium behavior. Differentiating the objective function and rearranging delivers,

$$a_i = \theta\alpha_i + (1-\theta)\,E(\overline{a}_R^d \mid P^d, \overline{\mu}) \tag{4}$$

where $E(\overline{a}_R^d \mid P^d, \overline{\mu}) = \lambda E(\overline{a} \mid P^d, \overline{\mu}) + (1-\lambda) E(\overline{a}^d \mid P^d, \overline{\mu})$.

Expression (4) shows that each individual's action is a weighted average of own preferences and expectations about the reference action. Through this mechanism, the model generates strategic complementarities between individual behavior and mean behavior at both the local and country levels.[8]

Proposition 1 *In the unique linear symmetric equilibrium of the game, we have*

$$a_i = \theta\alpha_i + \gamma P^d + (1-\theta-\gamma)\overline{\mu} \tag{5}$$

where $\gamma \equiv \frac{\theta(1-\theta)(\Theta+1-\lambda)}{\theta+\lambda(1-\theta)+\theta\Theta} < 1-\theta$ *increasing in* Θ.

***Proof*:** The proof is provided in the Appendix.

Corollary 1 (i) *Mean countrywide behavior is* $\overline{a} = (\theta+\gamma)\,\mu + (1-\theta-\gamma)\,\overline{\mu}$.

Individuals form their expectations about countrywide behavior from their beliefs about countrywide preferences: $E(\overline{a}|P^d, \overline{\mu}) = (\theta+\gamma)E(\mu|P^d, \overline{\mu}) + (1-\theta-\gamma)\overline{\mu}$, where $E(\mu\,|P^d, \overline{\mu})$ is given by (2). Substituting for this and for $\overline{a}^d$ in expression (4) we obtain (5).

Corollary 1 (ii) *Mean behavior in area d is* $\overline{a}^d = (\theta+\gamma)\,P^d + (1-\theta-\gamma)\,\overline{\mu}$.

The term $(\theta+\gamma)P^d$ in $\overline{a}^d$ reflects behavioral heterogeneity across areas. The first component, θ, captures people's desire to conform to their own (area-specific) preferences, while the second component, γ, captures cross-area heterogeneity in beliefs about μ.

II.B. Effect of Information Shock on Behavior

We now consider the effect of an information shock that releases new public information about μ, the mean countrywide preferences. Let $\overline{\mu}_0 \sim N(\mu, \Theta_0)$ be the pre-shock publicly

[8] For analytical convenience, we consider a continuum of areas.

observable signal about μ, so that $E(\mu \mid \overline{\mu}_0) = \overline{\mu}_0$, $Var(\mu \mid \overline{\mu}_0) = \Theta_0 > 0$. The post-shock public information is denoted as $\overline{\mu}_1$, with $E(\mu \mid \overline{\mu}_1) = \overline{\mu}_1$, $Var(\mu \mid \overline{\mu}_1) = \Theta_1 < \Theta_0$.[9] Without loss of generality, we let higher values of a (resp., α) capture more pro-immigrant behavior (resp., preferences). We consider the case $\overline{\mu}_1 < \overline{\mu}_0$, i.e. the new information indicates that countrywide preferences are less pro-immigrant than previously thought.

We start by assessing cross-area differences in the effect of the information shock on beliefs. As shown in (15), beliefs about μ are heterogeneous across areas. This implies that the change in beliefs generated by the shock is also heterogeneous.

Lemma 1 *The difference in beliefs in area d following the shock is*

$$E(\mu \mid P^d, \overline{\mu}_0) - E(\mu \mid P^d, \overline{\mu}_1) = \frac{1}{1+\Theta_0}\overline{\mu}_0 - \frac{1}{1+\Theta_1}\overline{\mu}_1 + \frac{\Theta_0 - \Theta_1}{(1+\Theta_1)(1+\Theta_0)}P^d, \tag{6}$$

increasing in P^d.

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

Intuitively, information indicating that the country as a whole is anti-immigrant is less surprising to individuals living in anti-immigrant areas, because their prior beliefs already suggested that the country may be anti-immigrant.

The next result highlights the implications of Lemma 1 for the cross-area change in behavior triggered by the shock. Since the new information points to more anti-immigrant preferences at country level than previously thought, behavior generally becomes more anti-immigrant, due to strategic complementarity between individual and countrywide behavior.[10] However, the *size* of this shift differs across areas. Denote as γ_t, $t \in \{0, 1\}$, the value of γ when $\Theta = \Theta_t$. Note that, since $\Theta_0 > \Theta_1$, we have $\gamma_0 > \gamma_1$.

Corollary 2 *The change in mean behavior in area d following the shock is*

$$\overline{a}_0^d - \overline{a}_1^d = (1 - \theta - \gamma_0)\overline{\mu}_0 - (1 - \theta - \gamma_1)\overline{\mu}_1 + (\gamma_0 - \gamma_1)P^d, \tag{7}$$

increasing in P^d.

In words, the model predicts that more anti-immigrant areas experience a smaller behavioral response to the information shock. That's because individuals in anti-immigrant

[9] Formally, let the additional signal released by the information shock be $\overline{\mu}_2 = \mu + \xi_2$ where $\xi_2 \sim N(0, \Theta_2)$. By Bayesian updating, the posterior belief conditional on both $\overline{\mu}_0$ and $\overline{\mu}_2$ is that μ is normally distributed around $\overline{\mu}_1 \equiv \frac{\Theta_2 \overline{\mu}_0 + \Theta_0 \overline{\mu}_2}{\Theta_0 + \Theta_2}$ with variance $\Theta_1 \equiv \frac{\Theta_0 \Theta_2}{\Theta_0 + \Theta_2}$.

[10] We are implicitly focusing on the empirically relevant case where $E(\mu \mid P^d, \overline{\mu}_1) < E(\mu \mid P^d, \overline{\mu}_0)$ and $\overline{a}_1^d < \overline{a}_0^d$.

areas are less surprised by the new information than those in immigrant loving areas and therefore adjust their behavior to a lesser extent.

II.C. Implications for Hate Crime

In this section we describe how the setup of the previous section maps into hate crime. We think of hate crime as an extreme expression of overt dislike towards immigrants (or, more generally, people from a minority background). The simplest way to capture this is to consider a threshold $-\mathbf{a} < 0$, such that, when behavior is below this threshold, it is classified as hate crime. Recall that, in our setup, utility is

$$\begin{aligned} u_i &= -\theta\left(a_i - \alpha_i\right)^2 - (1-\theta)\left(a_i - \overline{a}_R^d\right)^2 && (8) \\ &= 2a_i\left[\theta\alpha_i + (1-\theta)\,\overline{a}_R^d\right] - a_i^2 + K && (9) \end{aligned}$$

where $K \equiv -\theta\alpha_i^2 - (1-\theta)\left(\overline{a}_R^d\right)^2$. This makes clear that, although we have expressed utility as a combination of two loss functions, we can equivalently think of it as the result of the trade-off between the net psychological and social return from selecting action a_i — given by $2a_i\left[\theta\alpha_i + (1-\theta)\,\overline{a}_R^d\right]$ — and the expected material cost arising from the legal sanctions associated with a_i — given by a_i^2. A possible micro-foundation for the expected cost of legal sanctions is the following. All extreme behaviors —whether strongly anti-minority ($a_i < -\mathbf{a}$) or strongly anti-majority ($a_i > \mathbf{a}$) — are considered crimes, with the former categorized as hate crimes, and people are uncertain about the exact value of $\mathbf{a}$.[11]

The effect of information shock on hate crime.– Let $\widetilde{a}_t^d \equiv (-\mathbf{a} - \overline{a}_t^d)/\theta\sqrt{\sigma}$ where $\overline{a}_t^d = (\theta+\gamma_t)P^d + (1-\theta-\gamma_t)\overline{\mu}_t$, $t \in \{0,1\}$, and denote as $F\left(.\right)$ the cdf of the standard normal distribution. In area d, the change in hate crime generated by the shock is equal to $F(\widetilde{a}_1^d) - F(\widetilde{a}_0^d)$. Differentiating with respect to P^d we obtain

$$\frac{1}{\theta\sqrt{\sigma}}\left[f(\widetilde{a}_0^d)\left(\gamma_0 - \gamma_1\right) + \left(\theta + \gamma_1\right)\left(f(\widetilde{a}_0^d) - f(\widetilde{a}_1^d)\right)\right] \tag{10}$$

where, as already observed, $\gamma_0 > \gamma_1$. There are two effects at play. The first is that, as highlighted in Corollary 2, the behavioral shift induced by the shock is less pronounced in anti-immigrant areas. In (10), this is captured by $f(\widetilde{a}_0^d)\left(\gamma_0 - \gamma_1\right) > 0$. The second effect relates to the probability density of the preference parameters of those people who are pushed below the hate crime threshold by the information shock. In the Appendix, we prove that if the within-area variance of individual preferences sufficiently large, expression (10) is always positive. The increase in hate crime following the shock is thus

[11] To fix ideas, consider $a_i < 0$ (the case $a_i > 0$ is analogous). Suppose that, from the public's perspective, $\mathbf{a} \sim U[0, \overline{\mathbf{a}}]$, where $\overline{\mathbf{a}}$ is "large", and let the cost of being convicted for a hate crime be $b\left(-a_i - \mathbf{a}\right)$. When $b = 2\overline{\mathbf{a}}$, the expected cost associated with a_i is $b\Pr(-\mathbf{a} > a_i)\,E(-\mathbf{a} - a_i \mid -\mathbf{a} > a_i) = a_i^2$ for all $a_i \geq -\overline{\mathbf{a}}$.

unambiguously larger in more immigrant loving areas.

III. Field Setting and Data

The UK provides an ideal setting for our study because it witnessed two prominent nationwide information shocks that exposed previously hidden anti-immigrant sentiment at the national level. We provide an overview of these events below, followed by data on hate crime and attitudes toward immigrants. Subsequently, we establish four motivating facts in support of our conceptual framework.

III.A. National Political Events

UKIP Election.– The first information shock is the unexpected victory of the UK Independence Party's (UKIP) in the European Parliament elections held in May 2014. UKIP increased its vote share by more than 66 percent, from 16.5% in 2009 to 26.6% in 2014. It was the first time since 1910 that a party other than Labour or Conservative had won the largest number of seats in any election. A leading newspaper in the UK, the Guardian, reported this as a "political earthquake."[12] UKIP was the only party that explicitly advocated for leaving the European Union and used xenophobic rhetoric. Its surprise victory signaled more widespread anti-immigrant sentiment in the population than previously expected.

Brexit Referendum.– The second information shock is the Brexit referendum that was held in June 2016 and which resulted in the UK eventually leaving the European Union (EU) on 31 Jan 2020. The referendum outcome was unexpected. On voting day, British bookmakers placed the likelihood of Leave victory between 13% and 17%.[13] The BBC described it as "a huge surprise".[14] The Leave campaign made explicit use of anti-immigrant rhetoric, framing immigration as a central issue in its appeal to voters

III.B. Data

Our unit of analysis is a Community Safety Partnerships (CSP). We focus on England and Wales, as data from Scotland and Northern Ireland are not available at the local level. In England and Wales, crime data are recorded by 43 Police Force Areas. Each police force typically oversees several CSPs within its jurisdiction. The CSPs are empowered under the Crime and Disorder Act (1998) to formulate and implement strategies to tackle local

[12] Available at The Guardian, May 2014. UKIP wins EU elections with ease to set off political earthquake.
[13] The following newspaper article provides an overview of different bookmakers: New Statesman, June 2016. Latest Brexit betting: what are the odds on the EU referendum?
[14] Available at BBC News, June 2016. Brexit: Europe stunned by UK Leave vote.

crime, disorder, and antisocial behavior in their communities. Our sample comprises 304 of the 315 CSPs in England and Wales.[15]

We collect data on hate crimes, attitudes towards immigrants, beliefs on the likelihood of UKIP victory and Brexit success, British national identity, and other crimes. We supplement these data with a primary survey to measure individual propensity to express personal views toward immigrants following the Brexit referendum. We describe these data below, except for data from the primary survey, which we present in section VI.

Hate Crimes.– A criminal offense is recorded as hate crime if it is perceived to be motivated by hostility or prejudice towards someone based on 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 data on hate crimes are publicly available from the Office for National Statistics. It includes violence without injury, public order offenses, criminal damage, and violence with injury.[16] Our data on hate crimes covers 71 quarters and 18 years, from Q2 2002 to Q4 2019. We do not use data from Q1 2020 onward because of confounding with the COVID pandemic.[17]

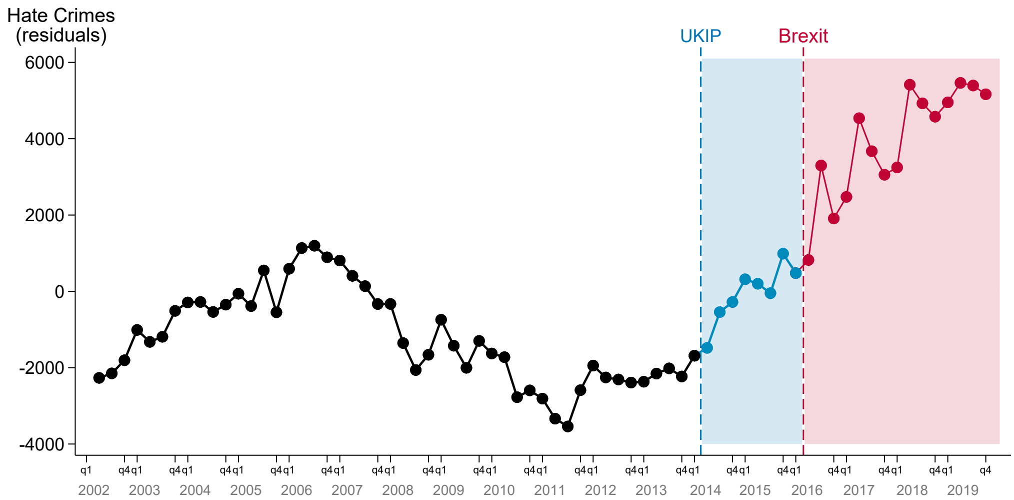

Figure 1: Evolution of Hate Crimes over Time (Residuals)

Notes. The figure depicts the evolution of hate crimes from Q2 2002 - Q4 2019. The Y-axis displays the residuals from a CSP-level regression of hate crimes on season indicators and an indicator variable for the 7/7 London Bombing (2005). The blue solid line corresponds to the post-UKIP period, while the red line corresponds to the post-Brexit period. Data are from Office for National Statistics.

Figure 1 plots total hate crime in England and Wales over time after accounting for season fixed effects and an indicator for the 7/7 London bombing in 2005. We observe

[15] Five local authorities comprise multiple CSPs, so we assign data in these authorities to the most populous CSP.

[16] Hate crime also includes crimes related to grievous bodily harm. These crimes make up less than 3 percent of all crimes. We exclude these crimes because they do not relate to our mechanism. However, given their small share, our results remain unchanged when we include them.

[17] Hate crimes in one CSP in Wales (Merthyr Tydfil) are not reported separately from Q2 2018 onwards, as it was merged with another CSP (Rhondda Cynon Taf) in April 2018.

a steep and steady surge in hate crimes in post-UKIP (shaded in blue) and post-Brexit (shaded in red) periods. Panel A of Table A.1 reports summary statistics. The average hate crime per CSP per quarter is 31.33 (s.d. 42.78). In the econometric analysis, we use standardized measure of hate crime, with mean 0 and standard deviation of 1.

Attitudes Toward Immigrants.– We capture sentiments towards immigrants using data on attitudes toward immigrants from Wave 1 of the British Election Study (BES) – an individual level panel survey with approximately 23,000 respondents from England and Wales (on average 75 respondents per CSP). Wave 1 was conducted two months before the UKIP event and it is the earliest date for which we have data on attitudes.

We measure attitudes toward immigrants (hereafter, attitudes) using responses to two questions that encompass economic and cultural dimensions, and can be considered as capturing important cultural traits like tolerance and respect (Tabellini, 2010). These include: (a) "Do you think immigration is good or bad for Britain's economy?" The respondents could choose their answer on a scale of 1-7, where 1 implies "bad for the economy" and 7 implies "good for the economy"; and (b) "Do you think that immigration undermines or enriches Britain's cultural life?". The respondents could choose their answer on a scale of 1-7, where 1 implies "undermines" and 7 implies "enriches." Since both questions use the same scale, we take the average of individual responses at the CSP level. The summary statistics in Table A.1 show that the average attitude is 3.43 (s.d. 0.48). However, there is large variation between CSPs, which ranges from 1.98 to 5.05. In the econometric analysis, we use standardized measure of attitudes with mean 0 and standard deviation of 1. Our results hold when we consider each question separately.

Figure A.1 shows that attitudes are fairly stable over time and do not decline in response to the political events. If at all, they increase slightly in magnitude over time. Given the centrality of attitudes to our analysis, we validate our approach by showing that attitudes have meaningful associations with political outcomes. Table A.2 shows a strong negative and statistically significant association between attitudes and vote share of UKIP in column 1 and support for Leave campaign of Brexit in column 2.

III.C. Motivating Facts

We establish four motivating facts that form the basis for our conceptual framework and subsequent empirical analyses.

Anti-Immigrant Sentiment in Pro-Immigrant CSPs.– Our conceptual framework requires the presence of individuals with anti-immigrant attitudes in immigrant friendly CSPs. We show in Figure 2 that in CSPs with above the median attitudes a non trivial share of individuals hold anti-immigrant attitudes, with nearly 15% holding the strongest that is possible.

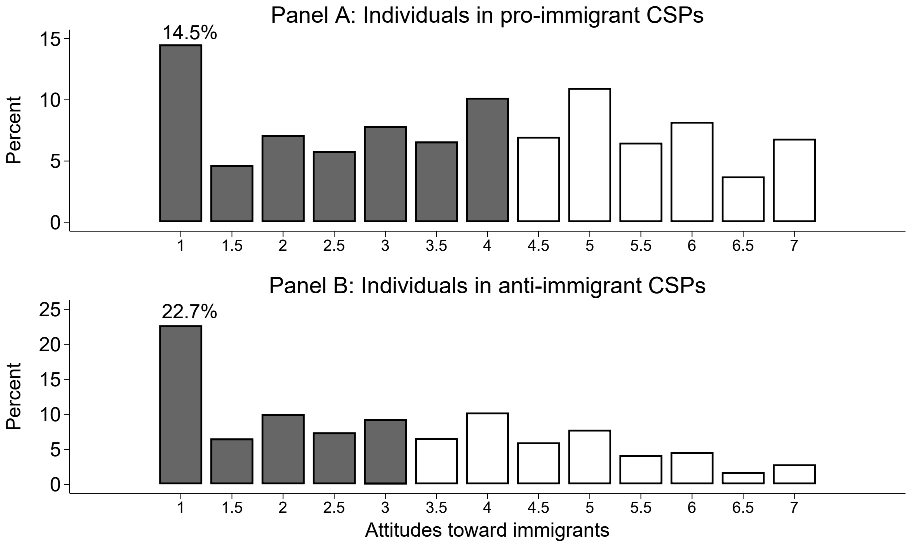

Figure 2: Distribution of Individual Attitudes Toward Immigrants by CSP Type

Notes. The figure reports the distribution of individual attitudes toward immigrants using data from Wave 1 of BES. Panel A shows the distribution of attitudes of individuals living in pro-immigrant CSPs (above-median average attitudes). Panel B shows the distribution of attitudes of individuals living in anti-immigrant CSPs (with below-median average attitudes). In each panel, shaded bars correspond to attitude values below or equal to the median of the distribution, while unshaded bars correspond to values above the panel-specific median.

Awareness of Local Sentiment toward Immigrants.– Our conceptual framework relies on people using local sentiment toward immigrants to form beliefs about nationwide sentiment. This assumes that people are aware of sentiments toward immigrants in their CSP. We provide evidence in support of this argument using the context of Brexit, where support for remain in EU can be interpreted as reflecting pro-immigrant attitude. We use data from Wave 8 of BES in which households were asked the question "Most people I know will vote to *remain* in the European Union." We compare the share of respondents reporting *yes* on this question (local belief) with *actual* remain vote share at the CSP level. If people are aware of the local sentiment toward immigrants, then the correlation between these two variables should be large and positive. Figure 3 shows that this is indeed the case. There is a strong positive association of 0.82 (s.e. = 0.03) between these two variables, which is statistically significant at the 1-percent level. This strong correlation suggests that people are aware of sentiments toward immigrants in their CSP.

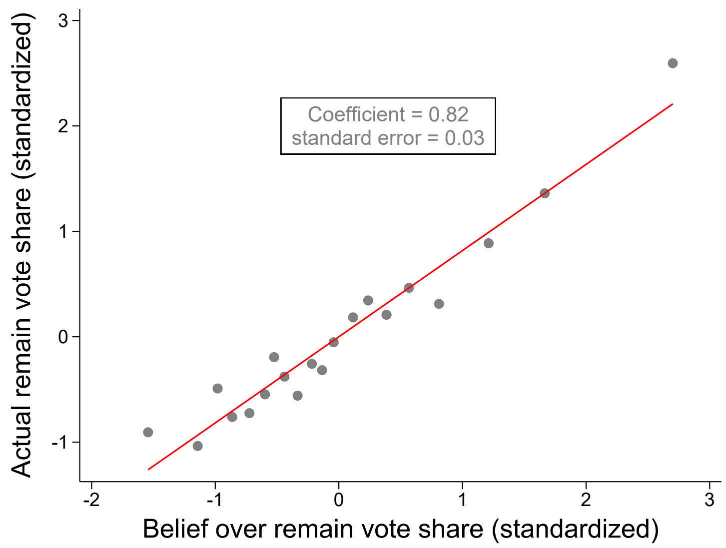

Figure 3: Awareness of Local Anti-Immigrant Sentiment at Local Level

Notes. The figure shows a CSP level bin-scatter plot on the association between beliefs over remain vote share and actual remain vote share in Brexit referendum. All variables are standardized to have mean 0 and standard deviation 1. Data on beliefs is from Wave 8 of BES. Data on actual remain vote share is from UK Electoral Commission (2019).

Attitudes and Belief Shocks.– We argue that when national events reveal widespread anti-immigrant sentiment, individuals residing in pro-immigrant areas experience larger belief shocks due to heterogeneous priors. Evidence in support of this argument warrants data on prior beliefs about the likelihood of the event happening. The BES offers data on prior beliefs about the likelihood of Brexit success, but data on the likelihood of UKIP victory were unfortunately collected ex post. However, as the study by Kuran (1991) points out, this can still yield meaningful insights.

For the Brexit event, we use data from Wave 8 of BES. In the survey, individuals were asked to rate the likelihood that Brexit will happen on a scale of 0-100, where 0 implies that "UK will definitely vote to remain" and 100 implies "UK will definitely vote to leave." Since the Leave campaign won the Brexit referendum, we measure belief shock as 100 minus the perceived likelihood of the Leave campaign winning. The average belief shock at the CSP level is 48.7% (3.4 s.d.).

For the UKIP election, we use responses in Waves 4 and 5, which were conducted between 4 March 2015 and 6 May 2015, 10-12 months after the UKIP election. In the survey, individuals were asked the question "Which of these parties do you think has [NO] real chance of being part of the next UK government?" UKIP was one of the listed parties. Respondents chose between 0 indicating "No" and 1 indicating "Yes". We measure belief shock as 1 minus the perceived chance of UKIP winning (Yes). The average belief shock from UKIP at CSP level is 36.5% (6.6 s.d.).

We compare belief shocks among people with *anti-immigrant* attitudes residing in immigrant friendly CSPs to their counterparts in anti-immigrant CSPs using the following equation:

$$D_1 = \mathbb{E}[\text{Belief Shock}_i \mid \text{Anti-Immigrant}_i = 1,\ \text{CSP}_i = \textbf{Immigrant-Friendly}] - \mathbb{E}[\text{Belief Shock}_i \mid \text{Anti-Immigrant}_i = 1,\ \text{CSP}_i = \textbf{Anti-Immigrant}] \tag{11}$$

We carry out a similar exercise for people with *pro-immigrant* attitudes residing in immigrant friendly CSPs to their counterparts in anti-immigrant CSPs using the following equation:

$$D_2 = \mathbb{E}[\text{Belief Shock}_i \mid \text{Pro-Immigrant}_i = 1,\ \text{CSP}_i = \textbf{Immigrant-Friendly}] - \mathbb{E}[\text{Belief Shock}_i \mid \text{Pro-Immigrant}_i = 1,\ \text{CSP}_i = \textbf{Anti-Immigrant}] \tag{12}$$

We define anti and pro-immigrant individuals / CSPs using a median split, whereby those below the median are classified as anti-immigrant and those above the median as pro-immigrant. Figure 4 plots D_1 and D_2 to show that people with anti-immigrant and pro-immigrant attitudes residing in *immigrant friendly* CSPs experience a significantly larger belief shock by 0.09-0.12 standard deviations than their counterparts residing in anti-immigrant CSPs.[18]

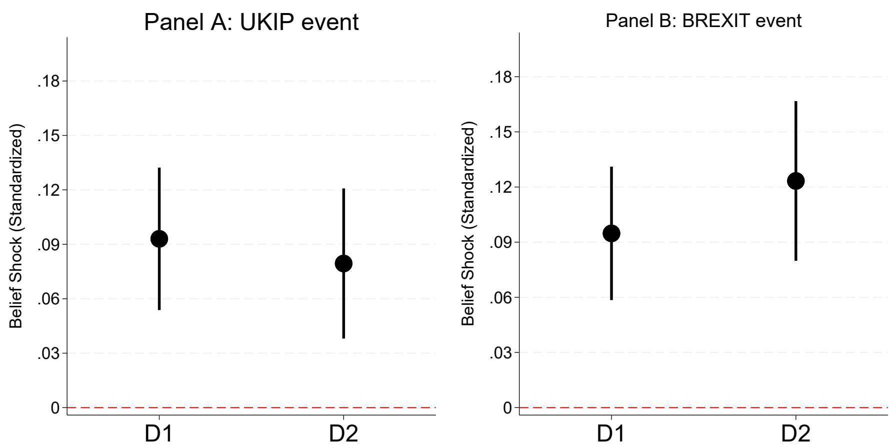

Figure 4: Belief Shock by Individual Attitudes and CSP Type

Notes. The figure reports D_1 and D_2 from equation (11) and equation (12) together with 95% confidence intervals separately for UKIP and Brexit events. Belief shock is standardized to have mean 0 and standard deviation 1. Data are from Wave 4-5 (Panel A) and Wave 8 (Panel B) of BES.

National Identity.– We argue that individuals care about complying with national norms because of national identity considerations. This matters in a context where national sentiments are important. To unpack this, we use data from a question in Wave 1 of BES on Britishness. In the survey, the respondents were asked to rate on a seven-point scale: "How strongly do you feel British?" The possible options ranged from "Not at all

[18]The two coefficients are not significantly different from each other either within UKIP (p-value=0.65) or within Brexit (p-value=0.28) events

strongly British" (point 1) to "Very strongly British" (point 7). Figure 5 shows a very strong sense of national identity among the respondents. Both the mean and median are close to 6 and there is not much variation.[19]

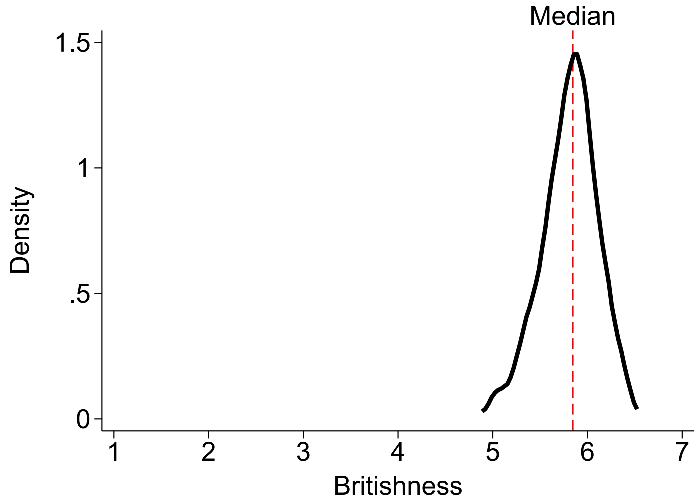

Figure 5: Distribution of British National Identity

Notes. The figure plots the kernel density of responses to the question on Britishness, whereby 1 implies weak and 7 implies strong. The vertical dashed red line indicates the sample median (5.84). Data are from Wave 1 of the BES.

IV. Empirical Strategy

We use difference-in-differences strategy to examine whether CSPs with pro-immigrant attitudes experience larger surges in hate crime than CSPs with anti-immigrant attitudes in post-event periods. In our setting, every unit is treated at the same time, but there is a heterogeneous response depending on attitudes (see for instance Korovkin and Makarin, 2023). For each event, we separately estimate the following equation:

$$hate_{it} = \alpha + \beta(Attitudes_i \times Post\text{-}Event_t) + \gamma \mathbf{X}_{it} + \tau_t + \eta_i + \epsilon_{it} \tag{13}$$

where $hate_{it}$ is hate crime in CSP i in quarter t. *Attitudes* are measured using survey responses in BES. *Event* refers to either UKIP election to the European Parliament or Brexit referendum. For the UKIP election, we use *Post-UKIP* – a binary indicator that takes a value of 1 for quarters from May 2014 onward, otherwise 0. For the Brexit referendum, we use *Post-Brexit* – a binary indicator that takes a value of 1 for quarters from June 2016 onward, otherwise 0. τ_t is a fixed effect for quarter. It absorbs changes that affect all CSPs equally in that quarter, such as macroeconomic conditions in England

[19] We choose British because our sample includes CSPs from both England and Wales. However, if we consider English identity, which was also elicited in the same survey, the mean and median for England are also close to 6.

and Wales. Crucially, it also absorbs seasonal variations in hate crime, which could result in higher hate crimes in summer than in winter. η_i is the fixed effect for CSP, which absorbs time-invariant differences between CSPs including slow moving factors like large stable differences in stock of population and immigrants. ϵ_{it} is the idiosyncratic error term. We cluster standard errors by CSP. While conducting robustness checks, we show that our results hold when we account for spatial correlation of errors. $\mathbf{X}$ is a vector of CSP level covariates that vary over time. Hate crime may depend on changes in the size and composition of the population in a CSP. Accordingly, we control for population size (yearly), flow of EU migrants (quarterly), and flow of non-EU migrants (quarterly). Hate crimes may also stem from economic, social, and cultural environment in CSPs, as these could have given rise to support for the political events under consideration. Accordingly, we control for CSP level index of education deprivation, austerity, generalized trust, and trust in politicians. Panels B of Table A.1 offers summary statistics on control variables.

The coefficient of interest is β, which captures the association of attitudes with hate crime in the Post-Event periods. In line with our conceptual framework, we expect β to have a positive sign, that is, steeper rise in hate crime in CSPs with more pro-immigrant attitudes. We estimate (13) separately for both the events. For the UKIP event, we restrict the sample to periods before Brexit, that is, from Q2 2002 to Q1 2016. For the Brexit event, we use the entire sample. We might under or over estimate β in the Post-Brexit period if we fail to account for the surge in hate crimes in the pre-Brexit period caused by UKIP event. So, we also present results in which we include at the same time the interaction of attitudes both with UKIP, as well as Brexit events.

Note that our focus is on changes in the levels of hate crime in the post-event periods rather than on proportional increases. So, we use number of hate crimes as the preferred dependent variable. This is also of interest from the policy point of view, as a larger increase in the number of hate crimes warrants more attention from the police and policy makers rather than larger proportional increases which could be reflecting very small numerical changes. That being said, while conducting robustness checks we also consider specifications with log of hate crime. Following Roth and Sant'Anna (2023) we also conduct a test and find that we cannot reject the null hypothesis that parallel trends are insensitive to the functional form.

We are worried about three major concerns. First, the identifying assumption is that in the absence of the event, the evolution of hate crimes will be similar across CSPs with different attitudes. We present event studies to test for pre-trends. Second, it is known that economic and social conditions in the pre-event period led to stronger support for UKIP election and Brexit success. Thus, it is plausible that we are capturing the effect of these factors rather than the effect of information shocks from the events. So, we consider several proxies of these variables. Third, our strategy assumes actual increase in hate crimes in the post-event periods, but we might be capturing changes in the reporting behavior of victims in the post-event periods. To mitigate this concern, we start by accounting for police specific factors affecting recording of crime. In addition, we use data from the Crime Survey for England and Wales (CSEW) to track changes in reporting behavior in response to the events. We also test if our results change when we adjust our measure of hate crime for under-reporting. We discuss each of these three strategies in detail below.

IV.A. Event Study

We examine the evolution of association between attitudes and hate crime before and after the events using the following equation:

$$\text{Hate}_{it} = \sum_{k=2002Q2}^{2019Q4} \beta_k \cdot T_{it}^k \times \text{Attitudes}_i + \gamma \mathbf{X}_{it} + \tau_t + \eta_i + \epsilon_{it} \tag{14}$$

T_{it}^k are indicators for the event-time quarters k. $k = 0$ is Q2 2014 in which the European Parliament elections were held. The omitted category ($k = -1$) is Q1 2014. The remaining terms are the same as in Equation (13). For UKIP, we use the time window of 48 quarters before the event (starting from Q2 2002) and eight quarters afterward (up to Q1 2016). For Brexit, we consider fourteen additional quarters (from Q2 2016-Q4 2019). Figure 6 presents the results.

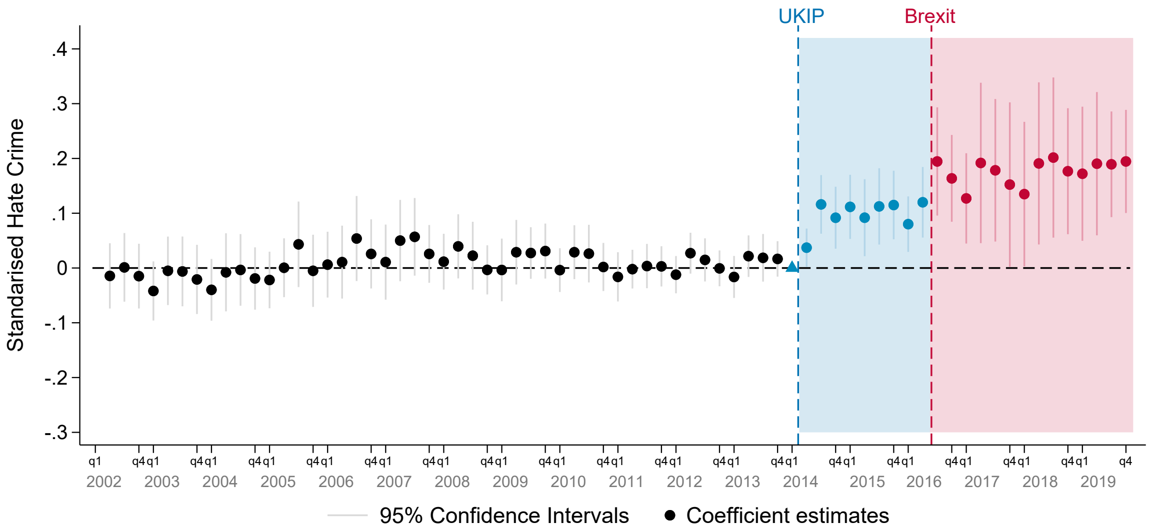

Figure 6: Event Study

Notes: The figure plots coefficients on $T_{it}^k \times \textit{Attitudes}$ from the estimation of Equation 14 with 95% confidence bands. The Y-axis denotes standardized measure of hate crime. Black markers correspond to pre-UKIP period, blue markers to post-UKIP event, and red markers to post-Brexit event. The blue triangle corresponds to omitted period of Q1 2014, the blue vertical line to UKIP event, and the red vertical line to Brexit event. On the X-axis, q1 and q4 refer to quarter 1 and 4. Standard errors are clustered at the CSP level.

Prior to the UKIP event, the trend in hate crime is similar across CSPs. However, in the post-UKIP period, there is a steep rise in hate crime in areas with pro-immigrant attitudes (see solid blue circles). The effect peaks immediately after the Brexit referendum in Q3 2016 and remains high on average in the Post-Brexit period (see red solid circles). The coefficients are individually and jointly statistically significant at the 1-percent level. Figure A.2 shows very similar patterns when we use hate crime per capita.

IV.B. Economic and Social Conditions in Pre-Event Period

Economic Conditions.– Guriev and Papaioannou (2022), Becker, Fetzer and Novy (2017), and Fetzer (2019) show that poor economic conditions in the form of education deprivation and austerity were responsible for the rise in UKIP popularity and support for Brexit. Therefore, one concern could be that the surge in hate crime in the post-event period is not because of information shocks from UKIP or Brexit events but is actually due to the persistent effect of economic conditions from the pre-event periods. However, an empirical regularity observed across many countries is that areas with poor economic conditions tend to exhibit anti-immigrant rather than pro-immigrant attitudes (see for example, De Bromhead, Eichengreen and O'Rourke, 2013; Algan et al., 2017). This implies that, if anything, CSPs with anti-immigrant rather than pro-immigrant attitudes would be expected to experience larger hate crime surges in the post-event periods. Our estimates are therefore plausibly unlikely to be confounded by pre-existing economic conditions.

Nevertheless, we consider three strategies to address this concern. First, to the extent that changes in economic conditions occur at the regional level, we account for such changes by including region-specific time trends. Second, austerity was a result of the UK welfare reforms starting 2010. So, we control for an interaction between post-2009 indicator and austerity, measured using data on financial loss per working-age adult from Beatty et al. (2013). Third, we also control for an interaction of education deprivation with a post-event indicator. We measure education deprivation in 2014-15 using the Education, Skills and Training domain of the English Indices of Deprivation (Ministry of Housing, Communities & Local Government, 2015).[20]

A related concern could be that the surge in hate crime is the result of poor economic conditions arising from the events themselves. Notice, however, that the patterns in Figure 1 and Figure 6 clearly show an immediate rise in hate crime in the post-event periods. This concern requires economic outcomes to change immediately after the events. However, evidence suggests that changes in economic outcomes occurred with a lag of several years and especially after Brexit was actually implemented. In fact Bakker et al. (2022) show that consumer prices remained unchanged right after the Brexit referendum in 2016.

[20]Our results remain unchanged when we use education deprivation from 2010-11.

Social and Cultural Conditions.– Guriev and Papaioannou (2022) argue that several social and cultural factors gave rise to surges in populism. These factors include generalized trust in strangers and trust in politicians. We capture these variables through Wave 1 and 7 of BES. For generalized trust, we use the question "Generally speaking, would you say that most people can be trusted, or that you can't be too careful in dealing with people?" We compute the share of respondents in a CSP who choose "most people can be trusted" to measure generalized trust. For trust in government, participants are asked "How much trust do you have in Members of Parliament in general?". They could choose answer on a scale of 1-7, where 1 implies "no trust" and 7 implies a "great deal of trust." We include an interaction of post-event indicator with trust in strangers, as well as with trust in government.

IV.C. Recording and Reporting Hate Crime

Recording Hate Crime.– Another potential concern could be changes in the recording of hate crimes in CSPs. This poses a concern to the extent that it disproportionately affected CSPs with pro-immigrant attitudes. We consider two strategies to deal with this concern. First, recording practices may have changed during the pre-event periods, but the effect persists in the post-event periods. To address this concern, we control for police force-specific time trends.

Second, police recording practices may have changed in the post-event period, especially due to increased sensitivity among the police or directives to record hate crimes more accurately. To address this concern, we control for an interaction between police force and post-event indicator. This allows us to account for changes in recording that may have resulted from the events themselves.

Reporting Hate Crime.– A related concern is that the frequency of hate crimes has remained the same, but reporting has increased in the post-event periods. Addressing this concern fully is very difficult, so we attempt to limit its scope. We use survey data from Crime Survey of England and Wales (CSEW) in which respondents are asked: a) whether they have been victim of a crime in the past year; b) whether the crime has been reported to the police.[21] CSEW uses these data to compute the *propensity to report* different types of crimes. We focus only on those crimes that also have a hate counterpart. Table A.3 lists the average reporting rates over the study period by different types of crimes with hate counterparts, such as wounding, assault with and without injury, arson, and vehicle damage. Note that these data are not available at the CSP level but are aggregated at England and Wales level. Nevertheless, they help us shed some light on the scope of reporting bias in three ways.

[21] Available online at Hate crime, England and Wales, 2019 to 2020, published 13$^{\text{th}}$ October 2020 and also at Crime in England and Wales: Annual Trend and Demographic Tables, published 22$^{\text{nd}}$ July 2021.

First, we test in Figure 7 whether reporting rates have changed in response to the events. We find that the average reporting rate is flat over time and there is no difference in pre-and post-event periods.

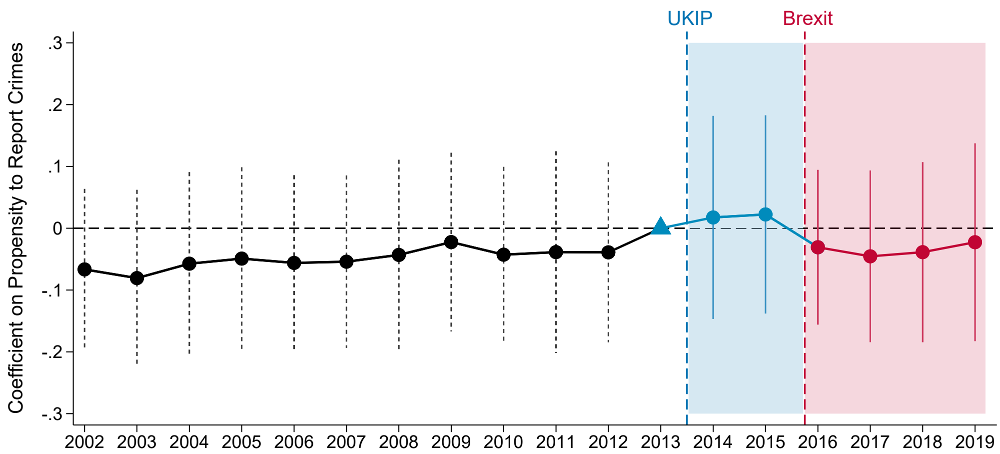

Figure 7: Propensity to Report Crimes over Time

Notes. The figure shows an event-study of the average propensity to report crimes over time. The Y-axis reports estimated coefficients from a regression of the reporting rate on calendar-year indicators, relative to the omitted UKIP baseline year (2013). The X-axis reports calendar years. Shaded areas indicate the post-UKIP (blue) and post-Brexit (red) periods. The solid blue vertical line corresponds to the post-UKIP period, while the solid red line corresponds to the post-Brexit period. Confidence intervals are at 95% level. Data are from CSEW.

Second, we construct an adjusted measure of hate crime following a methodology in the spirit of Soares (2004), exploiting variation in the composition of types of hate crime across CSPs. As an example, in some CSPs hate crimes are mostly due to wounding in a given year, whereas in others assaults are more common in that year. For each crime type, we use crime specific reporting rates from the relevant time period to construct an adjusted measure. For instance, if the reporting rate of wounding crime is 50 percent in time t, we adjust hate crimes related to wounding in each CSP by factor of 2 in that time period. We then test if our results hold when we use this adjusted measure as our dependent variable.

Third, we also run our main specification separately for different types of crimes that differ in their reporting rates, from 30 percent to 60 percent. The idea is that if our results are impacted by systematic under-reporting of hate crime then this should affect especially those with severe under-reporting. An added advantage of this exercise is that it allows us to rule out competing mechanisms like rise in polarization in which partisan group members engage in hate crimes toward each other in the form of racial and religious abuse.

IV.D. Falsification Test using Other Crimes

The information shocks from the UKIP and Brexit events are unlikely to affect crimes such as murder and burglary. However, if hate crime changes are due to changes in economic conditions as well as reporting practices then we should see a surge in all kinds of crimes. Accordingly, we construct falsification tests using data on these crimes to confirm that our results are not spurious. We do not expect substitution between hate crime and other crimes unless murder and burglary were specifically targeted towards immigrants, which seems unlikely. Nevertheless, while conducting robustness checks we additionally control for these crimes.

V. Results

V.A. Main Result

We present our main results from the estimation of equation 11 in Table 1 separately for the UKIP event in Panel A and Brexit event in Panel B. Column 1 includes only fixed effects for CSPs and quarter. We find that one standard deviation increase in attitudes toward immigrants (0.458) is associated with a rise in hate crime in post-UKIP period by 0.17 SD and in the post-Brexit period by 0.16 SD. These estimates are not only economically meaningful but are also statistically significant at the 1-percent level. In terms of raw differences, it amounts to an increase in hate crime by 6 incidents, which is 20 percent of the mean (31 incidents).

In column 2, we include CSP level controls like population, number of EU migrants, number of non-EU migrants, and region-specific time trends. The coefficient on *Attitudes* $\times$ *Post-UKIP* declines in magnitude to 0.10 SD whereas that on *Attitudes* $\times$ *Post-Brexit* to 0.13 SD, but both remain statistically significant at the 1-percent level. In column 3, we introduce police force-specific time trends to account for possible changes in crime recording behavior of the police. In both panels, the coefficients increases slightly in magnitude and their standard error declines. Finally, in column 4, we control for an interaction between police force and post-event indicator to account for possible changes in crime recording behavior that may have been induced by the event itself. The coefficients increase further in magnitude and are now comparable to those reported in column 1.

We proceed by including economic variables in Table 2. We consider an interaction of post-event indicator with an index of education deprivation in column 1 and an interaction of post-2009 indicator with austerity in column 2. Next, we include an interaction of post-event indicator with generalized trust in column 3 and with trust in politicians in column 4. We find that education deprivation is consistently positively and significantly associated with hate crime in the post-event periods. Nevertheless, the introduction of these controls does not lead to any major changes in the coefficient on *Attitudes* $\times$ *Post-*

UKIP or *Attitudes* × *Post-Brexit*, which remain robust in both magnitude and significance throughout.

Table 1: Information Shock, Attitudes, and Hate Crime

	Dependent Variable: Standardized Hate Crime			
	Quarter & CSP Fixed Effects (1)	Controls & Region × Time Trend (2)	Controls & Police × Time Trend (3)	Controls & Police × Post-Event (4)
	Panel A: UKIP Event			
Attitudes × Post-UKIP	0.169	0.101	0.113	0.120
	(0.024)	(0.022)	(0.021)	(0.020)
R^2	0.89	0.91	0.93	0.89
Observations	17,024	17,024	16,912	17,024
	Panel B: Brexit Event			
Attitudes × Post-Brexit	0.160	0.132	0.164	0.169
	(0.034)	(0.038)	(0.038)	(0.036)
R^2	0.86	0.89	0.91	0.88
Observations	21,577	21,577	21,435	21,577
Number of CSPs	304	304	302	304
CSP Fixed Effects	Yes	Yes	Yes	Yes
Quarter Fixed Effects	Yes	Yes	Yes	Yes
Controls	No	Yes	Yes	Yes
Region × Quarter	No	Yes	No	No
Police × Quarter	No	No	Yes	No
Police × Post-Event	No	No	No	Yes

Notes: OLS estimates with standard errors clustered at the CSP level. Control variables include population, number of EU migrants, and number of non-EU migrants. Attitudes is toward immigrants and is standardized to have mean of zero and standard deviation of 1. Panel A shows results using information shock from the UKIP event. Panel B shows results using information shock from the Brexit event. All specifications incorporate quarter and CSP fixed effects. Panel A covers a sample period from 2002 Q2 to 2016 Q1, whereas Panel B covers a sample period from 2002 Q2 to 2019 Q4. The sample size is larger in Panel B because it includes fifteen additional quarters. Column (3) has fewer observations because two CSPs have singleton observations.

V.B. Robustness

We implement a number of robustness checks using the specification with full set of controls from column 4 of Table 2.

Simultaneously Considering UKIP and Brexit Events.– We test whether our results hold when we simultaneously consider interaction of attitudes with Post-UKIP and with Post-Brexit. Table A.4 reports the results. We find that the coefficient on *Attitudes* × *Post-UKIP* and that on *Attitudes* × *Post-Brexit* are 0.10; both are statistically significant at the 1 percent level.

Table 2: Information Shock, Attitudes, and Hate Crime:
Controlling for Economic and Trust Variables

	Dependent Variable: Standardized Hate Crime			
	Economic Variables		Trust in	
	Education	Austerity	Strangers	Politicians
	(1)	(2)	(3)	(4)
	Panel A: UKIP Event			
Attitudes × Post-UKIP	0.167	0.170	0.171	0.170
	(0.027)	(0.028)	(0.028)	(0.029)
Education × Post-UKIP	0.041	0.066	0.064	0.065
	(0.023)	(0.025)	(0.025)	(0.024)
Austerity × Post-2009		-0.054	-0.054	-0.053
		(0.016)	(0.016)	(0.016)
Generalized Trust × Post-UKIP			-0.037	-0.036
			(0.059)	(0.059)
Trust Politicians × Post-UKIP				0.023
				(0.108)
R^2	0.89	0.89	0.89	0.89
Observations	17,024	17,024	17,024	17,024
	Panel B: Brexit Event			
Attitudes × Post-Brexit	0.188	0.190	0.191	0.190
	(0.038)	(0.037)	(0.038)	(0.038)
Education × Post-Brexit	0.181	0.187	0.185	0.186
	(0.040)	(0.043)	(0.044)	(0.044)
Austerity × Post-2009		-0.016	-0.016	-0.016
		(0.023)	(0.022)	(0.023)
Generalized Trust × Post-Brexit			-0.045	-0.044
			(0.089)	(0.087)
Trust Politicians × Post-Brexit				0.024
				(0.168)
R^2	0.87	0.87	0.87	0.87
Observations	21,577	21,577	21,577	21,577
Number of CSPs	304	304	304	304
CSP Fixed Effects	Yes	Yes	Yes	Yes
Quarter Fixed Effects	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes

Notes: OLS estimates with standard errors clustered at the CSP level. All columns use the baseline specification from column 1 of Table 1, adding standard controls: population, EU migrants, and non-EU migrants. Column (1) controls for an interaction between post-event indicator and education deprivation; column (2) includes an interaction between post-2009 indicator and austerity. Column (3) includes generalized trust × post-event; column (4) includes trust in Members of Parliament × post-event. All variables are standardized to have mean zero and unit standard deviation to allow for comparison of coefficients.

Adjusted Hate Crime by Reporting Rates.– We conduct two checks. First, we show in A.5 that our results hold when we compute an adjusted measure of hate crime that takes into consideration reporting rates of different types of crime. The coefficient on *Attitudes* × *Post-UKIP* turns out to be smaller in magnitude (0.09) than when we use the unadjusted measure, but it stays statistically significant at the 1-percent level. In contrast, the coefficient on *Attitudes* × *Post-Brexit* is similar in magnitude (0.22) and significance to the unadjusted measure.

We also run our main specification separately for different types of hate crimes that differ in their reporting rates. We report the results in Figure 8 where we plot the coefficient on *Attitudes* × *Post-UKIP* in Panel A and *Attitudes* × *Post-Brexit* in Panel B for different hate crime types listed on the x-axis. We find that the coefficients are comparable in magnitude to each other and their confidence intervals overlap. Together, these results suggest under-reporting alone cannot explain our results. In addition, they show that polarization alone cannot explain our results as the effect persists across different types of crimes.

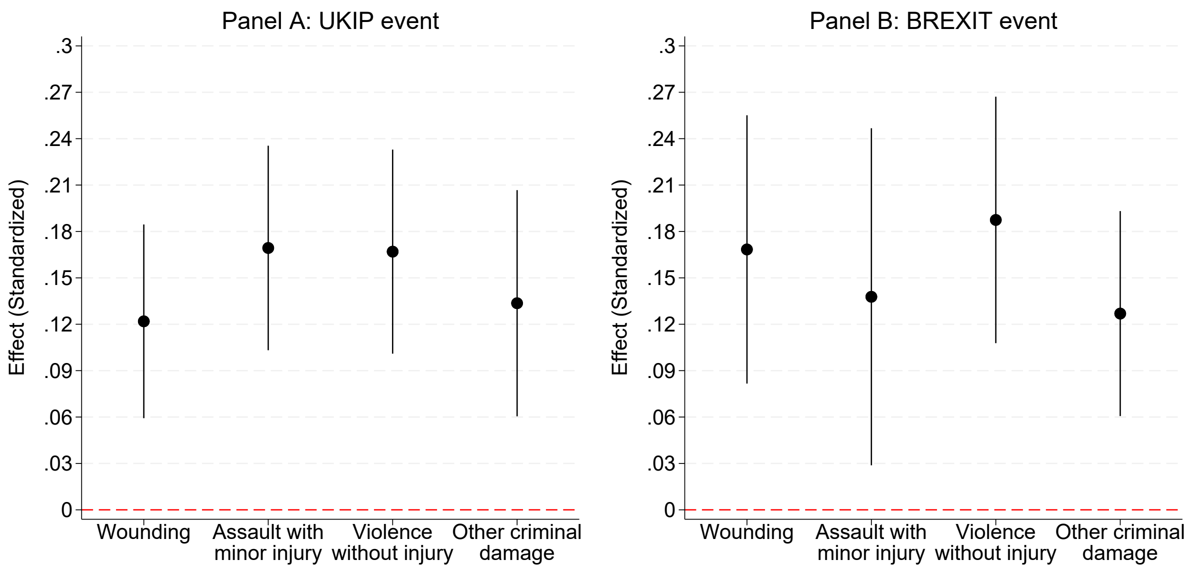

Figure 8: Attitudes toward Immigrants and Type of Hate Crime

Notes. The figure reports coefficient estimate with 95% confidence intervals on the interaction of attitudes with a post-event indicator. The coefficients are obtained from a separate regression of each hate crime type on the interaction term after controlling for population size, EU migrants, non-EU migrants, interaction between austerity and post-2009 indicator, as well as separate interaction of post-event indicator with education deprivation, generalized trust, and trust in politicians. Panel A corresponds to the UKIP event and Panel B to the Brexit event. Hate crimes are grouped into four categories: wounding; assault with minor injury; violence without injury; and other criminal damage.

Spatial Correlation of Standard Errors.– Column 1 of Table A.6 shows that our results hold when we consider spatial correlation of standard errors using 50 km distance as cutoffs. The coefficients on both *Attitudes* × *Post-UKIP* and *Attitudes* × *Post-Brexit* remain statistically significant at the 1-percent level.

Including Grievous and Aggravated Bodily Harm.– Column 2 of Table A.6 shows that our results do not change much in magnitude and significance when we include these types of hate crimes, which make up a very small percentage of hate crimes.

Terrorist Attacks.– There could be a concern that our results are confounded with terrorist attacks that occurred in the second quarter of 2017 in London and Manchester, which were perpetrated by immigrants and individuals of foreign descent. If these attacks led to a surge in hate crime in CSPs with more positive attitudes towards immigrants, this could result in overestimation of the effect of information shock from Brexit event. Column 3 of Table A.6 shows that our results hold when we control for an interaction between attitudes and post-indicator for Q2 2017 to account for changes in hate crime that are possibly due to these terrorist attacks.

Hate Crime per Capita.– Column 4 of Table A.6 shows that our results hold when use hate crime per-capita (normalized by population).

Log of Hate Crime per Capita.– Column 5 of Table A.6 shows that our results hold when we use log of hate crime.

Using One Attitude at a Time.– So far, our results are based on a measure of attitudes that take both economic and cultural importance of immigrants. We show in columns 1-2 of Table A.7 that our results are robust to using only the economic or the cultural dimension.

Alternative Wave to Measure Attitudes.– We test whether our results also hold when we use responses from Wave 7 to measure attitudes for our Post-Brexit analysis. This wave was conducted two months before Brexit. Column 3 of Table A.7 reports the results. The coefficients on attitudes in the post-event periods are similar in magnitude and statistical significance to our main findings.

Falsification Test using Other Crimes.– We provide further evidence in support of our results by conducting a falsification test using data on other crimes: murder and burglary. These crimes are unlikely to be affected by information shocks from UKIP and Brexit events, but may evolve in response to other confounding changes, such as those related to recording of crime or deterioration in the economic environment. Table A.8 shows the results. In both panels, the effect of attitudes on other crimes in the post-event period is mostly very small in magnitude and is also statistically insignificant. We further show in Table A.9 that our main results are robust to controlling for other crimes.

VI. Mechanism

We argue that national events like UKIP election and Brexit referendum revealed new information on previously unknown country-wide support for anti-immigrant sentiment. This generated a belief shock for individuals residing in pro-immigrant CSPs. This shock was particularly instrumental for anti-immigrant individuals residing in pro-immigrant CSPs who now realized that their attitudes align with that of the country, felt emboldened, and resorted to hate crime.

Earlier, we saw in section III. that pro-immigrant and anti-immigrant individuals residing in immigrant friendly CSPs are the ones who experience a larger belief shock. We now proceed by showing that these belief shocks map onto hate crimes. Ideally we would like to show this mapping for anti-immigrant individuals living in immigrant friendly CSPs. However, individual level data on crime are never publicly available. So we start by showing this mapping at the CSP level. Subsequently, we also establish this link at the individual level by conducting a proof-of-concept survey in which our outcome is not hate crimes but vocalization about immigrants.

VI.A. Belief Shocks and Hate Crime in Post-Event Periods

We examine the association between belief shocks and hate crimes in the post-event period. The event study in Figure 9 reveals a strong positive and statistically significant association between belief shocks and hate crimes in post-event periods.

Table 3 presents results from an econometric analysis. Column 1 includes only CSP and quarter fixed effects. It shows that a one SD increase in belief shock is associated with a rise in hate crimes by 0.09 SD in Post-UKIP period and by 0.11 SD in Post-Brexit period. These results hold when we include one after the other economic variables, trust variables, and all of them together in columns 2-5.

VI.B. Attitudes, Belief Shock, and Behavior

We designed an exploratory study to test for the relationship between attitude, beliefs, and tendency to engage in anti-immigrant behavior. The survey was run on platform called Prolific between February and March, 2021. Admittedly, the survey was carried out with a lag of almost 5 years in the post-event period, so our purpose here is to offer exploratory evidence. 1448 individuals from England and Wales participated in the survey. These individuals were selected on the bases of participation in the Brexit referendum regardless of whether they supported leave or remain. We ensured that the share of participants supporting leave and remain were aligned with the actual shares in Brexit referendum, that is, 52 percent who supported leave. In the survey, we asked respondents three key questions:

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

2. 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?

3. 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. Choose from: Much more likely; more likely; neither more nor less likely; less likely; much less likely.

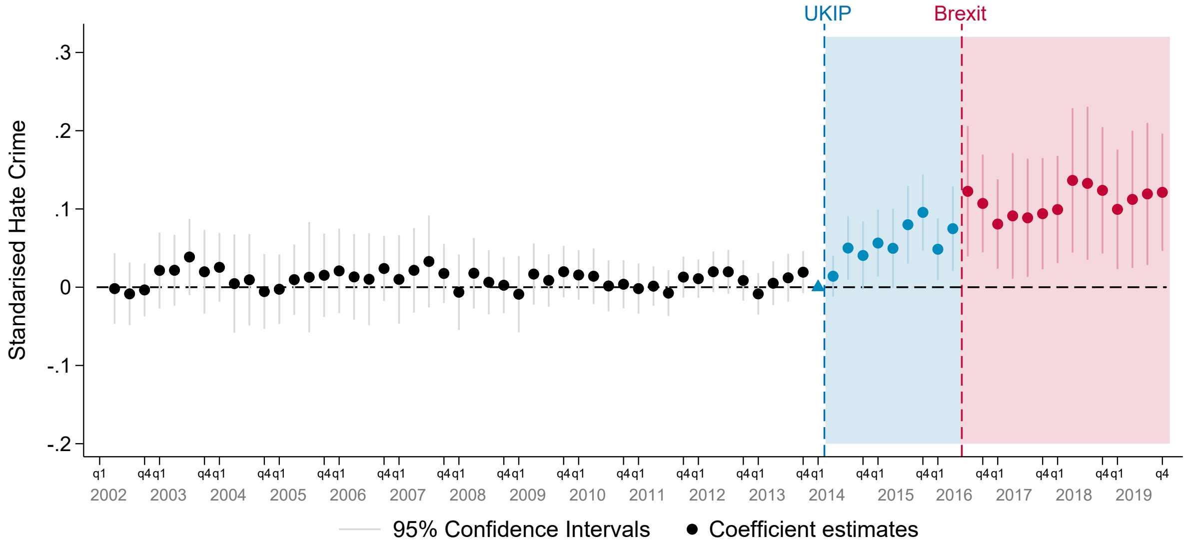

Figure 9: Event Study: Belief Shock and Hate Crime

Notes: The figure plots coefficients on T_{it}^{k} × *Belief shock* with 95% confidence bands, where T_{it}^{k} are event-time indicators that the event happened k quarters away from UKIP event. The Y-axis denotes standardized measure of hate crime. Black markers correspond to pre-UKIP period, blue markers to post-UKIP event, and red markers to post-Brexit event. The blue triangle corresponds to omitted period of Q1 2014, the blue vertical line to UKIP event, and the red vertical line to Brexit event. On the X-axis, q1 and q4 refer to quarter 1 and 4. Standard errors are clustered at the CSP level.

Table 3: Belief Shock and Hate Crime

	Dependent Variable: Standardized Hate Crime				
	Quarter &	Economic Variables		Trust in	
	CSP FE	Education	Austerity	Strangers	Politicians
	(1)	(2)	(3)	(4)	(5)
	Panel A: UKIP Event				
Belief Shock × Post-UKIP	0.094	0.076	0.079	0.080	0.075
	(0.019)	(0.017)	(0.017)	(0.016)	(0.016)
Education × Post-UKIP		-0.014	0.009	0.008	0.025
		(0.020)	(0.023)	(0.024)	(0.022)
Austerity × Post-2009			-0.052	-0.052	-0.050
			(0.017)	(0.016)	(0.017)
Generalized Trust × Post-UKIP				-0.021	-0.015
				(0.060)	(0.059)
Trust Politicians × Post-UKIP					0.290
					(0.115)
R^2	0.88	0.89	0.89	0.89	0.89
Observations	17,024	17,024	17,024	17,024	17,024
	Panel B: Brexit Event				
Belief Shock × Post-Brexit	0.111	0.144	0.144	0.145	0.138
	(0.032)	(0.031)	(0.031)	(0.032)	(0.031)
Education × Post-Brexit		0.161	0.165	0.164	0.178
		(0.039)	(0.043)	(0.045)	(0.044)
Austerity × Post-2009			-0.012	-0.012	-0.010
			(0.024)	(0.024)	(0.024)
Generalized Trust × Post-Brexit				-0.027	-0.022
				(0.092)	(0.089)
Trust Politicians × Post-Brexit					0.278
					(0.184)
R^2	0.85	0.87	0.87	0.87	0.87
Observations	21,577	21,577	21,577	21,577	21,577
Number of CSPs	304	304	304	304	304
CSP Fixed Effects	Yes	Yes	Yes	Yes	Yes
Quarter Fixed Effects	Yes	Yes	Yes	Yes	Yes
Controls	No	Yes	Yes	Yes	Yes

Notes: OLS estimates with standard errors clustered at the CSP level. Column 1 uses the baseline specification from column 1 of Table 1 with CSP and quarter fixed effects. Column (2) adds education deprivation × post-event. column (3) includes austerity × post-2009. Column (4) includes generalized trust × post-event. Column (5) includes trust in politicians × post-event. All variables are standardized to have mean zero and unit standard deviation to allow for comparison of coefficients.

The first question elicits attitudes toward immigrants on a scale from 0 (very negative) to 10 (very positive). We invert the variable so that higher values correspond to anti-immigrant sentiment. The second question asks respondents their belief on how other people in their local area view immigrants on the same scale. We keep the scale as it is so that now higher values correspond to pro-immigrant views in one's area. Finally, since we cannot ask participants their inclination to engage in hate crime, we asked them if Brexit referendum caused them to become *more or less likely* to publicly express views on immigrants, on a scale from 1 (much less likely) to 5 (much more likely). Note that these views could be positive or negative; we did not emphasize either. In addition to these questions, the respondents also provided data on socio-economic and demographic characteristics.

We use these data to identify individuals for whom the propensity to publicly express views on immigrants changed most markedly in the post-Brexit period. For this purpose, we regress changes in public expression of views toward immigrants on individuals' attitudes, pro-immigrant sentiment in one's area, and an interaction between the two. In line with our conceptual framework, we hypothesize the interaction term to have a positive effect in the post-event period, that is, those with anti-immigrant attitudes in immigrant friendly areas are more likely to vocalize their concern in the post-event period. Table A.10 shows that the coefficient on the interaction term is indeed positive; it as a magnitude of 10 percent and is statistically significant at the 1-percent level. This result implies that individuals with anti-immigrant attitudes in pro-immigrant areas are more likely to increase their expression of views on immigrants after the Brexit referendum.

VII. Conclusions

In this paper, we study whether and how heterogeneous priors in the context of national information shocks give rise to surges in xenophobic behavior and hate crimes. Our study takes place in the context of anti-immigrant sentiment revealed by two political events in the UK: UKIP's 2014 election to the European Parliament and the 2016 Brexit referendum. Using a difference-in-differences strategy, we show a large surge in hate crimes in areas with pro-immigrant attitudes. This result is robust to including a variety of economic, social and cultural variables as controls, as well as accounting for differences in recording and reporting of hate crimes.

We conceptualize and substantiate empirically that these results arise because people have heterogeneous priors which are not randomly distributed but rooted in sentiments towards immigrants in one's own local community. This causes differences in belief shocks from the new information generated by the political events. Using data on the perceived likelihood of the events, we construct measures of belief shocks and show that these are positively associated with hate crimes in the post-event periods. A proof-of-concept survey reveals that anti-immigrant individuals from pro-immigrant areas are much more likely to voice their concerns in the post-event periods.

These findings challenge conventional narratives, highlighting how populist shocks erode pluralistic ignorance not in places where xenophobia is entrenched but also in liberal areas, potentially amplifying societal divisions. This has important implications for policymakers aiming to foster social cohesion and address hate crimes.

References

Adena, Maja, Ruben Enikolopov, Maria Petrova, Veronica Santarosa, and Ekaterina Zhuravskaya. 2015. "Radio and the Rise of the Nazis in Prewar Germany." *Quarterly Journal of Economics*, 130(4): 1885–1939.

Adriani, Fabrizio, and Silvia Sonderegger. 2015. "Trust, trustworthiness and the consensus effect: An evolutionary approach." *European Economic Review*, 77: 102–116.

Adriani, Fabrizio, and Silvia Sonderegger. 2019. "A theory of esteem based peer pressure." *Games and Economic Behavior*, 115: 314–335.

Akerlof, George A. 1980. "A theory of social custom, of which unemployment may be one consequence." *Quarterly Journal of Economics*, 94(4): 749–775.

Akerlof, George A, and Rachel E Kranton. 2000. "Economics and identity." *Quarterly Journal of Economics*, 115(3): 715–753.

Algan, Yann, Sergei Guriev, Elias Papaioannou, and Evgenia Passari. 2017. "The European trust crisis and the rise of populism." *Brookings Papers on Economic Activity*, 2017(2): 309–400.

Bakker, Jan, Nikhil Datta, Richard Davies, and Joshua De Lyon. 2022. "Non-tariff barriers and consumer prices: evidence from Brexit." Centre for Economic Performance, LSE.

Beatty, Christina, Tony Gore, Ryan Powell, and Elizabeth Sanderson. 2013. "The impact of welfare reform in Hampshire." Centre for regional economic and social research, Sheffield Hallam University.

Becker, S.O., T. Fetzer, and D Novy. 2017. "Who voted for Brexit? A comprehensive district-level analysis." *Economic Policy*, 32(92): 601–650.

Beknazar-Yuzbashev, George, Rafael Jiménez-Durán, Jesse McCrosky, and Mateusz Stalinski. 2025. "Toxic content and user engagement on social media: Evidence from a field experiment." CESifo Working Paper.

Bénabou, Roland, and Jean Tirole. 2011. "Identity, morals, and taboos: Beliefs as assets." *Quarterly Journal of Economics*, 126(2): 805–855.

Bernheim, B Douglas. 1994. "A Theory of Conformity." *Journal of Political Economy*, 102(5): 841–877.

Bicchieri, Cristina, Eugen Dimant, and Silvia Sonderegger. 2023. "It's not a lie if you believe the norm does not apply: Conditional norm-following and belief distortion." *Games and Economic Behavior*, 138: 321–354.

Bursztyn, Leonardo, Georgy Egorov, and Stefano Fiorin. 2020. "From extreme to mainstream: The erosion of social norms." *American Economic Review*, 110(11): 3522–3548.

Bursztyn, Leonardo, Georgy Egorov, Ruben Enikolopov, and Maria Petrova. 2019. "Social media and xenophobia: evidence from Russia." National Bureau of Economic Research.

Cantoni, Davide, David Y Yang, Noam Yuchtman, and Y Jane Zhang. 2019. "Are Protests Games of Strategic Complements or Substitutes? Experimental Evidence from Hong Kong's Democracy Movement." *The Quarterly Journal of Economics*, 1021: 1077.

De Bromhead, Alan, Barry Eichengreen, and Kevin H O'Rourke. 2013. "Political extremism in the 1920s and 1930s: Do German lessons generalize?" *The Journal of Economic History*, 73(2): 371–406.

Fages, Diego Marino, and Agustina Martínez-Pozo. 2026. "Hate in the Tropics: Political Leaders and the Social Acceptability of Online Hate Speech." University of Durham.

Fetzer, Thiemo. 2019. "Did austerity cause Brexit?" *American Economic Review*, 109(11): 3849–86.

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

Guriev, Sergei, and Elias Papaioannou. 2022. "The political economy of populism." *Journal of Economic Literature*, 60(3): 753–832.

Jiménez-Durán, Rafael. 2023. "The economics of content moderation: Theory and experimental evidence from hate speech on Twitter." George J. Stigler Center for the Study of the Economy & the State Working Paper 324.

Jiménez Durán, Rafael, Karsten Müller, and Carlo Schwarz. 2024. "The effect of content moderation on online and offline hate: Evidence from Germany's NetzDG." *Available at SSRN 4230296*.

Kalra, Aarushi. 2024. "Hate in the time of algorithms: Evidence from a large-scale experiment on online behavior." Working Paper.

Korovkin, Vasily, and Alexey Makarin. 2023. "Conflict and intergroup trade: Evidence from the 2014 Russia-Ukraine crisis." *American Economic Review*, 113(1): 34–70.

Kuran, Timur. 1991. "The East European revolution of 1989: is it surprising that we were surprised?" *American Economic Review*, 81(2): 121–125.

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

Michaeli, Moti, and Daniel Spiro. 2017. "From peer pressure to biased norms." *American Economic Journal: Microeconomics*, 9(1): 152–216.

Ministry of Housing, Communities & Local Government. 2015. "The English Indices of Deprivation 2015." UK Government.

Roth, Jonathan, and Pedro HC Sant'Anna. 2023. "When is parallel trends sensitive to functional form?" *Econometrica*, 91(2): 737–747.

Shayo, Moses. 2009. "A model of social identity with an application to political economy: Nation, class, and redistribution." *American Political Science Review*, 103(2): 147–174.

Soares, Rodrigo. 2004. "Development, crime and punishment: accounting for the international differences in crime rates." *Journal of Development Economics*, 73(1): 155–184.

Tabellini, Guido. 2010. "Culture and institutions: economic development in the regions of Europe." *Journal of the European Economic Association*, 8(4): 677–716.

Vanberg, Christoph. 2019. "A short note on the rationality of the false consensus effect." University of Heidelberg Working Paper 662.

Voigtländer, Nico, and Hans-Joachim Voth. 2012. "Persecution perpetuated: the medieval origins of anti-Semitic violence in Nazi Germany." *Quarterly Journal of Economics*, 127(3): 1339–1392.

ONLINE APPENDIX:
Information Shocks, Attitudes toward Immigrants, and Hate Crime

Jake Bradley ⓡ Facundo Albornoz ⓡ Silvia Sonderegger ⓡ
Jesús Rodríguez ⓡ Devesh Rustagi ⓡ

Appendix A Theory

Proof Proposition 1 As discussed in the text, standard Bayesian updating delivers,

$$E(\mu \mid P^d, \overline{\mu}) = \frac{\Theta}{1+\Theta} P^d + \frac{1}{1+\Theta} \overline{\mu}. \tag{15}$$

Consider a generic linear symmetric equilibrium where, for all i, $a_i = k\alpha_i + \gamma P^{d_i} + (1-k-\gamma)\,\overline{\mu}$ where d_i indicates i's area. Aggregating all individuals across all areas yields $\overline{a} = (k+\gamma)\,\mu + (1-k-\gamma)\,\overline{\mu}$ and aggregating all individuals within area d yields $\overline{a}^d = (k+\gamma)\,P^d + (1-k-\gamma)\,\overline{\mu}$. The optimal action for individual j living in area d is

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

Substituting for $E_j\left(\mu|P^d,\overline{\mu}\right)$ we obtain the following system

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

which, after substituting $k=\theta$ delivers $\gamma = \frac{\theta(1-\theta)(\Theta+1-\lambda)}{\theta+\lambda(1-\theta)+\theta\Theta}$ as described in proposition 1. QED

Proposition 2 *For sufficiently large σ, the difference in hate crime rates before and after the event is an increasing function of P^d.*

Proof of Proposition 2 Let f denote the standard normal density. A second-order Taylor expansion of $f(\tilde{a}_1^d)$ around $\tilde{a}_0^d$ gives

$$f(\tilde{a}_1^d) \approx f(\tilde{a}_0^d) + f'(\tilde{a}_0^d)\big(\tilde{a}_1^d - \tilde{a}_0^d\big) + \frac{1}{2} f''(\tilde{a}_0^d)\big(\tilde{a}_1^d - \tilde{a}_0^d\big)^2.$$

Rearranging,

$$f(\tilde{a}_0^d) - f(\tilde{a}_1^d) \approx -f'(\tilde{a}_0^d)\left(\tilde{a}_1^d - \tilde{a}_0^d\right) - \frac{1}{2}f''(\tilde{a}_0^d)\left(\tilde{a}_1^d - \tilde{a}_0^d\right)^2.$$

Using the identities $f'(x) = -xf(x)$ and $f''(x) = (x^2 - 1)f(x)$, we obtain

$$f(\tilde{a}_0^d) - f(\tilde{a}_1^d) \approx f(\tilde{a}_0^d)\left(\tilde{a}_1^d - \tilde{a}_0^d\right)\left[\tilde{a}_0^d - \frac{1}{2}\left((\tilde{a}_0^d)^2 - 1\right)\left(\tilde{a}_1^d - \tilde{a}_0^d\right)\right].$$

Substituting in (10) we get

$$\frac{f\left(\widetilde{a}_0^d\right)}{\theta\sqrt{\sigma}}\left[\widehat{\varkappa}(\theta + \gamma_1) + \gamma_0 - \gamma_1\right] \tag{16}$$

where $\widehat{\varkappa} \equiv \frac{\overline{a}_1^d - \overline{a}_0^d}{\theta^2\sigma}\left[\mathbf{a} + \overline{a}_0^d - \frac{1}{2}\left(\frac{\left(\mathbf{a}+\overline{a}_0^d\right)^2}{\theta^2\sigma} - 1\right)\left(\overline{a}_1^d - \overline{a}_0^d\right)\right]$. Recall that $\overline{a}_0^d$, $\overline{a}_1^d$, γ_1 and γ_0 are independent of σ. If $\sigma \to \infty$, $\widehat{\varkappa} \to 0$ and hence the expression in the square bracket in (16) is unambiguously positive (since $\gamma_0 > \gamma_1$). By continuity, this implies that there exists a value $\widetilde{\sigma}$ such that expression (16) is unambiguously positive when $\sigma > \widetilde{\sigma}$. QED

Appendix B Empirical Strategy and Results

Summary Statistics.— Table A.1 reports summary statistics on hate crime, attitudes toward immigrants, and control variables.

Summary Statistics (CSP Level)

	Mean (1)	Standard Deviation (2)
Hate Crime per quarter	31.33	42.78
Attitudes toward Immigrants	3.43	0.48
Population (1000s)	180.30	118.07
EU migrants per quarter	271.20	452.55
Non-EU migrants per quarter	177.30	329.21
Education Deprivation	21.01	8.53
Austerity	455.82	124.92
Generalized Trust	0.53	0.14
Trust in Politicians	3.11	0.24

Notes: The table reports the mean (column 1) and standard deviation (column 2). All variables are measured at the CSP level. The sample covers 304 CSPs. Hate Crime is the average count of police-recorded hate offenses per CSP per quarter. Data on hate crime are from Office of National Statistics (ONS) and cover Q2 2002 — Q4 2019. Attitudes toward immigrants capture views over economic and cultural contributions of immigrants and are from Wave 1 of the British Election Study (BES). Population is mid-year estimate and is expressed in thousands. Population data is from ONS and it is measured annually. EU migrants and Non-EU migrants capture migrant flows measured using quarterly National Insurance Number (NINO) registrations. Education Deprivation uses Education, Skills and Training domain of the English Indices of Deprivation in 2014-15 based on data from Ministry of Housing, Communities and Local Government (2015). Austerity is financial loss per working-age adult from (Beatty et al., 2013). This data is available for 2010 only. Generalized Trust captures trust in strangers (binary variable) and is available at the earliest from Wave 7 of BES. Trust in Politicians captures trust in politicians in the UK (1-7 scale) and is Wave 1 of BES.

Attitudes over Time.- Figure A.1 shows the stability of attitudes over time.

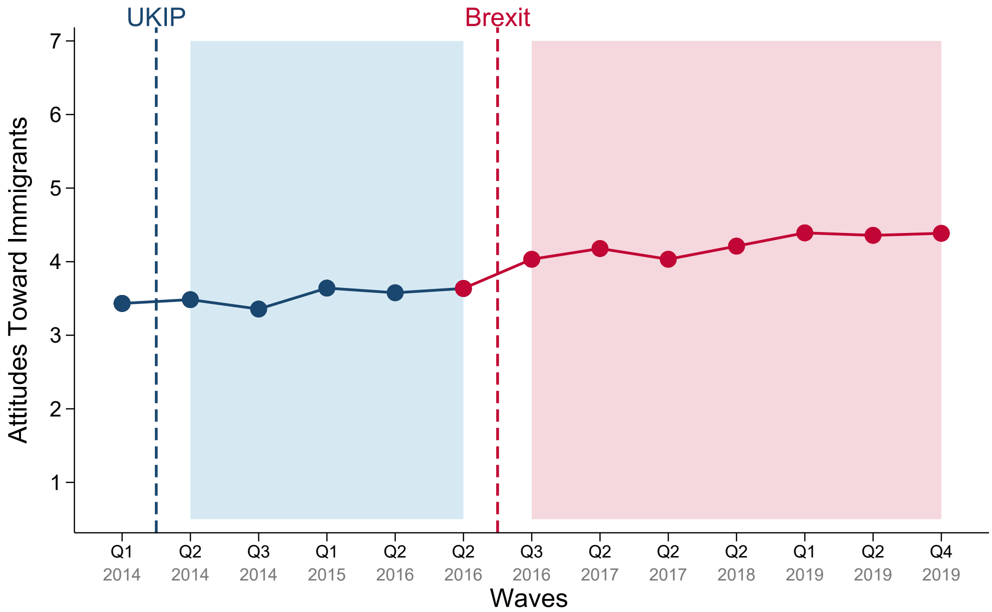

Figure A.1: Evolution of Attitudes Toward Immigrants Over Time

Notes. The figure shows the evolution of attitudes toward immigrants over time. The Y-axis reports attitudes toward immigrants on a scale of 1–7, where 1 indicates very negative and 7 very positive. The X-axis reports calendar quarters corresponding to survey waves. Vertical dashed lines mark UKIP (blue) and Brexit (red) events. Shaded areas indicate the post-UKIP (blue) and post-Brexit (red) periods. Data are from British Election Study.

Attitudes toward Immigrants.— Table A.2 validates our measure of attitudes by showing that these measures are, as expected, negatively correlated with vote share for both UKIP and Leave campaign of Brexit.

Table A.2: Attitudes Toward Immigrants and Political Outcomes

	Dependent Variable: Vote Share (Standardized)	
	UKIP	Brexit
	(1)	(2)
Attitudes toward Immigrants	-0.347	-0.488
	(0.054)	(0.047)
R^2	0.66	0.69
Observations	304	304
Controls	Yes	Yes

Notes: OLS estimates with robust standard errors in parentheses. Column (1) uses standardized UKIP vote share as the dependent variable; Column (2) uses standardized leave share in Brexit referendum. Controls include average age, gender, education level, income, ethnicity and religion.

Alternative Event Study.— Figure A.2 shows event study with hate crime per capita. As before, we find a large surge in hate crime in the post-event periods but a clear absence of trends in the pre-event periods.

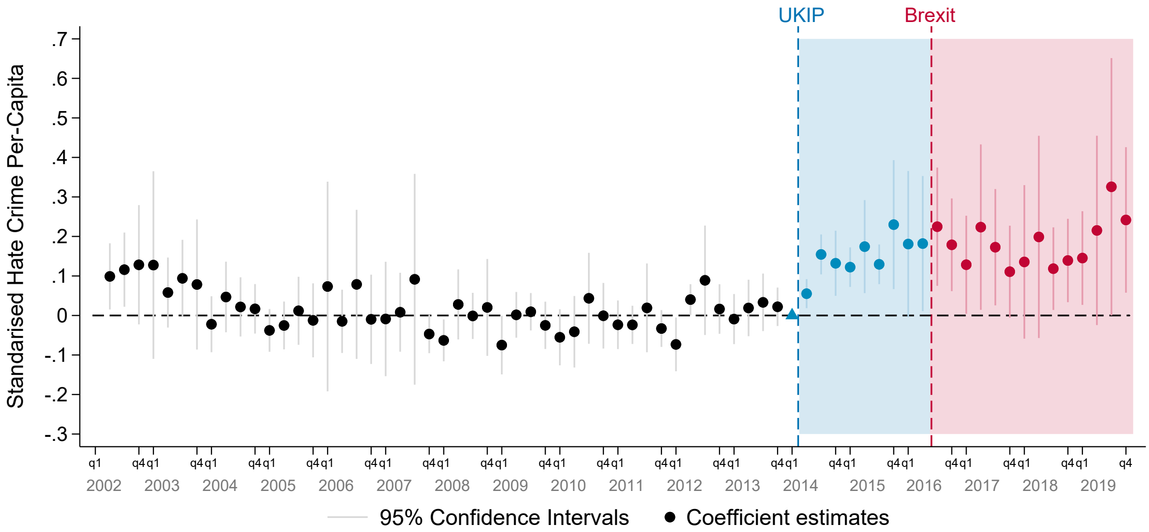

Figure A.2: Event Study Hate Crime Per-Capita

Notes: The figure plots coefficients on $T_{it}^{k} \times Attitudes$ from the estimation of Equation 14 with 95% confidence bands. The Y-axis denotes standardized measure of hate crime per-capita. Black markers correspond to pre-UKIP period, blue markers to post-UKIP event, and red markers to post-Brexit event. The blue triangle corresponds to omitted period of Q1 2014, the blue vertical line to UKIP event, and the red vertical line to Brexit event. On the X-axis, q1 and q4 refer to quarter 1 and 4. Standard errors are clustered at the CSP level.

Reporting Rates by Crime Category.— We match five types of crime in CSEW to 12 different types of hate crime listed in ONS. Table A.3 provides an illustration. Note that in the table column 3 reports average reporting rate over 18 years to demonstrate our point, but this rate actually varies over time.

Table A.3: Crime Equivalency and Reporting Rates to Construct Adjusted Hate Crime Measures

CSEW Offense Type	**Corresponding Hate Crime**	**Average Reporting Rate (%)**
(1)	(2)	(3)
Wounding	• Inflicting grievous bodily harm without intent • Actual bodily harm and other injury • Assault with injury	59.8
Assault with minor injury	• Less serious wounding	39.5
Violence without injury	• Assault without injury • Harassment • Public fear, alarm or distress	38.7
Arson and other criminal damage	• Criminal damage to a building other than a dwelling • Criminal damage to a dwelling • Other criminal damage • Criminal damage	42.5
Criminal damage to a vehicle	• Criminal damage to a vehicle	28.5

Notes: CSEW offense type lists five crime types (see column 1) that have hate equivalents (see column 2). Reporting rates in column 3 are average over 18 years and are for illustrative purposes only. While computing adjusted crime statistics, we use year specific weights. Data are from Table D10 of *"Crime in England and Wales: Annual Trend and Demographic Tables"*

Simultaneously Considering UKIP and Brexit Events.— Table A.4 presents results after considering simultaneously the interaction of attitudes with post-UKIP and post-Brexit events. Our results hold in both magnitude and significance.

Table A.4: Attitudes Toward Immigrants and Hate Crime:
Considering UKIP and Brexit Simultaneously

	Dependent Variable: Standardized Hate Crime
Attitudes × Post-UKIP	0.102
	(0.021)
Attitudes × Post-Brexit	0.106
	(0.033)
R^2	0.87
Observations	21,577
Number of CSPs	304
CSP Fixed Effects	Yes
Quarter Fixed Effects	Yes
Controls	Yes

Notes: OLS estimates with standard errors clustered at the CSP level. Controls include population size, EU migrants, Non-EU migrants, interactions between austerity and the post-2009 indicator, as well as separate interaction of post-event indicator with education deprivation, generalized trust, and trust in politicians.

Reporting Hate Crime.– Table A.5 reports results using adjusted hate crime as the dependent variable. Our results are comparable in magnitude and significance to those obtained using the unadjusted hate crime measure.

Table A.5: Attitudes Toward Immigrants and Hate Crime: Using Adjusted Measure of Hate Crime

	Dependent Variable: Standardized Adjusted Hate Crime
	Panel A: UKIP Event
Attitudes × Post-UKIP	0.085
	(0.023)
R^2	0.88
Observations	17,024
	Panel B: Brexit Event
Attitudes × Post-Brexit	0.222
	(0.042)
R^2	0.85
Observations	21,577
Number of CSPs	304
CSP Fixed Effects	Yes
Quarter Fixed Effects	Yes
Controls	Yes

Notes: OLS estimates with standard errors clustered at the CSP level. The dependent variable is standardized adjusted hate crime. The specification includes controls for population size, EU migrants, Non-EU migrants, interactions between austerity and the post-2009 indicator, as well as separate interactions of post-event indicator with education deprivation, generalized trust, and trust in politicians.

Additional Robustness Checks.— Table A.6 shows that our results are robust to spatial standard errors using 50 km distance (column 1), including grievous and aggravated bodily harm in our measure of hate crime (column 2), terrorist attacks (column 3), hate crime per capita (column 4), and log of hate crime (column 5).

Table A.6: Attitudes Toward Immigrants and Hate Crime:
Additional Robustness Checks

	Spatial Errors (1)	All Hate Crimes (2)	Terrorist Attacks (3)	Hate per capita (4)	Log Hate per capita (5)
	Panel A: UKIP Event				
Attitudes × Post-UKIP	0.170	0.157	–	0.168	0.020
	(0.053)	(0.028)	–	(0.025)	(0.003)
Observations	17,024	17,024	–	17,024	17,024
	Panel B: Brexit Event				
Attitudes × Post-Brexit	0.190	0.181	0.173	0.184	0.018
	(0.039)	(0.037)	(0.031)	(0.056)	(0.004)
Observations	21,577	21,577	21,577	21,577	21,577
Number of CSPs	304	304	304	304	304
CSP Fixed Effects	Yes	Yes	Yes	Yes	Yes
Quarter Fixed Effects	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes
Police × Post-Event	Yes	Yes	Yes	Yes	Yes

Notes: OLS estimates with standard errors clustered at the CSP in parentheses except in column 1 which reports standard errors adjusted for spatial correlation of errors using 50 km distance. In columns 1 to 3 the dependent variable is standardized hate crime. In column 2 we consider all types of hate crimes including grievous and aggravated bodily harm. Column 3 reports results after controlling for an interaction between attitudes and a post-indicator for the terrorist attacks in London and Manchester (2017 Q2). Column 4 uses standardized hate crime per capita. Column 5 uses the log of hate crime per capita. Controls include population size, EU migrants, non-EU migrants, and interaction between austerity and post-2009 indicator, as well as separate interactions of post-event indicator with education deprivation, generalized trust, trust in politicians.

Alternative Measures of Attitudes.— Table A.7 shows that our results are robust to considering economic and cultural attitudes separately (column 1-2), and using attitudes toward immigrants from Wave 7 of BES (column 3).

Table A.7: Attitudes Toward Immigrants and Hate Crime:
Alternative Measures of Attitudes

	Dependent Variable: Standardized Hate Crime		
	Attitudes (Economy) (1)	Attitudes (Culture) (2)	Attitudes (Wave 7) (3)
	Panel A: UKIP Event		
Attitudes × Post-UKIP	0.162	0.166	
	(0.030)	(0.029)	
Observations	17,024	17,024	
	Panel B: Brexit Event		
Attitudes × Post-Brexit	0.183	0.184	0.167
	(0.037)	(0.039)	(0.038)
Observations	21,577	21,577	21,577
Number of CSPs	304	304	304
CSP Fixed Effects	Yes	Yes	Yes
Quarter Fixed Effects	Yes	Yes	Yes
Controls	Yes	Yes	Yes

Notes: OLS estimates with standard errors clustered at the CSP in parentheses. Columns (1) and (2) use economic and cultural components of attitudes toward immigrants, respectively. Column (3) uses measures of attitudes toward immigrants from Wave 7 of the BES. It includes controls for population size, EU migrants, non-EU migrants, and interaction between austerity and post-2009 indicator, as well as separate interactions of post-event indicator with education deprivation, generalized trust, trust in politicians.

Falsification Test.— Table A.8 reports results from a falsification test in which we use other crimes like burglary and murder as the dependent variable.

Table A.8: Attitudes Toward Immigrants and Hate Crime: Falsification Test

	Dependent Variable	
	Standardized Murder (1)	Standardized Burglary (2)
	Panel A: UKIP Event	
Attitudes × Post-UKIP	-0.001	-0.046
	(0.030)	(0.026)
R^2	0.674	0.897
Observations	17,024	17,024
	Panel B: Brexit Event	
Attitudes × Post-Brexit	0.015	0.001
	(0.042)	(0.026)
R^2	0.63	0.89
Observations	21,577	21,577
Number of CSPs	304	304
CSP Fixed Effects	Yes	Yes
Quarter Fixed Effects	Yes	Yes
Controls	Yes	Yes

Notes: OLS estimates with standard errors clustered at the CSP level. Columns (1) and (2) use standardized murder and standardized burglary, respectively. Controls are for population size, EU migrants, non-EU migrants, and interaction between austerity and post-2009 indicator, as well as separate interactions of post-event indicator with education deprivation, generalized trust, trust in politicians.

Table A.9 shows that our results hold when we control for other crimes that are unrelated to hate.

Table A.9: Attitudes Toward Immigrants and Hate Crime: Controlling for Other Crimes

	Dependent Variable: Standardized Hate Crime
	Panel A: UKIP Event
Attitudes × Post-UKIP	0.164
	(0.030)
R^2	0.89
Observations	17,024
	Panel B: Brexit Event
Attitudes × Post-Brexit	0.189
	(0.038)
R^2	0.87
Observations	21,577
Number of CSPs	304
CSP Fixed Effects	Yes
Quarter Fixed Effects	Yes
Controls	Yes
Other Crimes	Yes

Notes: OLS estimates with standard errors clustered at the CSP level. The specification controls for other crimes, defined as the aggregation of murder and burglary. It includes controls for population size, EU migrants, non-EU migrants, and interaction between austerity and post-2009 indicator, as well as separate interactions of post-event indicator with education deprivation, generalized trust, trust in politicians. Data on other crimes is from ONS.

Survey Data.— Table A.10 reports results from the survey on expression of views toward immigrants.

Table A.10: Changes in Public Expression of Views on Immigrants in the Post-Event Period

	Dependent variable: Change in Expression	
	Without controls (1)	With controls (2)
Anti-Immigrant Attitude (Individual)	0.262	0.241
	(0.032)	(0.032)
Pro-Immigrant Area	0.098	0.094
	(0.031)	(0.030)
Anti-Immigrant Attitude × Pro-Immigrant Area	0.101	0.098
	(0.027)	(0.027)
R^2	0.057	0.067
Observations	1,448	1,448
Controls	No	Yes

Notes: OLS estimates with robust standard errors in parentheses. The dependent variable "Change in expression" and measures the change in how likely respondents are to express their opinions on immigrants following the Brexit referendum. Controls include individual age, gender, education level, income and employment. The dependent variable, anti-immigrant attitude (individual), pro-immigrant area, and their interactions are standardized so the coefficient can be interpreted as elasticities.